

國立臺灣大學電機資訊學院資訊網路與多媒體研究所

碩士論文

Graduate Institute of Networking and Multimedia

College of Electrical Engineering and Computer Science

National Taiwan University

Master Thesis

3-D HD-YOLOv4: 三維超感知域雙分支預測

YOLOv4 於肺部電腦斷層掃描影像結節偵測

3-D HD-YOLOv4: 3-D Hyper Receptive Field and Dual
Head YOLOv4 for Pulmonary Nodule Detection in Lung

CT

陳子安

Tzu-An Chen

指導教授：張瑞峰 博士

Advisor: Ruey-Feng Chang, Ph.D.

中華民國 110 年 7 月

July 2021

口試委員會審定書



國立臺灣大學碩士學位論文 口試委員會審定書

3-D HD-YOLOv4: 三維超感知域雙分支預測 YOLOv4
於肺部電腦斷層掃描影像結節偵測

3-D HD-YOLOv4: 3-D Hyper Receptive Field and Dual
Head YOLOv4 for Pulmonary Nodule Detection in Lung
CT

本論文係陳子安君（學號 R08944043）在國立臺灣大學資訊網路與多媒體研究所完成之碩士學位論文，於民國一百一十年七月三十日承下列考試委員審查通過及口試及格，特此證明

口試委員：

張 溢 峰

（簽名）

（指導教授）

羅 崇 鈞

陳 良 碩

所 長：

施 吉 昇

致謝



以跨領域身份進入電資學院學習實屬不易，所幸在過程中遇見了許多貴人，因為有你們的協助，我才能如此順利的完成此篇論文，並帶著這份追夢的憧憬與專業知識繼續邁進。首先要感謝的是指導教授張瑞峰博士，老師沒有因為非本科系的身份而拒收我當研究生，並提供了良好的研究環境及頂尖的運算資源，讓我在學習、研究的過程中可以竭盡所能、盡情嘗試。接著要感謝的是學長黃耀賢博士，有了實驗室大家長的你坐鎮，讓大家心安了不少，經過你的指點，任何問題都可以迎刃而解，凡事總是親力親為又平易近人，在此預祝你未來的教職生涯一帆風順。再來則是要感謝實驗室的好夥伴們（天毅、聖智、以尊、子傑、張鈞、秉儒、子欽、戴勤、易鑫），剛踏入碩一時有你們的暖心相待，讓這兩年的碩士生涯非常精彩，平常閒聊、八卦、運動、遊戲的時光歷歷在目，希望大家都有璀璨的未來發展，飛黃騰達的時候要記得還有小弟我啊！最後要感謝我的家人，毅然決然的支持我念碩班的決定，並讓我在學習的過程中無後顧之憂，真的很感謝你們！

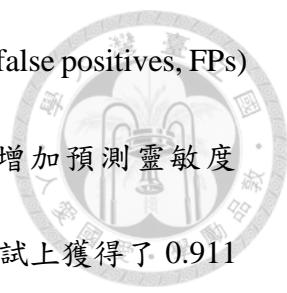
最後的期勉，希望自己能夠秉持著踏入臺灣大學電機資訊學院資訊網路與多媒體研究所時的初心，儘管將來會遇到許多困難，仍能以不辭辛勞的意志與毅力克服一切，充實一切相關能力及技術，並對生活事物保持著熱情及好奇，繼續走下去，成為一個更好的人。

摘要



低顯影劑量電腦斷層掃描造影 (Low-dose Computed Tomography, LDCT) 技術被廣泛應用於肺癌初期的結節偵測，以提高病患的存活率。然而，觀察數百張的 LDCT 影像及每張掃描影像內部的小型結節非常耗時。因此，電腦輔助偵測 (Computer-aided Detection, CADe) 系統被用於加速檢測流程及減輕放射科醫師的負擔。由於近年來深度學習技術於電腦視覺領域的成功發展，越來越多以卷積神經網路 (Convolutional Neural Network, CNN) 為基礎的 CADe 系統被應用於醫學使用上。

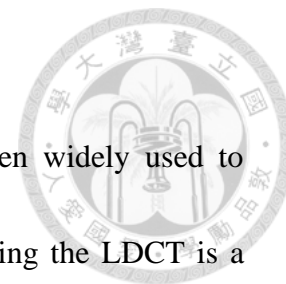
此研究提出了一以 CNN 為基礎的肺部結節偵測 CADe 系統，該系統包含了資料前處理、肺部結節偵測及初步結果後處理，共三個階段。在資料前處理的階段中，影像的間距被標準化為相同，並對去除肺部外區域後的影像進行了正規化處理，再將影像裁切為具有相同格式的數個部分。接著，經過前處理的影像會作為偵測模型的輸入，並由模型輸出一序列的候選結節及其位置、直徑大小。所提出的 CNN 偵測模型被取名為 3-D Hyper Receptive Field and Dual Head YOLOv4 (3-D HD-YOLOv4)，是修改自 YOLOv4 網路的架構。為了提升原 YOLOv4 所萃取特徵的代表性，加入了壓縮激勵 (Squeeze-and-excitation, SE) 模塊和感知域 (Receptive Field Block, RFB) 模塊。此外，雙頭 (Dual Head, D-head) 模塊和跨階平行分支 (cross stage parallel branch) 機制也被用於增強原 YOLOv4 的結節偵測能力。最後，在初步結果後處理的部分，對於相同結節的重複預測將會被去除。



根據實驗結果顯示，SE 和 RFB 模塊可以大幅降低誤報 (false positives, FPs) 的數量，以及 D-head 模塊和跨階平行分支機制可以大幅增加預測靈敏度 (Sensitivity)。除此之外，所提出的 CADe 系統在公開的資料集測試上獲得了 0.911 的競爭績效指標分數 (Competition Performance Metric, CPM)，在誤報率 (FPs per Scan) 為 8 的靈敏度為 0.982。相較於其他最新技術的系統，所提出的 CADe 系統展現了其優異的表現。整體結果指出，我們針對模型的改進可以大幅提升其對於結節的偵測能力，並達到目前最先進技術的表現。

關鍵詞：低顯影劑量電腦斷層掃描，電腦輔助偵測系統，卷積神經網路，單階段物件偵測，YOLOv4

Abstract



Low-dose computed tomography (LDCT) screening has been widely used to detect lung nodules and reduce early mortality. However, reviewing the LDCT is a time-consuming process due to hundreds of slices and the small size of the nodule in each scan. For lightening the burden of radiologists, the computer-aided detection (CADe) system is proposed as the second review tool to accelerate the procedure. In recent years, the convolutional neural network (CNN) based CADe system becomes popular in medical usage because of the success of deep learning in computer vision.

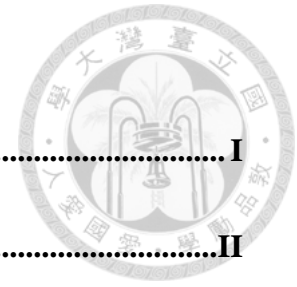
In this study, a CNN-based CADe system containing the data preprocessing, the lung nodule detection, and the initial result post-processing is proposed for the pulmonary nodule detection. In the data preprocessing, the spacing equalization standardizes the images to a consistent spacing, the lung area regularization removes the area outside the lungs and normalizes the intensity of the voxels, and the volume of interest (VOI) extraction divides the LDCT scan into numerous VOIs with consistent format. Then, the extracted VOIs are fed into the proposed detection model, three-dimensional (3-D) HD-YOLOv4 modified from the YOLOv4 for the nodule detection. To increase the representation of the extracted feature in YOLOv4, the squeeze-and-excitation (SE) module and the receptive field block (RFB) module are embedded. Furthermore, the dual head (D-head) module and the cross stage parallel branch

architecture are adopted to enhance the nodule detection capability of YOLOv4. Lastly, the initial result post-processing eliminates the duplicate detections on the same nodule.

Our experiment results show that the SE and RFB modules significantly reduce the number of false positives (FPs). The D-head module and the cross stage parallel branch mechanism could dramatically increase the sensitivity. Moreover, the proposed system obtains the 0.911 competition performance metric (CPM) score on the public dataset, and the sensitivity at eight false positives per scan is 0.982. Compared with other state-of-the-art systems, the proposed CADe system shows its outperformance. These results indicate that the modifications to the detection model can remarkably improve its nodule detection capability and achieve state-of-the-art performance.

Keywords: Low-dose computed tomography, Computer-aided detection system, Convolutional neural network, One-stage object detection, You Only Look Once version 4.

Table of Contents



口試委員會審定書	I
致謝.....	II
摘要.....	III
Abstract.....	V
Table of Contents	VII
List of Figures.....	IX
List of Tables	XI
Chapter 1 Introduction.....	1
Chapter 2 Materials	5
Chapter 3 Method	7
3.1 Data Preprocessing.....	9
3.2 Lung Nodule Detection.....	10
3.2.1 3-D YOLOv4	10
3.2.2 3-D HD-YOLOv4	13
3.2.2.1 SE-CSP Stage Block.....	14
3.2.2.2 Hyper Receptive Field	15
3.2.2.3 Dual Head and Cross Stage Parallel Branch.....	17
3.3 Initial Result Post-processing	20

3.4	Loss Function.....	21
3.5	Model Training	22
Chapter 4	Results and Discussion.....	24
4.1	Experiment Environment.....	24
4.2	Evaluation	24
4.3	Experiment Results	25
4.3.1	Ablation Study	25
4.3.2	Comparison with SOTA	27
4.4	Discussion	28
Chapter 5	Conclusions and Future Works	34
Reference	36



List of Figures



Fig. 2-1 The distribution of nodule size (mm).....	6
Fig. 3-1 The flowchart of the proposed CADe system.	8
Fig. 3-2 The flowchart of the data preprocessing.	10
Fig. 3-3 The fundamental operation units: (a) CSP stage block, (b) CSPBlock, and (c) ConvBlock.....	12
Fig. 3-4 The architecture of the proposed 3-D HD-YOLOv4.	13
Fig. 3-5 The structures of the (a) SE module, the (b) SE-CSP stage block, and the (c) SE-CSPBlock.	15
Fig. 3-6 The structure of the RFB module.....	17
Fig. 3-7 The structures of the (a) Conv-head module and the (b) D-head module..	19
Fig. 3-8 The architecture of the cross stage parallel branch.	20
Fig. 3-9 The flowchart of the initial result post-processing.....	21
Fig. 4-1 The FROC performance comparison of different models.....	26
Fig. 4-2 The comparison of different models with the number of false negatives grouped by the diameter d (mm).....	27
Fig. 4-3 The detected cases: (a) small nodule with size 3.83 mm, (b) medium nodule with size 6.27 mm surrounded by lung tissues, and (c) large nodule	

with size 25.87 mm. (The ground truth is the bounding box in red, and the predicted bounding box is in blue.).....32

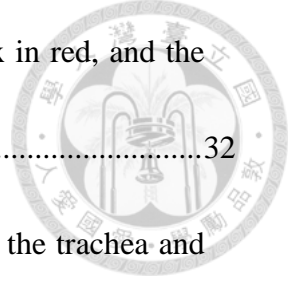


Fig. 4-4 The misjudgment detections: (a) suspicious region near the trachea and (b) unclear edge small nodule with size 4.70 mm. (The ground truth is the bounding box in red, and the predicted bounding box is in blue.).....33

List of Tables

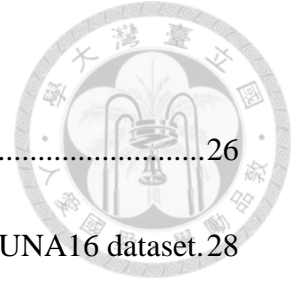
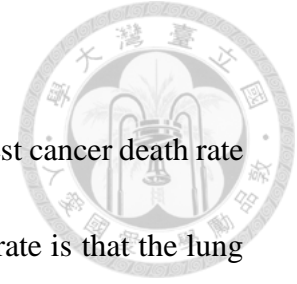


Table 4-1 The CPM scores comparison of the different models.26

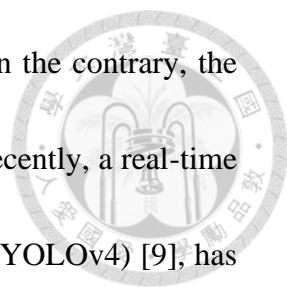
Table 4-2 The comparison with state-of-the-art methods on the LUNA16 dataset.28

Chapter 1 Introduction



Lung cancer is one of the most common cancers, with the highest cancer death rate (18%) in the world [1]. The primary reason for the high mortality rate is that the lung cancer symptoms only become apparent after the advanced stage [2]. The best approach to increasing the survival rate is to detect and diagnose early; after that, the medical interventions could treat the patients. Computed tomography (CT) screening is a common approach to detect lung nodules, and the effectiveness of annual low-dose computed tomography (LDCT) screening for lung cancer has been verified in clinical trials [3-5]. However, making large volumes of LDCT scans screening accurate and efficient is a time-consuming process. In this case, the computer-aided detection (CADe) system is a helpful tool for assisting radiologists in accelerating the LDCT review.

Computer vision has advanced by leaps and bounds in the past few years due to the great success of deep learning adaptation on convolutional neural networks (CNNs) [6]. According to the reports [7, 8], developing a CAdE system with CNN architecture in medical images is a trend because of the automatic abilities of feature extraction for object detection and the outstanding performance. These powerful CNN detection models are usually categorized into the one-stage and two-stage object detectors. In general, two-stage detectors are composed of a region proposal stage and a false positives (FPs) reduction stage. This configuration which needs more parameters and



training time, leads to both higher accuracy and model latency. On the contrary, the one-stage object detectors have lower accuracy and faster speed. Recently, a real-time one-stage object detection model, You Only Look Once version 4 (YOLOv4) [9], has been verified that it has the state-of-the-art (SOTA) performance for the task of object detection on the common objects in context (COCO) dataset. The most distinguishing features of YOLOv4 are its real-time latency, lightweight network, and high enough accuracy. Due to the remarkable performance of YOLOv4 on common object detection, it is chosen as the main network architecture for nodule detection in this study.

However, the original YOLOv4 is designed and optimized for 2-D common object detection, which is considerably different from the lung nodule detection in 3-D LDCT. In this study, we focus on the following three perspectives to modify the architecture of YOLOv4. First, there are RGB three channels in the common image while there is only one grayscale channel in the 3-D LDCT. Therefore, more representative and informative features extracted by the CNN model from the LDCT are required to obtain better nodule detection performance. Second, compared to the common object detection, the area ratio of the nodule in the whole 3-D LDCT is extremely small. In this case, the down-sample operations in the YOLOv4 might easily cause subtle information vanishing during the dimension reduction. Last, the common object detection is usually a more than hundreds-of-classes classification task, but the lung

nodule detection is a two-class classification problem. The channel size adjustment is necessary to lower the memory demand of 3-D LDCT and make the optimization process more efficient.



In 2017, Hu *et al.* proposed the squeeze-and-excitation network (SENet) constructed by the squeeze-and-excitation (SE) module for the image tasks in the ImageNet large scale visual recognition challenge (ILSVRC) 2017 [10]. The SENet reduces the top-5 error rate to 2.25% of the 2016 winning entry in the classification task with the channel-wise attention mechanism. The result shows that the SE module could enhance the representative of the extracted features and obtain better classification results. In 2018, the receptive field block (RFB) module was designed by Liu *et al.*, and the RFB network (RFB Net) delivered a significant performance on the Pascal VOC and COCO dataset [11]. The RFB module is a lightweight structure that can expand the receptive field size of the feature map to obtain more informative and discriminative features. These two modules could significantly improve the quality of the feature extraction in the CNN model.

In 2020, the dual head (D-head) module, which combined the fully connected head (FC-head) for the classification task and the convolution head (Conv-head) for the regression task, was delivered by Wu *et al.* [12]. The D-head module gains 3.5 average precision (AP) improvement on the COCO dataset from the feature pyramid network

(FPN). This result indicates that the D-head module could boost the detection ability significantly with only a few parameters increased. On the other hand, the down-sample operation might cause the primitive information to vanish, especially small nodules with sizes less than five millimeters. To address this issue, the cross stage parallel branch mechanism is proposed in our detection model. The extra feature map generated from this mechanism could remarkably increase the sensitivity of the nodule detection task. These two modules could reinforce the nodule detection ability for the CNN model.

In this study, a CNN-based CADe system, 3-D Hyper Receptive Field and Dual Head YOLOv4 (3-D HD-YOLOv4) modified from the YOLOv4, is developed for nodule detection in 3-D LDCT scans. The introduction of the architecture of the proposed CADe system is in the following chapters. The materials used for the experiments and evaluations are mentioned in Chapter 2. Next, the details of the proposed CADe system are elaborated in Chapter 3. In Chapter 4, the experiment results evaluation, including the ablation study and the comparisons with other SOTA methods, are presented. Lastly, the conclusions and the future works are stated in Chapter 5.

Chapter 2 Materials



The Lung Nodule Analysis 2016 (LUNA16) [13] is a public online challenge organized by several researchers from the Netherlands and Italy in 2016. The purpose of this challenge is to enable participants to test their algorithms or proposed methods on a common database under a standardized evaluation procedure. The LUNA16 dataset is a subset of the Lung Image Database Consortium image collection (LIDC-IDRI) [14] which consists of thoracic low-dose computed tomography (LDCT) scans with lesion annotations determined independently by four experienced radiologists. The lesions contain nodules and non-nodule, and each nodule is also annotated with coordinates and diameter.

The LUNA16 dataset comprises the LDCT scans with slice thickness ≤ 2.5 mm and the nodules with diameter ≥ 3 mm. Besides, each nodule annotation is accepted by at least three radiologists. Hence, there are 888 LDCT scans, including 1,186 nodules with a mean size of 8.31 ± 4.76 mm, in the LUNA16 dataset. The distribution of nodule size is illustrated in Fig. 2-1.

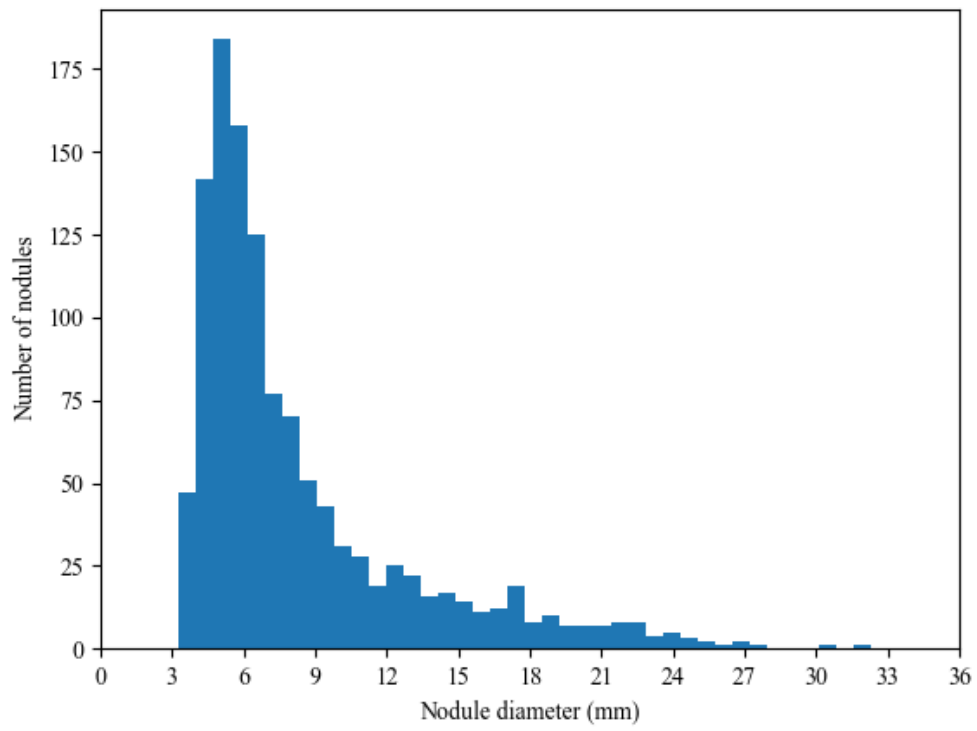
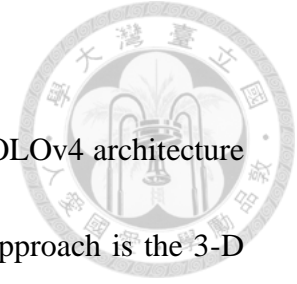


Fig. 2-1 The distribution of nodule size (mm).

Chapter 3 Method



This study proposes an end-to-end CADe system based on YOLOv4 architecture for lung nodule detection in 3-D LDCT scans. The input of our approach is the 3-D LDCT scan, and the output is the coordinates and diameter of each detected nodule candidate. The proposed CADe system is composed of data preprocessing, lung nodule detection, and initial result post-processing. Firstly, the data preprocessing is performed to divide each 3-D LDCT scan into volumes of interest (VOIs) with a consistent format. Then, the VOIs are fed into the nodule detection model 3-D HD-YOLOv4, to generate a list of nodule candidates with confidence scores, coordinates, and diameters as an initial detection result. However, these nodule candidates may contain duplicate detections of the same nodule, overlapping with each other. Thus, the initial result post-processing is applied to the initial detection result to eliminate the overlapping candidates and obtain the final result. The flowchart of the CADe system is illustrated in Fig. 3-1.

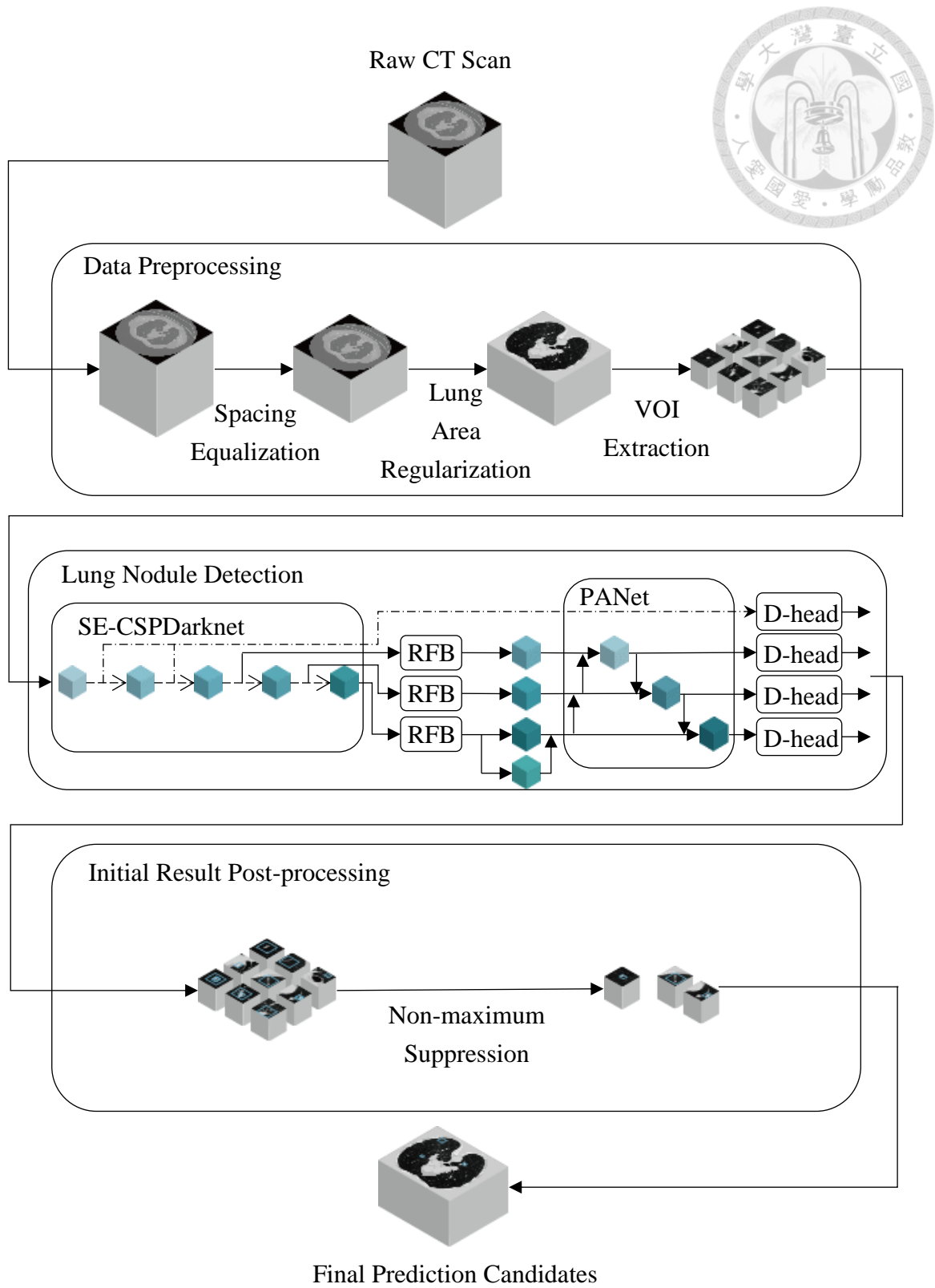
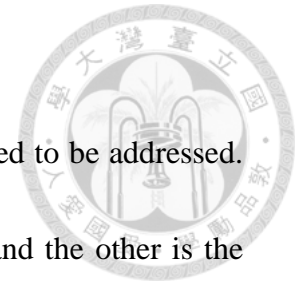


Fig. 3-1 The flowchart of the proposed CADe system.

3.1 Data Preprocessing

Before executing the lung nodule detection, two problems need to be addressed.

One is the slice spacing inconsistency of the used LDCT scans, and the other is the memory demand of the lung nodule detection model with 3-D input. Therefore, the data preprocessing, including the spacing equalization and the image modulation, is performed to overcome the problems. The slice spacing problem is addressed in the spacing equalization by adopting the nearest neighbor interpolation method [15]. The memory demand problem is subdued in image modulation by performing the lung area regularization and the VOI extraction in each 3-D LDCT scan sequentially. The lung area regularization is composed of noise area elimination and voxel intensity normalization. The noise area elimination is to remove the useless areas where is outside the lungs. Then, the voxel intensity normalization is applied to convert the intensity of each voxel into the range of 0 to 255. After the lung area regularization, the VOI extraction is adopted to divide the regularized LDCT scan into several VOIs with a fixed size of $80 \times 80 \times 80$ pixels. The flowchart of the data preprocessing is illustrated in Fig. 3-2.



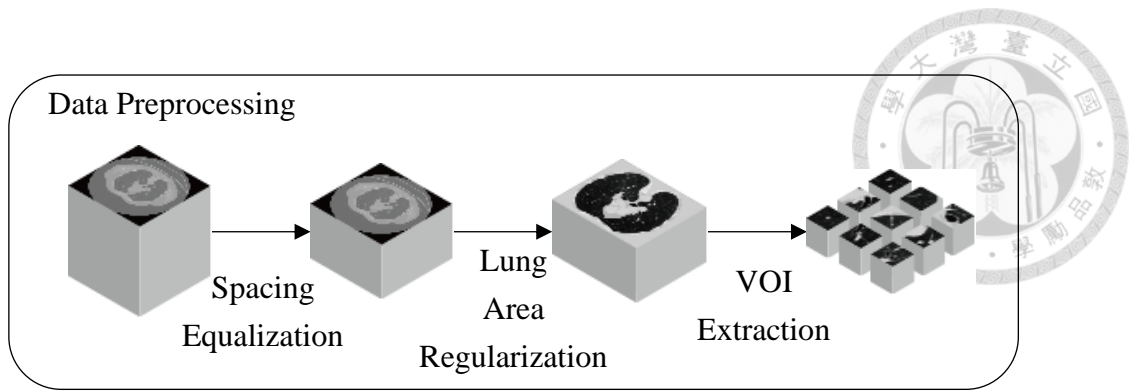


Fig. 3-2 The flowchart of the data preprocessing.

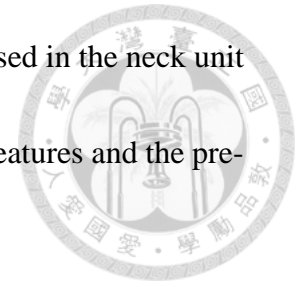
3.2 Lung Nodule Detection

After performing the data preprocessing, the VOIs are fed to our nodule detection model, 3-D HD-YOLOv4, to generate nodule candidates. The 3-D HD-YOLOv4 is modified from the YOLOv4 [9] architecture by integrating hyper receptive field and dual head structure to enhance the nodule detection capability of YOLOv4. The details of the 3-D HD-YOLOv4 model are described in the following sections.

3.2.1 3-D YOLOv4

YOLOv4 is a one-stage anchor-based detection method that predicts the location and category of the detected object simultaneously with the pre-defined anchors. The framework of YOLOv4 is composed of the backbone, the neck, and the head units. First, the input image is encoded as the different level representative feature maps in the backbone unit by the cross stage partial (CSP) Darknet [9], constructed based on

five CSP stage blocks. Then, the different level feature maps are fused in the neck unit to generate more informative feature maps. Finally, the combined features and the pre-defined anchors are utilized to predict objects in the head unit.



The CSP stage block, responsible for reducing the computation cost by down-sample operation and enhancing the model capability in learning, is the critical feature map generator in the backbone unit. As shown in Fig. 3-3(a), a CSP stage block includes N CSP blocks (CSPBlock) and four convolution blocks (ConvBlock). The CSPBlock illustrated in Fig. 3-3(b) is made up of two ConvBlocks and the skip connection, and the ConvBlock presented in Fig. 3-3(c) is composed of a convolution layer, a batch normalization layer [16], and a Mish activation function [17].

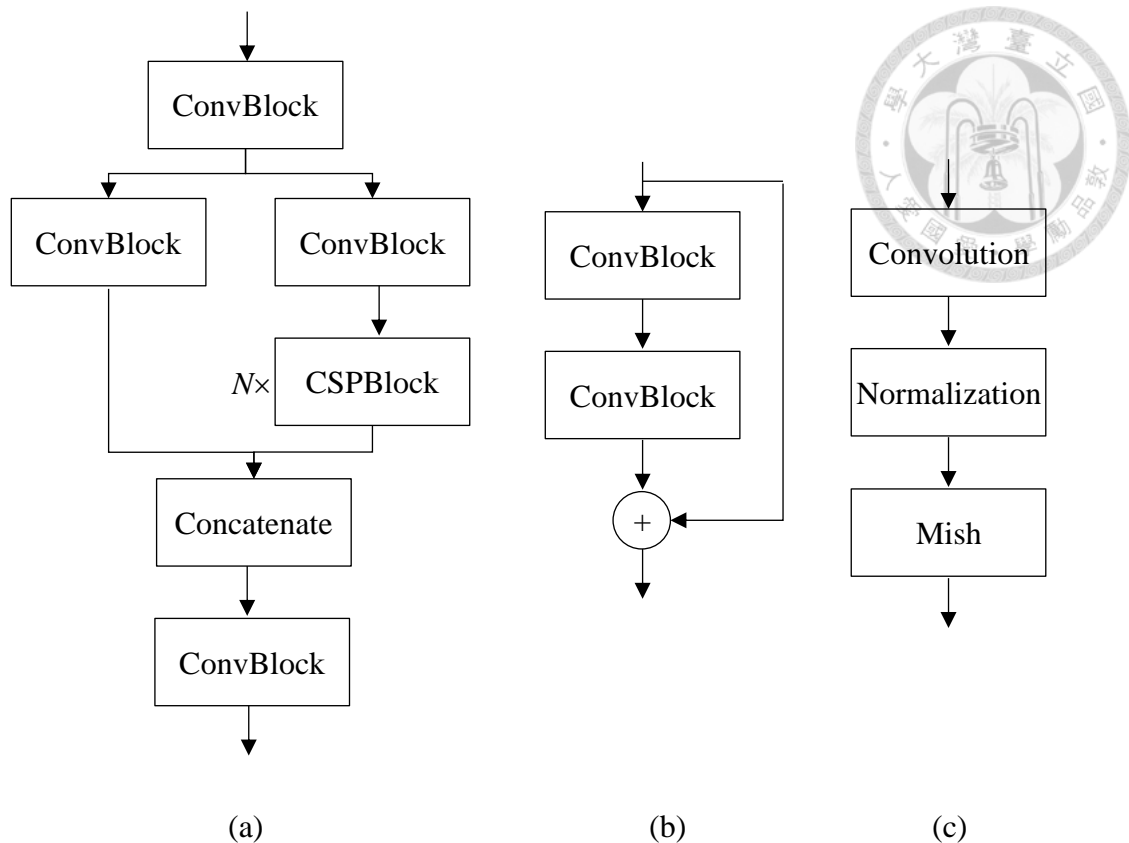
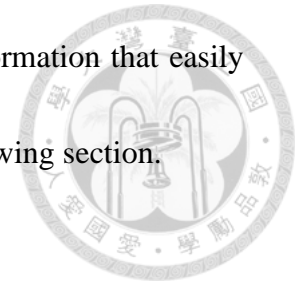


Fig. 3-3 The fundamental operation units: (a) CSP stage block, (b) CSPBlock, and (c) ConvBlock.

For detecting nodule in the LDCT image, the primitive YOLOv4 is reconstructed and modified by using the 3-D convolution and declining the channel size of the CSP stage blocks from [64, 128, 256, 512, 1024] to [64, 64, 64, 64, 64] in the backbone unit. However, the capability of 3-D YOLOv4 might be insufficient if adopting 3-D YOLOv4 directly to observe suspicious nodules in the 3-D LDCT image only through the above modifications. Hence, in this study, two approaches are employed to reinforce the detection ability. The additional modules are embedded in the backbone unit and the neck unit to highlight and absorb more discriminative information while performing feature extraction and feature fusion. The other is that partial routes of the

framework are split into parallel branches to retain the subtle information that easily vanished during the process. The details are introduced in the following section.



3.2.2 3-D HD-YOLOv4

In this study, a 3-D HD-YOLOv4 integrates with the SE module [10], the hyper receptive field module [11], and the (D-head structure [12] in the different units of 3-D YOLOv4 is proposed for nodule detection. The SE-CSP stage block replaces the CSPBlock with the SE module in the backbone unit, which replaces the CSP stage block to highlight the critical features. The hyper receptive field module [11] is utilized in the neck unit to broaden the receptive field [18] of convolution to absorb more information. The dual head and the cross stage parallel branch are adopted in the head unit to improve nodule localization and determination capability. The architecture of the proposed 3-D HD-YOLOv4 is illustrated in Fig. 3-4.

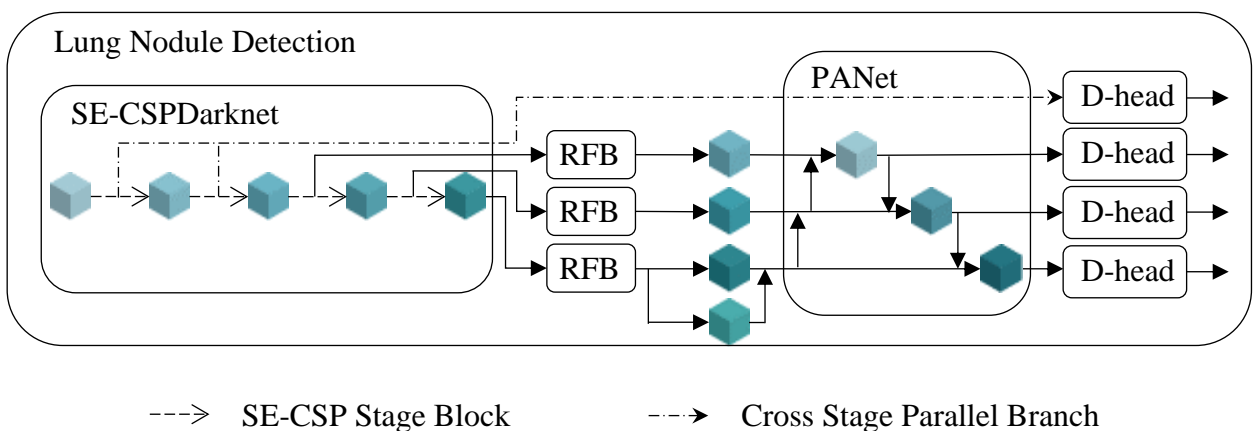


Fig. 3-4 The architecture of the proposed 3-D HD-YOLOv4.

3.2.2.1 SE-CSP Stage Block



An attention mechanism is an approach that distributes the allocation of available computational resources to the most informative usage [19]. The attention operator can generally be incorporated with several layers representing higher-level cognition for adaptation between modalities. The SE module is a lightweight attention operator used to reweight the features with gate mechanism during extraction by performing dynamic channel-wise feature recalibration [10]. The critical features contributing to detection or discrimination could be sieved out for performance improvement with the reweighting. The structure of the SE module shown in Fig. 3-5(a) is composed of an adaptive average pooling (AvgPooling) layer, two fully connected (FC) layers, a rectified linear unit (ReLU) activation function [20], a sigmoid activation function [21], and a multiplication operation. The AvgPooling layer is used to obtain the global average value in each channel of the feature map. Then the channel-wise weight is produced by transforming the average value through the following two FC layers and two activation functions. After the multiplication operation, the reweighted feature containing more discriminative information is generated. Therefore, the SE-CSP stage block constructed with the SE-CSPBlocks is used to create and highlight the essential feature maps in this study. The SE-CSPBlock combines the SE module and CSPBlock and is built by inserting the SE module before performing the addition operation of skip



connection. The structures of the SE-CSP stage block and SE-CSPBlock are presented in Fig. 3-5(b) and Fig. 3-5(c), respectively.

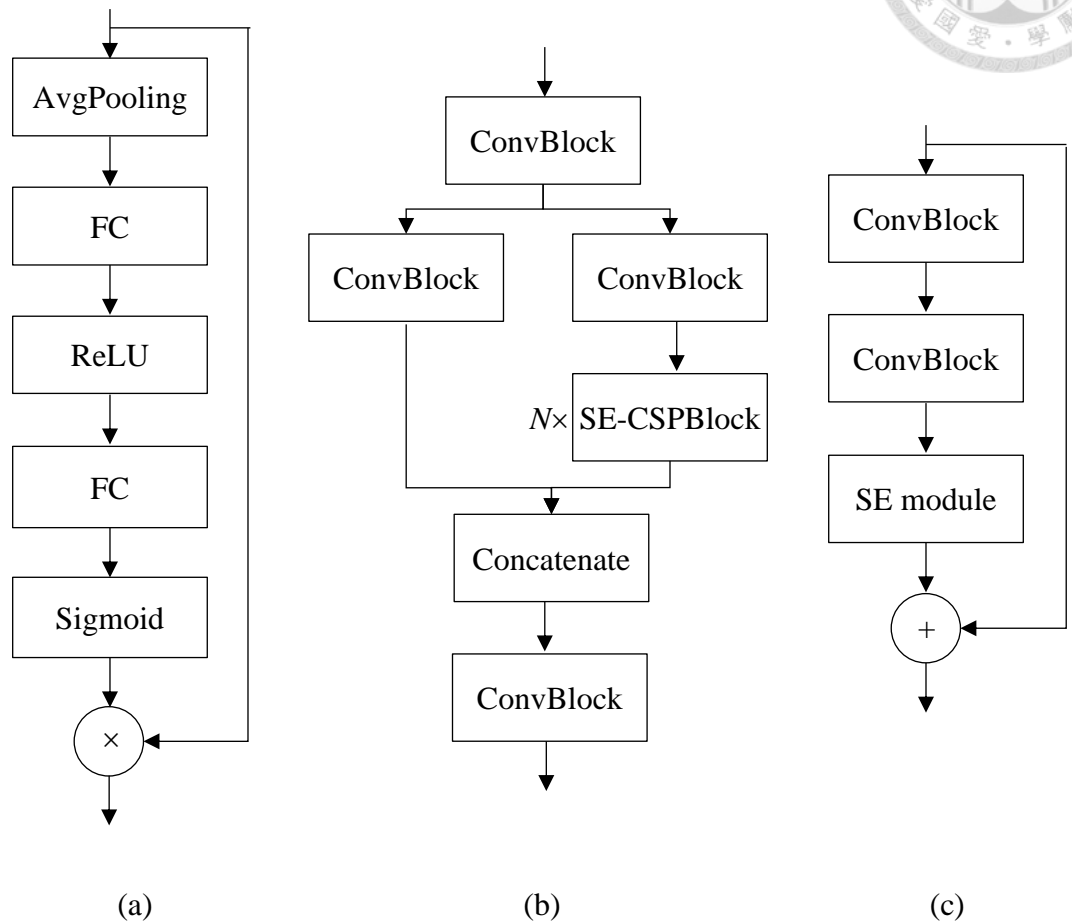
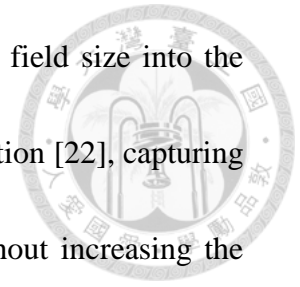


Fig. 3-5 The structures of the (a) SE module, the (b) SE-CSP stage block, and the (c) SE-CSPBlock.

3.2.2.2 Hyper Receptive Field

The receptive field is the region in the input image that can be perceived by the model [18]. The greater receptive field size the model receives, the more context information from the entire image the model can receive. Thus, increasing the receptive field might improve the model performance [9]. Therefore, the RFB module [11] is

cascaded behind the backbone unit to apply the greater receptive field size into the detection model. The RFB module is built with the dilated convolution [22], capturing more context information with different receptive field sizes without increasing the parameters. As illustrated in Fig. 3-6, the RFB module mainly comprises multi-branch ConvBlocks with a trailing ConvBlock with different dilated rates and the skip connection where the dilation rate is the distance between two positions in the convolution layer. In the RFB module, the dilation rate in each trailing ConvBlock is set to the number of ConvBlocks in each branch. With the usage of the RFB module, the receptive field size of the feature map is significantly increased.



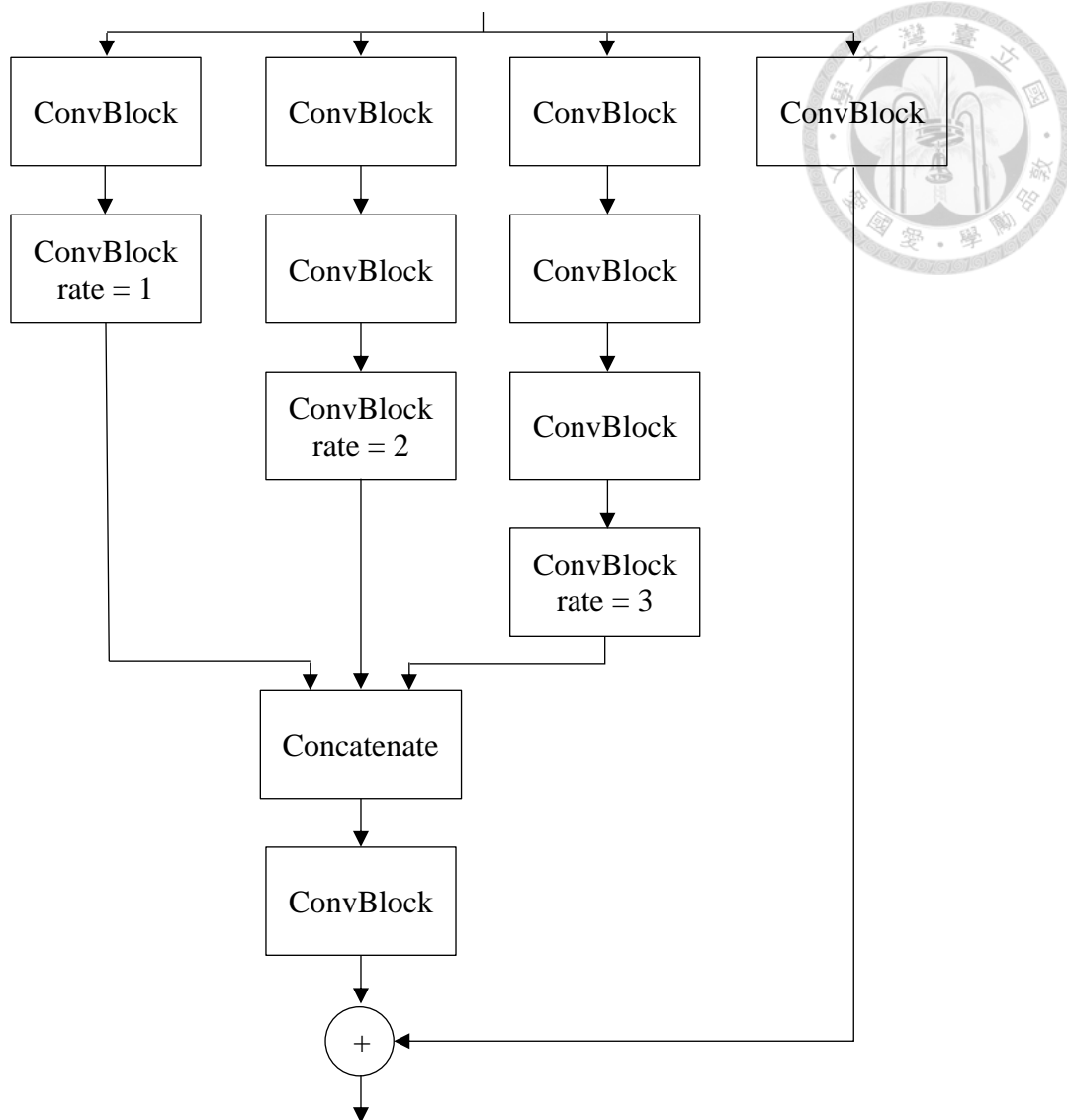


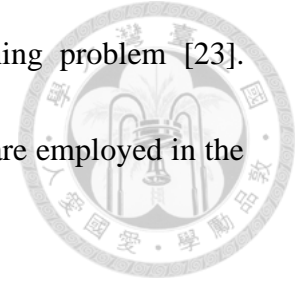
Fig. 3-6 The structure of the RFB module.

3.2.2.3 Dual Head and Cross Stage Parallel Branch

In the CNN detection model, the localization and determination of the Conv-head module in the head unit and the dimension reduction in the backbone unit are often adopted. However, the Conv-head module and the dimension reduction will result in

the weaker recognition ability [12] and the information vanishing problem [23].

Therefore, the D-head module and the cross stage parallel branch are employed in the head unit to overcome the issues.



First, to overcome the weaker recognition ability, the D-head module is developed for the ability improvement. The notion of D-head is to exploit the features obtained from the neck unit more properly than Conv-head. The structure of the Conv-head module, as illustrated in Fig. 3-7(a), is built with a ConvBlock and a convolution layer. Without the FC layer, the Conv-head is more proper to capture the entire object context for bounding box regression but weaker for object recognition. Therefore, to improve the performance of localization and recognition simultaneously, the D-head module, which includes a Conv-head and an FC-head module, is employed to replace the Conv-head module in the head unit. Fig. 3-7(b) shows the architecture of the D-head module. The FC-head module is constructed only with two FC layers and a ReLU activation function. Each head of D-head will generate a set of coordinates, diameter, and a confidence score for a detection result, respectively. Because the purpose of the D-head is to improve the recognition capability, the coordinates and diameter are determined only by Conv-head, but the confidence score is determined by the following definition:

$$S = S_{FC} + (1 - S_{FC}) \times S_{Conv} \quad (1)$$

where S , S_{FC} , and S_{Conv} , are the final confidence score, the FC-head confidence score, and the Conv-head confidence score, respectively. With the D-head module, the performance of localization and recognition could be raised in the meanwhile.

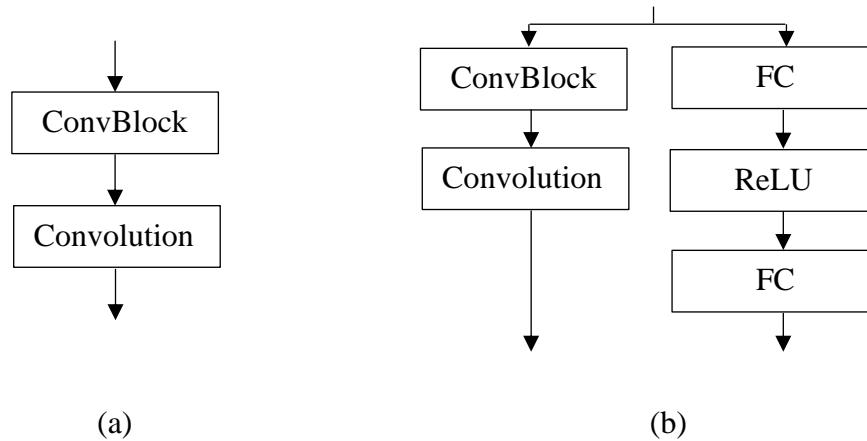


Fig. 3-7 The structures of the (a) Conv-head module and the (b) D-head module.

Second, our proposed detection model uses cross stage parallel branch splitting to address the information vanishing problem. The notion of cross stage parallel branch splitting is that an additional route which is the branch of the outputs of the first two SE-CSP stage blocks in the backbone unit is outstretched to deliver more primitive information to a D-head module in the head unit. As shown in Fig. 3-8, the output route from the second SE-CSP stage block in the backbone unit obtains cross stage attention information by two SE modules and passes a CSP stage block, an RFB module, and a ConvBlock. The cross stage parallel branch mechanism has neither part nor lot in the feature fusion to conquer information vanishing. With the cross stage parallel branch, the nodule detection performance, smaller nodules especially, will be advanced.

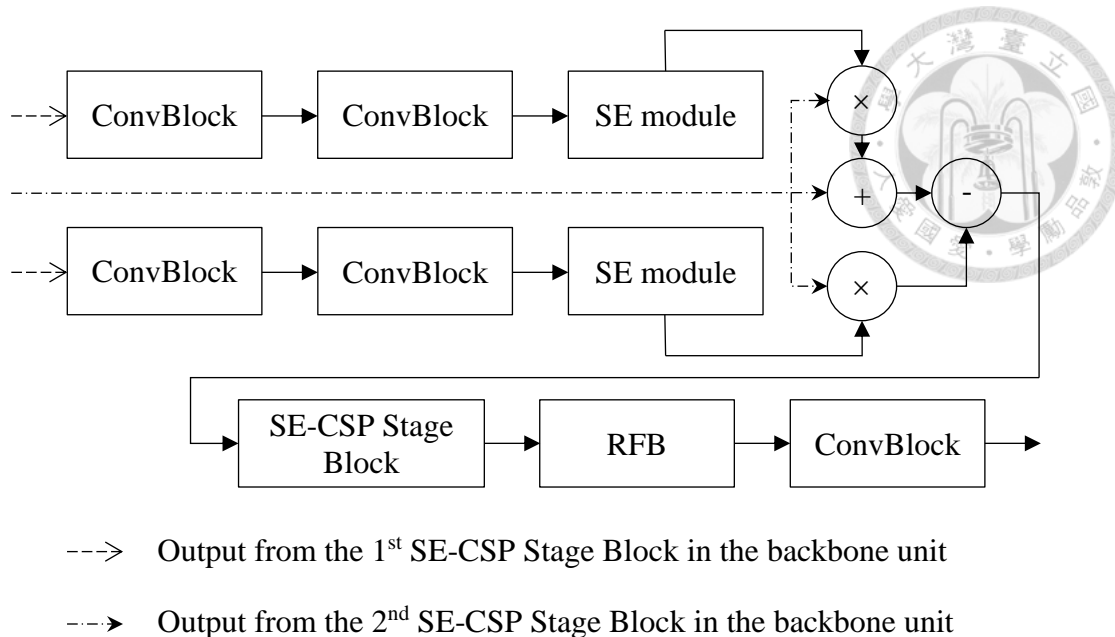


Fig. 3-8 The architecture of the cross stage parallel branch.

3.3 Initial Result Post-processing

The initial results with confidence scores, coordinates, and diameters are generated after the lung nodule detection. However, these results may contain duplicate predictions on the same nodule, overlapping with each other. The non-maximum suppression (NMS) method [24] is used to combine them to deal with this problem. Suppose the intersection over union (IoU) of any two predicted bounding boxes is greater than 0.1. In that case, the NMS will only preserve the one with the greater confidence score until no IoU conforms to the criterion. The flowchart of the initial result post-processing is illustrated in Fig. 3-9.

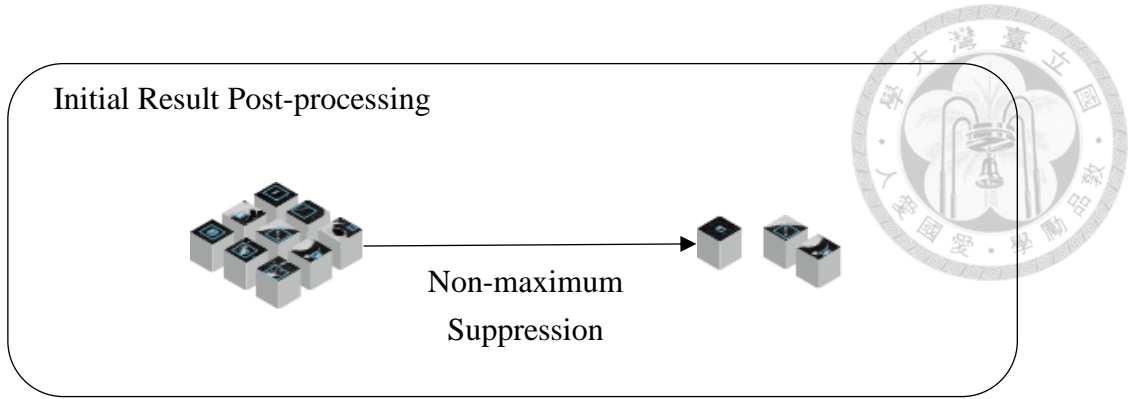


Fig. 3-9 The flowchart of the initial result post-processing.

3.4 Loss Function

The loss function is a critical criterion for training a model. Therefore, in this study, a loss function which is the sum of the two losses calculated from Conv-head and FC-head is proposed to train out the 3-D HD-YOLOv4 detection model. The proposed loss function is defined as:

$$L_{\text{Total}} = L_{\text{FC}} + L_{\text{Conv}} \quad (2)$$

which L_{Total} , L_{FC} , and L_{Conv} are the losses of the total loss, FC-head, and Conv-head, respectively. The loss of each head is composed of a classification term for confidence score prediction and a regression term for bounding box localization. The definitions of the L_{FC} and L_{Conv} are in the following:

$$L_{\text{FC}} = \lambda_{\text{FC}} \times L_{\text{FC}}^{\text{Cls}} + (1 - \lambda_{\text{FC}}) \times L_{\text{FC}}^{\text{Reg}} \quad (3)$$

$$L_{\text{Conv}} = (1 - \lambda_{\text{Conv}}) \times L_{\text{Conv}}^{\text{Cls}} + \lambda_{\text{Conv}} \times L_{\text{Conv}}^{\text{Reg}} \quad (4)$$

where L_{FC}^{Cls} and L_{FC}^{Reg} are the classification loss and the regression loss in L_{FC} , and L_{Conv}^{Cls} and L_{Conv}^{Reg} are the classification term and the regression term in L_{Conv} . λ_{FC} and λ_{Conv} are the weights that control the balance between two different terms in each loss, and they are set to 0.7 and 0.8, respectively. Moreover, the binary cross entropy loss is used to compute the classification loss, and the Smooth-L1 loss [25] is employed to calculate the regression loss.

3.5 Model Training

Before training the proposed detection model, two issues are addressed to avoid biased training and ensure the effectiveness of the detection model. One is the distribution imbalance in the nodule size and sample, and the other is the overfitting problem [26]. Firstly, the distribution imbalance problems are overcome by increasing the large nodules whose diameters are greater than 10 mm, and applying the hard negative mining [27] method during the training process, respectively. With the larger nodules increase, the detection ability on larger targets could be enhanced. On the other hand, the number of negative samples could be controlled by applying the hard negative mining method. The model could focus more on the hard negative samples rather than the easier ones. With the hard negative mining method, the false positives (FPs) are reduced significantly. Secondly, the overfitting problem is conquered by applying the

flipping data augmentation and the random crop [28] method during training. If the IoU of the cropped volume and the ground truth bounding box is greater than 0.5, the cropped volume is regarded as a positive sample in the random crop implementation.

On the other hand, if the IoU is less than 0.02, it will be considered a negative sample.

The other cases that the IoU is in the range of 0.02 and 0.5 are ignored. Furthermore, during the model training, the batch size is set to 8, the training epoch is set to 400, and the optimizer, AdaBelief [29], with weight decay 0.0001. The initial learning rate of the optimizer is set to 0.001, updated by the cosine annealing mechanism [30], and will be decreased by a factor of 10 for every 200 epochs.

Chapter 4 Results and Discussion



4.1 Experiment Environment

The proposed CADe system is implemented using the Pytorch 1.8.0 framework and the programming language Python 3.6.12. The entire experiment is run on a PC equipped with an Intel i7-10700 2.9GHz CPU (NVIDIA, Santa Clara, CA, USA) with 16 GB RAM and an NVIDIA GeForce RTX 3090 GPU with 24 GB memory (NVIDIA, Santa Clara, CA, USA).

4.2 Evaluation

In this study, the LUNA16 dataset is randomly divided into five equal-sized parts, and the 5-fold cross-validation [31] experiment is performed. The main criteria for evaluating the CADe system are the sensitivity and the corresponding FP per scan. The predicted nodule candidate from the CADe system is regarded as a true positive (TP) if the candidate is located within the radius of the ground truth nodule center. Otherwise, it is considered an FP. The performance of the CADe system is evaluated by calculating the competition performance metric (CPM) [32]. CPM is the average sensitivity at seven pre-defined false positives per scan: 1/8, 1/4, 1/2, 1, 2, 4, 8 on the free-response

receiver operating characteristic (FROC) curve [33]. A higher CPM score means better performance.



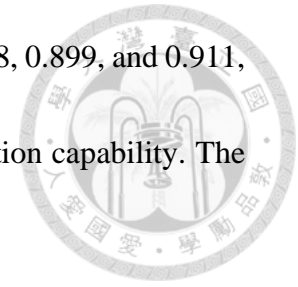
4.3 Experiment Results

In this study, two experiments are conducted to verify the proposed 3-D HD-YOLOv4 model's effectiveness. One is the ablation study of the integrated modules in the proposed model, and the other is the comparison between the proposed model and other state-of-the-art (SOTA) models. The details are described in the following sections.

4.3.1 Ablation Study

The proposed 3-D HD-YOLOv4 detection model is modified from the YOLOv4 [9] by integrating the SE module [10], the RFB module [11], the D-head module [12], and the cross stage parallel branch mechanism. To verify the enhancement of these modifications gradually, two ablation studies are executed. One is the CPM scores comparison between different methods, and the other compares the number of false negatives (FN) between them. Table 4-1 lists the results of the first ablation study. The CPM score of 3-D YOLOv4 is 0.887. Then, by applying the SE-CSP stage block, the RFB module, the D-head module, and the cross stage parallel branch mechanism to the

3-D YOLOv4 sequentially, the performance is raised to 0.896, 0.898, 0.899, and 0.911, respectively. The proposed modifications could improve the detection capability. The



FROC of different models is illustrated in Fig. 4-1.

Table 4-1 The CPM scores comparison of the different models.

Model	Param.	1/8	1/4	1/2	1	2	4	8	CPM
3-D YOLOv4	1.83 M	0.768	0.821	0.874	0.910	0.935	0.945	0.959	0.887
3-D YOLOv4+SE	2.12 M	0.777	0.829	0.877	0.927	0.947	0.957	0.959	0.896
3-D YOLOv4+SE+RFB	2.25 M	0.782	0.845	0.886	0.922	0.944	0.954	0.959	0.898
3-D YOLOv4+SE+RFB +D-head	2.26 M	0.785	0.843	0.895	0.917	0.937	0.957	0.962	0.899
3-D HD-YOLOv4	2.71 M	0.791	0.847	0.890	0.936	0.960	0.970	0.982	0.911

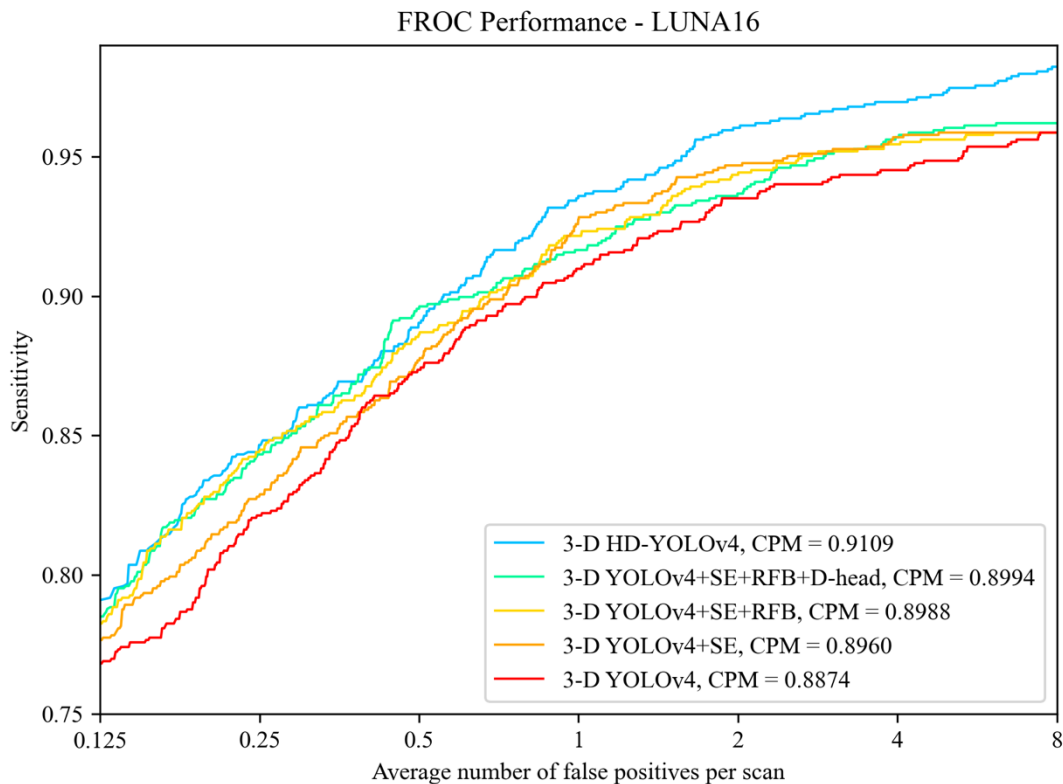


Fig. 4-1 The FROC performance comparison of different models.

Next, the comparison of the number of FNs is illustrated in Fig. 4-2. The FNs are divided into small, medium, and large groups according to the diameter d (mm). The

small, medium, and large groups are with $d \leq 5$, $5 < d \leq 10$, and $10 < d$, respectively. In comparing the small group, the number of FNs of the proposed model is eight while the other models are nearby 25. For the medium and large groups, the FNs omitted by the proposed model are 7 and 6, respectively, and those excluded by the other methods are greater or equal to 10. The number of the FNs by the proposed detection model is only 21, while those by the other models are more than 45.

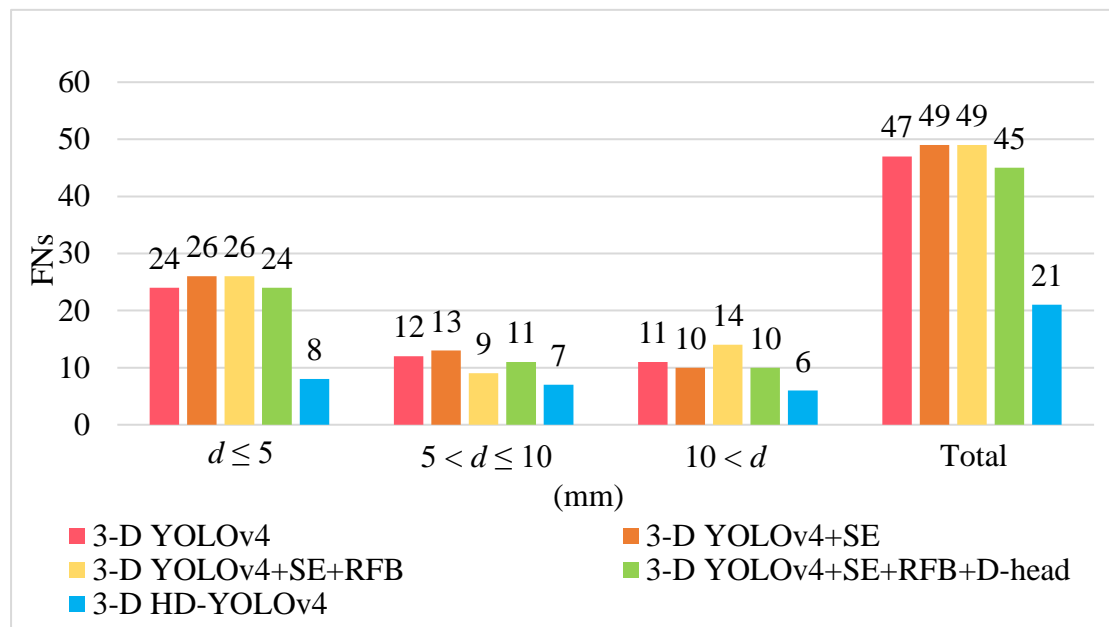
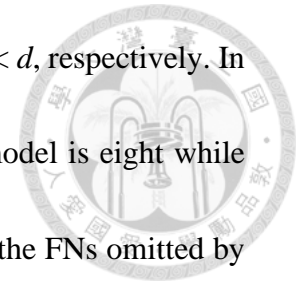


Fig. 4-2 The comparison of different models with the number of false negatives grouped by the diameter d (mm).

4.3.2 Comparison with SOTA

To analyze the performance of the proposed model, we further compare it with several SOTA methods on the LUNA16 dataset by using the CPM score shown in Table 4-2. The CPM scores obtained by the SOTA methods are in the range of 0.862 and

0.897, and that by the proposed model is 0.911, which surpasses the top two models by 0.014 and 0.020, respectively. Furthermore, the sensitivities of our detection model in FPs per scan 1/8, 2, 4, and 8 are 0.791, 0.960, 0.970, and 0.982, which are all in the first place.

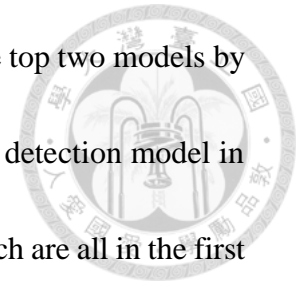


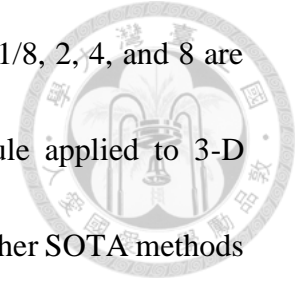
Table 4-2 The comparison with state-of-the-art methods on the LUNA16 dataset.

Methods	Param.	1/8	1/4	1/2	1	2	4	8	CPM
Li <i>et al.</i> [34]	-	0.739	0.803	0.858	0.888	0.907	0.916	0.920	0.862
Song <i>et al.</i> [35]	-	-	-	-	0.911	0.928	-	0.948	-
Mei <i>et al.</i> [36]	-	0.712	0.802	0.865	0.901	0.937	0.946	0.955	0.874
Wang <i>et al.</i> [37]	-	0.676	0.776	0.879	0.949	0.958	0.958	0.958	0.878
Ding <i>et al.</i> [38]	11.72 M	0.748	0.853	0.887	0.922	0.938	0.944	0.946	0.891
Khosravan <i>et al.</i> [39]	4.57 M	0.709	0.836	0.921	0.953	0.953	0.953	0.953	0.897
Proposed	2.71 M	0.791	0.847	0.890	0.936	0.960	0.970	0.982	0.911

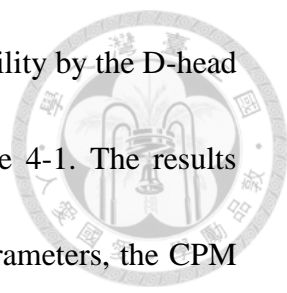
4.4 Discussion

The proposed method is integrated with four modifications in this study, including the SE module, RFB module, D-head module, and cross stage parallel branch. The first two modifications are the major factors of the predicted confidence score, and the remaining two are that of the detection capability. With the enhancements of predicted confidence score and detection capability, the CPM score boosts by 0.011 and 0.013, respectively. Furthermore, the number of FNs is significantly reduced from about 50 to 21 by adopting four modifications simultaneously, and all the sensitivities at the pre-defined FPs per scan are improved by about 0.025. With the comparison of other SOTA

methods, the sensitivities of the proposed method at FPs per scan 1/8, 2, 4, and 8 are the highest. These results show the improvement of each module applied to 3-D YOLOv4 and the outperformance of 3-D HD-YOLOv4 related to other SOTA methods on the same dataset.

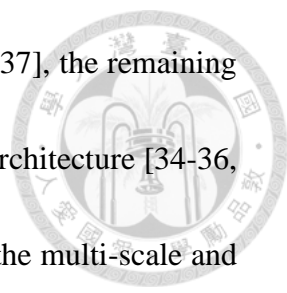


In this study, the SE and RFB modules are employed in the backbone and neck units to improve the predicted confidence score. Hence, the experiments by integrating the SE module and the RFB module are performed to verify the effectiveness. In the experiments, each module is applied to the 3-D YOLOv4 sequentially. With the attention mechanism provided by the SE module, all the sensitivities at the pre-defined FPs per scan are improved. The SE module is used to advance the quality of the feature representation by the self-gating mechanism [10]. Through the self-gating mechanism, the weight obtained from the SE module highlights the feature more representative. As shown in Fig. 4-1 and Fig. 4-2, the number of detected nodules decreases while the FROC performance rises. This result indicates that the detected nodules have higher confidence scores, and the attention mechanism suppresses the number of FPs. Next, the integration of the RFB module endues the feature larger receptive field size, and this modification slightly improves the CPM score and the sensitivities at FPs per scan less or equal to 1/2.



Additionally, the effectiveness of enhancement of detection ability by the D-head and the cross stage parallel branch is verified and listed in Table 4-1. The results indicate that with the D-head module and the slight increase in parameters, the CPM score is boosted by 0.001, and the sensitivity at 8 FPs per scan is increased to 0.962. The D-head module is constructed by integrating an FC-head module with a Conv-head module. The Conv-head is a local operation that can only obtain information from a restricted kernel size, while the FC-head constructed by fully connected architectures is without this constrain. Thus, the D-head module can make more exhaustive usage of the feature while keeping the localization ability of Conv-head. Furthermore, by performing the cross stage parallel branch, the proposed 3-D HD-YOLOv4 improves all the sensitivities at the FPs per scan \geq one and detects more nodules, especially small ones. The cross stage parallel branch delivers the more primitive information to the head unit to avoid losing information while performing more dimension reduction operations. The increasing number of detected nodules shows that the cross stage parallel branch can significantly retain the subtle information that easily vanishes during the process.

Next, the proposed system is compared with other SOTA methods on the same dataset. The results exhibited the Table 4-2 present that the performance of our proposed model outperforms that of SOTA methods. In the SOTA methods, excluding



the multi-scale with the two-stage model proposed by Wang *et al.* [37], the remaining detection models are designed with the single-scale and one-stage architecture [34-36, 39]. Compared to the single-scale and one-stage detection models, the multi-scale and two-stage models could achieve higher accuracy but have more detection time [40]. To pursue a faster algorithm and higher accuracy, the S4ND [39] proposed by Khosravan *et al.* adopted the dense block module [41] to extremely extract the feature. In contrast, this dense architecture made a significant increase in parameters. Relative to the dense block system, our model surpasses its CPM score by 0.014 and only uses about half the parameters of their proposed model. The proposed 3-D HD-YOLOv4, which adopts multi-scale detection, one-stage procedure, and lightweight architecture, outperforms the other SOTA methods.

Observations from the CT scan with the proposed system show that the proposed 3-D HD-YOLOv4 performs well in general cases. The detected cases from the small group, medium group, and large group are presented in Fig. 4-3. As illustrated in Fig. 4-3(a) and Fig. 4-3(b), the small nodule and the medium nodule with size 3.83 mm and 6.27 mm are both well detected, and the large nodule with size 25.87 mm shown in Fig. 4-3(c) is also observed as expected. Our method can adaptively choose the most proper size anchor to detect nodules with different sizes with multi-scale detection.

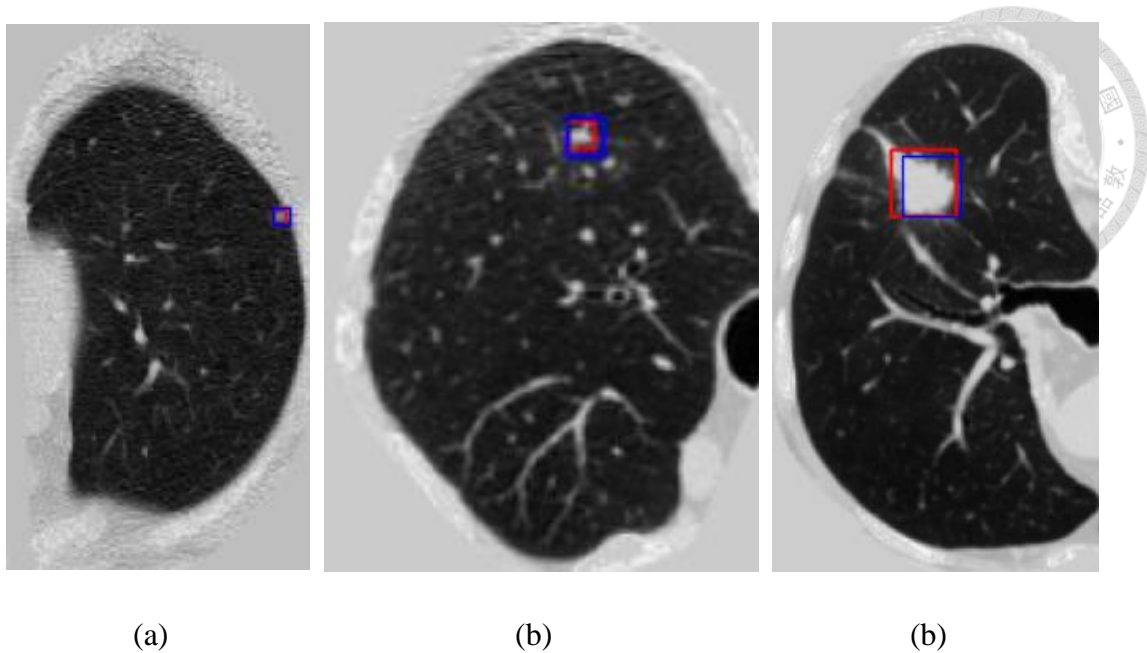


Fig. 4-3 The detected cases: (a) small nodule with size 3.83 mm, (b) medium nodule with size 6.27 mm surrounded by lung tissues, and (c) large nodule with size 25.87 mm. (The ground truth is the bounding box in red, and the predicted bounding box is in blue.)

However, the proposed method still has three shortcomings. First, despite the lightweight framework and the remarkable detection capability of the proposed method, the memory demand during the training process is enormous. This result might be attributed to the architecture design without optimization. Second, compared to the 3-D YOLOv4, the proposed model spends more time in the model training and inference process due to the extra gradient calculation of the integrated modules and the additional feature map prediction from the cross stage parallel branch mechanism. Third, there are still some hard cases for our model to detect. As presented in Fig. 4-4(a) and Fig. 4-4(b), the suspicious region near the trachea and the unclear edge small nodule with size 4.70 mm are treated as FP and FN, respectively. These misjudgments may

occur because of their misleading pathological features. In future works, some model compression techniques, such as network pruning [42], knowledge distillation [43], and network architecture search (NAS) [44], could be applied to reduce the redundant settings of the proposed model to deal with both memory demand and time-consuming issues. To improve the detection of hard cases, novel data augmentation and pre-train techniques are worth employing.

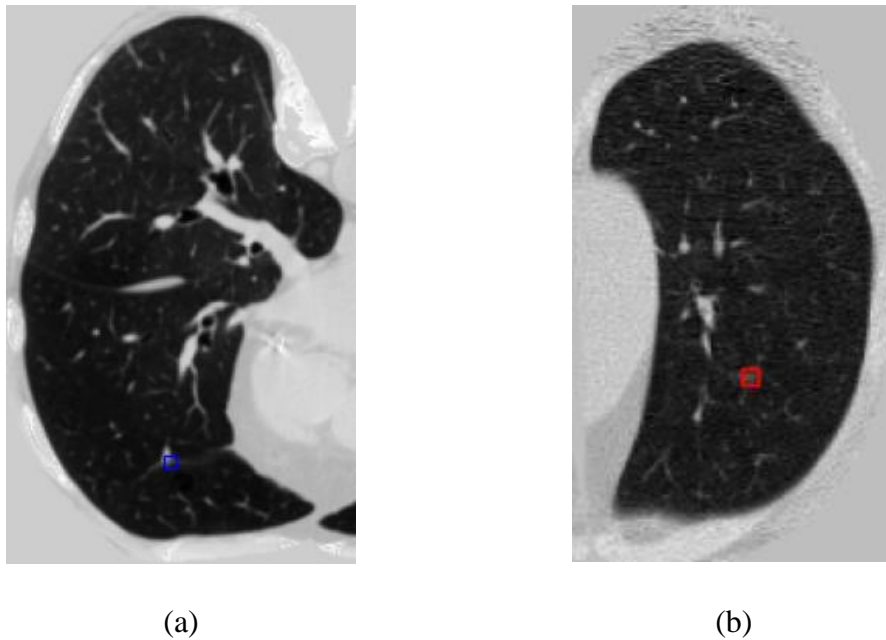
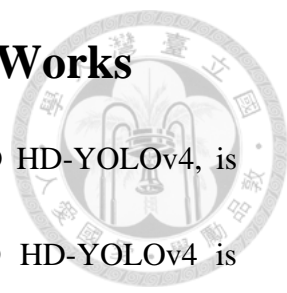


Fig. 4-4 The misjudgment detections: (a) suspicious region near the trachea and (b) unclear edge small nodule with size 4.70 mm. (The ground truth is the bounding box in red, and the predicted bounding box is in blue.)

Chapter 5 Conclusions and Future Works



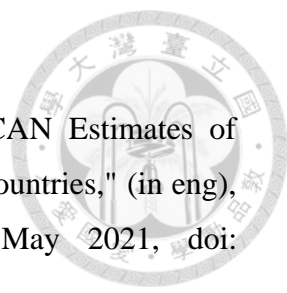
In this study, an end-to-end CNN-based CADe system, 3-D HD-YOLOv4, is developed for pulmonary nodule detection in LDCT. The 3-D HD-YOLOv4 is modified from the YOLOv4 by integrating the SE module, RFB module, D-head module, and cross stage parallel branch mechanism to enhance the nodule detection ability. The performance of the proposed CADe system is evaluated on the LUNA16 dataset and compared with the current SOTA methods. The result shows that our proposed system obtains a 0.911 CPM score which surpasses all the other SOTA methods more than 0.014. For the sensitivities at 1/8, 2, 4, and 8 FPs per scan, the performances of the proposed detection system are the highest than other methods. This result indicates that our approach with fewer parameters has lower FPs and higher sensitivities.

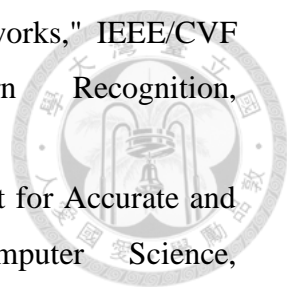
However, there is still room for the improvement of the CADe system. First, the framework of the CNN model is without optimization, which leads to a great amount of memory demand and time consumption during model training. Second, there are still some hard cases for our system to detect. Therefore, some network compression techniques, network pruning, knowledge distillation, NAS, etc., could be adopted to decrease the memory and time demand during the process in future works. Furthermore, experimenting with other methods, such as self-supervised pre-training with other

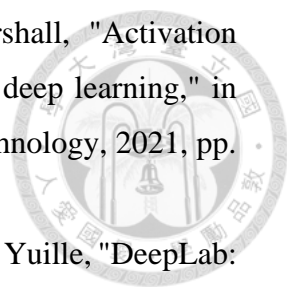
datasets, backbone switching, or different hyperparameter settings, might be necessary to improve the detection capability of the proposed system.

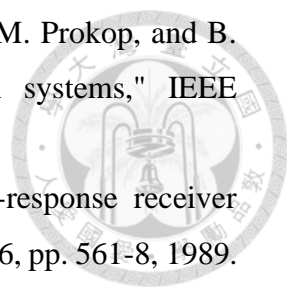


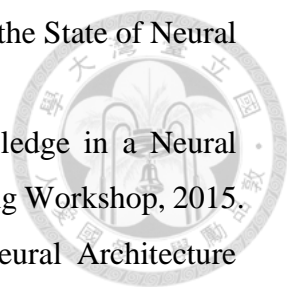
Reference

- 
- [1] H. Sung et al., "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," (in eng), *CA Cancer J Clin*, vol. 71, no. 3, pp. 209-249, May 2021, doi: 10.3322/caac.21660.
- [2] T. National Lung Screening Trial Research et al., "Reduced lung-cancer mortality with low-dose computed tomographic screening," *N Engl J Med*, vol. 365, no. 5, pp. 395-409, Aug 4 2011, doi: 10.1056/NEJMoa1102873.
- [3] T. National Lung Screening Trial Research, "Lung Cancer Incidence and Mortality with Extended Follow-up in the National Lung Screening Trial," (in eng), *J Thorac Oncol*, vol. 14, no. 10, pp. 1732-1742, Oct 2019, doi: 10.1016/j.jtho.2019.05.044.
- [4] U. Pastorino et al., "Prolonged lung cancer screening reduced 10-year mortality in the MILD trial: new confirmation of lung cancer screening efficacy," (in eng), *Ann Oncol*, vol. 30, no. 7, pp. 1162-1169, Jul 1 2019, doi: 10.1093/annonc/mdz117.
- [5] H. J. de Koning et al., "Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial," (in eng), *N Engl J Med*, vol. 382, no. 6, pp. 503-513, Feb 6 2020, doi: 10.1056/NEJMoa1911793.
- [6] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep Learning for Computer Vision: A Brief Review," (in eng), *Comput Intell Neurosci*, vol. 2018, p. 7068349, 2018, doi: 10.1155/2018/7068349.
- [7] L. Cai, J. Gao, and D. Zhao, "A review of the application of deep learning in medical image classification and segmentation," *Annals of Translational Medicine*, vol. 8, no. 11, 2020.
- [8] S. K. Zhou et al., "A Review of Deep Learning in Medical Imaging: Imaging Traits, Technology Trends, Case Studies With Progress Highlights, and Future Promises," *Proceedings of the IEEE*, <https://doi.org/10.1109/jproc.2021.3054390> vol. 109, no. 5, pp. 820–838-820–838, 2021.
- [9] C.-Y. W. Alexey Bochkovskiy, Hong-Yuan Mark Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *Computer Vision and Pattern Recognition*, 2020.

- 
- [10] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," IEEE/CVF Conference on Computer Vision and Pattern Recognition, <https://doi.org/10.1109/cvpr.2018.00745> 2018.
- [11] S. Liu, D. Huang, and Y. Wang, "Receptive Field Block Net for Accurate and Fast Object Detection," Lecture Notes in Computer Science, https://doi.org/10.1007/978-3-030-01252-6_24 pp. 404–419-404–419, 2018.
- [12] Y. Wu et al., "Rethinking Classification and Localization for Object Detection," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
- [13] 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," (in eng), *Med Image Anal*, vol. 42, pp. 1-13, Dec 2017, doi: 10.1016/j.media.2017.06.015.
- [14] S. G. Armato, 3rd et al., "The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans," *Med Phys*, vol. 38, no. 2, pp. 915-31, Feb 2011, doi: 10.1118/1.3528204.
- [15] R. C. González, R. Woods, and B. Masters, "Digital Image Processing, Third Edition," *Journal of Biomedical Optics*, vol. 14, pp. 029901-029901, 2009.
- [16] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," presented at the ICML, 2015.
- [17] D. Misra, "Mish: A Self Regularized Non-Monotonic Neural Activation Function," arXiv, 2019.
- [18] W. Luo, Y. Li, R. Urtasun, and R. Zemel, "Understanding the Effective Receptive Field in Deep Convolutional Neural Networks," in International Conference on Neural Information Processing Systems, 2016: Curran Associates Inc., pp. 4905–4913-4905–4913.
- [19] A. Vaswani et al., "Attention is All You Need," in International Conference on Neural Information Processing Systems, 2017: Curran Associates Inc., pp. 6000–6010-6000–6010.
- [20] V. Nair and G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," in ICML, J. Fürnkranz and T. Joachims, Eds., 2010: Omnipress, pp. 807-814.

- 
- [21] C. E. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation functions: comparison of trends in practice and research for deep learning," in International Conference on Computational Sciences and Technology, 2021, pp. 124 – 133-124 – 133.
- [22] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," IEEE Transactions on Pattern Analysis and Machine Intelligence, <https://doi.org/10.1109/tpami.2017.2699184> vol. 40, no. 4, pp. 834–848-834–848, 2018.
- [23] Y. Zhang and J. Schneider, "Projection Penalties: Dimension Reduction without Loss," in ICML, 2010.
- [24] J. Hosang, R. Benenson, and B. Schiele, "Learning Non-maximum Suppression," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), <https://doi.org/10.1109/cvpr.2017.685> 2017.
- [25] R. Girshick, "Fast R-CNN," IEEE International Conference on Computer Vision (ICCV), <https://doi.org/10.1109/iccv.2015.169> 2015.
- [26] X. Ying, "An Overview of Overfitting and its Solutions," Journal of Physics: Conference Series, vol. 1168, p. 022022, 2019, doi: 10.1088/1742-6596/1168/2/022022.
- [27] K. Sung and T. Poggio, "Example-based learning for view-based human face detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 1, pp. 39-51, 1998, doi: 10.1109/34.655648.
- [28] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in International Conference on Learning Representations, Y. Bengio and Y. LeCun, Eds., 2015.
- [29] J. Zhuang et al., "AdaBelief Optimizer: Adapting Stepsizes by the Belief in Observed Gradients," in Advances in Neural Information Processing Systems, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, Eds., 2020, vol. 33: Curran Associates, Inc., pp. 18795–18806-18795–18806.
- [30] I. Loshchilov and F. Hutter, "SGDR: Stochastic Gradient Descent with Warm Restarts," presented at the ICLR, 2017.
- [31] A. Sylvain and C. Alain, "A survey of cross-validation procedures for model selection," Statistics Surveys, vol. 4, no. none, pp. 40-79, 1/1 2010, doi: 10.1214/09-SS054.

- 
- [32] M. Niemeijer, M. Loog, M. D. Abramoff, M. A. Viergever, M. Prokop, and B. van Ginneken, "On combining computer-aided detection systems," *IEEE transactions on medical imaging*, vol. 30, pp. 215-23, 2011.
- [33] D. P. Chakraborty, "Maximum likelihood analysis of free-response receiver operating characteristic (FROC) data," *Medical physics*, vol. 16, pp. 561-8, 1989.
- [34] Y. Li and Y. Fan, "DeepSEED: 3D Squeeze-and-Excitation Encoder-Decoder Convolutional Neural Networks for Pulmonary Nodule Detection," in *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2020, pp. 1866-1869.
- [35] T. Song et al., "CPM-Net: A 3D Center-Points Matching Network for Pulmonary Nodule Detection in CT Scans," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Cham, A. L. Martel et al., Eds., 2020: Springer International Publishing, pp. 550-559.
- [36] J. Mei, M. M. Cheng, G. Xu, L. R. Wan, and H. Zhang, "SANet: A Slice-Aware Network for Pulmonary Nodule Detection," *IEEE Trans Pattern Anal Mach Intell*, vol. PP, Mar 9 2021, doi: 10.1109/TPAMI.2021.3065086.
- [37] B. Wang, G. Qi, S. Tang, L. Zhang, L. Deng, and Y. Zhang, "Automated Pulmonary Nodule Detection: High Sensitivity with Few Candidates," in *Medical Image Computing and Computer Assisted Intervention*, Cham, A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger, Eds., 2018: Springer International Publishing, pp. 759-767.
- [38] J. Ding, A. Li, Z. Hu, and L. Wang, "Accurate Pulmonary Nodule Detection in Computed Tomography Images Using Deep Convolutional Neural Networks," *Lecture Notes in Computer Science*, https://doi.org/10.1007/978-3-319-66179-7_64 pp. 559–567-559–567, 2017.
- [39] N. Khosravan and U. Bagci, "S4ND: Single-Shot Single-Scale Lung Nodule Detection," *Lecture Notes in Computer Science*, https://doi.org/10.1007/978-3-030-00934-2_88 pp. 794–802-794–802, 2018.
- [40] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, <https://doi.org/10.1109/cvpr.2017.106> 2017.
- [41] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, <https://doi.org/10.1109/cvpr.2017.243> 2017.

- 
- [42] D. Blalock, J. J. G. Ortiz, J. Frankle, and J. Gutttag, "What is the State of Neural Network Pruning?," in Machine Learning and Systems, 2020.
- [43] G. Hinton, O. Vinyals, and J. Dean, "Distilling the Knowledge in a Neural Network," in NIPS Deep Learning and Representation Learning Workshop, 2015.
- [44] M. Wistuba, A. Rawat, and T. Pedapati, "A Survey on Neural Architecture Search," presented at the arXiv, 2019.