**淡江大學資訊工程學系全英語碩士班**

**碩士論文**

指導教授：陳啓禎

基於YOLO架構和深度學習的胸腔CT影像肺癌檢測與分類研究

Research on Lung Cancer Detection and Classification in Chest CT Images based on YOLO Architecture and Deep Learning

研究生：吳昶漢 撰

Graduate Student: Angus Wu

**中華民國  114年  6月**

**June 2025**

|  |
| --- |
| 論文名稱：基於YOLO架構和深度學習的胸腔CT影像肺癌檢測與分類研究 頁數：56頁  校系(所)組別：淡江大學資訊工程學系全英語碩士班  畢業時間及提要別：113學年度第2學期碩士學位論文提要  研究生：吳昶漢 指導教授：陳啓禎 博士  論文提要內容：  肺癌仍是全球癌症相關死亡的主要原因之一，早期偵測對於改善病患預後至關重要。本研究探討YOLO（You Only Look Once）架構與深度學習技術於胸腔電腦斷層影像中肺部結節之檢測與分類應用。利用由多位放射科醫師註解的LIDC-IDRI資料集，我們提出一種改良型的YOLOv11模型，結合多重膨脹注意力機制與重參數化骨幹模組（RepC3、C2PSA），以提升特徵提取與結節定位能力。  本研究方法著重於將膨脹卷積與注意力模組整合至YOLOv11中，以捕捉多尺度空間特徵，針對小結節與模糊案例等挑戰進行優化。模型透過平均準確率（mAP）、精確度（precision）、召回率（recall）與F1分數等指標進行評估。結果顯示，所提出之YOLO模型在mAP@0.5達到81.34%，F1分數為79.12%，優於YOLOv8與原始YOLOv11模型。此模型在結節分類（良性、不明與惡性）方面亦表現出色，具備降低誤判與漏診風險的潛力。  本研究強調人工智慧系統在強化肺癌篩檢上的應用潛力，為臨床工作流程提供具擴展性之解決方案。未來方向包括進一步探討三維體積分析、整合臨床資料，以及在資源有限環境下的實地部署，以提升診斷準確性與普及性。  關鍵字：肺部腫瘤偵測、深度學習、電腦斷層影像、肺節結、注意力機制。 |

＊依本校個人資料管理規範，本表單各項個人資料僅作為業務處理使用，並於保存期限屆滿後，逕行銷毀。 表單編號：ATRX-Q03-001-FM030-03

|  |
| --- |
| Title: Research on Lung Cancer Detection and Classification in Chest CT Total pages: 56  Images based on YOLO Architecture and Deep Learning  Name of Institute: Master's Program, Department of Computer Science and Information Engineering, Tamkang University (English-Taught Program)  Graduation date: June, 2025 Degree conferred: Master degree  Name of Student: Angus Wu Supervisor: Dr. Chii-Jen Chen  吳昶漢 陳啓禎  Abstract:  Lung cancer remains a leading cause of cancer-related deaths globally, with early detection being critical for improving patient outcomes. This study explores the application of the YOLO (You Only Look Once) architecture and deep learning techniques for the detection and classification of pulmonary nodules in chest CT images. Leveraging the LIDC-IDRI dataset, which includes annotated CT scans reviewed by multiple radiologists, we propose an enhanced YOLOv11 model incorporating multi-dilation attention mechanisms and reparameterized backbone blocks (RepC3, C2PSA) to improve feature extraction and nodule localization.  Our methodology focuses on augmenting YOLOv11 with dilated convolutions and attention modules to capture multi-scale spatial features, addressing challenges such as small nodule detection and ambiguous cases. The model was evaluated using metrics including mean Average Precision (mAP), precision, recall, and F1-score. Results demonstrate that the proposed YOLO model achieves superior performance, with an mAP@0.5 of 81.34% and a balanced F1-score of 79.12%, outperforming baseline YOLOv8 and YOLOv11 architectures. The model excels in classifying nodules into benign, ambiguous, and malignant categories, offering a robust tool for reducing false positives and missed diagnoses.  This research highlights the potential of AI-driven systems to enhance lung cancer screening, providing a scalable solution for clinical workflows. Future work may explore 3D volumetric analysis, integration of clinical metadata, and deployment in resource-limited settings to further advance diagnostic accuracy and accessibility. |
| Keywords: Lung cancer detection, Deep Learning, CT imaging, pulmonary nodules, attention mechanism. |

According to “TKU Personal Information Management Policy Declaration“, the personal information collected on this form is limited to this application only. This form will be destroyed directly over the deadline of reservations. 表單編號：ATRX-Q03-001-FM031-03

**Table of Contents**

[**Chapter 1 Introduction** 1](#_Toc204004137)

[1.1 Background and Significance 1](#_Toc204004138)

[1.2 Introduction of YOLO 2](#_Toc204004139)

[1.3 Thesis Overview 4](#_Toc204004140)

[**Chapter 2: Related Works** 5](#_Toc204004141)

[2.1 Lung Nodules: Definition, Causes, and Clinical Significance 5](#_Toc204004142)

[2.2 Computed Tomography (CT) Imaging for Pulmonary Diagnosis 6](#_Toc204004143)

[2.3 YOLO-Based Detection of Pulmonary Nodules 6](#_Toc204004144)

[2.4 Lung Nodule Detection and Classification Using YOLO 7](#_Toc204004145)

[2.5 Lung Nodule Classification Categories 7](#_Toc204004146)

[**Chapter 3: Research Methodology** 13](#_Toc204004147)

[3.1 Overview of YOLOv11 Architecture 13](#_Toc204004148)

[3.2 Proposed Methodology for Pulmonary Nodule Detection 21](#_Toc204004149)

[**Chapter 4: Experimental Results and Discussions** 29](#_Toc204004150)

[4.1 Dataset and Preprocessing 29](#_Toc204004151)

[4.2 Training Configuration and Evaluation Metrics 32](#_Toc204004152)

[4.3 Results and Discussions 35](#_Toc204004153)

[4.3.1 Original YOLO Experiment Results 35](#_Toc204004154)

[4.3.2 Ablation Study Results 46](#_Toc204004155)

[4.3.3 Architectural Comparison: YOLOv11, YOLOv12, and the Proposed Model 49](#_Toc204004156)

[**Chapter 5: Conclusion and Future Works** 51](#_Toc204004157)

[5.1 Conclusion 51](#_Toc204004158)

[5.2 Future Works 52](#_Toc204004159)

[**References** 54](#_Toc204004160)

**List of Figures**

[Figure 1. Healthy Lung CT Image. 9](#_Toc204004209)

[Figure 2. Benign Lung Nodule CT Image. 10](#_Toc204004210)

[Figure 3. Ambiguous Lung Nodule CT Image. 10](#_Toc204004211)

[Figure 4. Malignant Lung Nodule CT Image. 11](#_Toc204004212)

[Figure 5. YOLOv11 Backbone Architecture Diagram. 16](#_Toc204004213)

[Figure 6. YOLOv11 Neck Architecture Diagram. 17](#_Toc204004214)

[Figure 7. C3K2 Module Diagram. 18](#_Toc204004215)

[Figure 8. C2PSA Module Diagram. 19](#_Toc204004216)

[Figure 9. YOLOv11 Architecture. 20](#_Toc204004217)

[Figure 10. MSDA Module. 25](#_Toc204004218)

[Figure 11. RepC3 Module. n indicates the number of times it will run RepConv. 26](#_Toc204004219)

[Figure 12. RepC3 + MSDA Module. 26](#_Toc204004220)

[Figure 13. The Proposed YOLO Attention Framework Neck. 27](#_Toc204004221)

[Figure 14. The Proposed YOLO Attention Framework. 28](#_Toc204004222)

[Figure 15. YOLOv11 Metrics and Training/Validation Loss Graphs. YOLOv11 achieves slightly higher early precision, the proposed architecture offers a more balanced and robust performance suitable for clinical deployment. 38](#_Toc204004223)

[Figure 16. Proposed Architecture Metrics and Training/Validation Loss Graphs. The proposed model demonstrates smoother and faster convergence across all loss curves, higher recall, and improved mAP scores—particularly mAP@0.5:0.95—indicating enhanced ability to detect small or ambiguous nodules. 39](#_Toc204004224)

[Figure 17. YOLOv11 Normalized Confusion Matrix. 40](#_Toc204004225)

[Figure 18. Proposed Architecture Normalized Confusion Matrix. 41](#_Toc204004226)

[Figure 19. YOLOv11 Precision-Recall Curve. 42](#_Toc204004227)

[Figure 20. Proposed Architecture Precision-Recall Curve. 43](#_Toc204004228)

[Figure 21. Validation label visualization for a batch of CT images. Ground truth bounding boxes are annotated with class indices (0: benign, 1: ambiguous, 2: malignant) and corresponding malignancy scores. 44](#_Toc204004229)

[Figure 22. Prediction output of the proposed YOLO model on the same validation batch. The bounding boxes display predicted classes and confidence scores, demonstrating accurate localization and malignancy classification of pulmonary nodules. 45](#_Toc204004230)

**List of Tables**

[Table 1. Comparison of Healthy Lung with Different Nodules. 12](#_Toc204004243)

[Table 2. Dataset Malignancy Interpretation. 31](#_Toc204004244)

[Table 3. YOLO Training Parameters. 34](#_Toc204004245)

[Table 4. Hardware Specification. 35](#_Toc204004246)

[Table 5. Ablation study comparing YOLOv8 and YOLOv11 with and without RepC3 and MSDA modules. 48](#_Toc204004247)

[Table 6. Comparison Results of Different YOLO versions and Proposed version. 50](#_Toc204004248)

# **Chapter 1 Introduction**

1.1 Background and Significance

Lung cancer remains one of the leading causes of cancer-related deaths worldwide. According to the World Health Organization (WHO), lung cancer accounts for approximately 1.8 million deaths annually, making it a critical public health concern [1]. The disease is characterized by the uncontrolled growth of abnormal cells in the lungs, which can metastasize to other organs if not detected early. The prognosis for lung cancer patients largely depends on early detection and accurate classification, which can significantly improve treatment outcomes and survival rates [2].

Lung cancer develops due to a combination of genetic, environmental, and lifestyle factors. The primary risk factor is tobacco smoking, which is responsible for approximately 85% of all lung cancer cases [3]. Secondhand smoke exposure, air pollution, occupational exposure to carcinogens (such as asbestos, radon, and arsenic), and genetic predisposition also contribute to lung cancer formation [4]. In recent years, studies have highlighted the role of genetic mutations and epigenetic alterations in lung cancer progression, further emphasizing the need for molecular-based detection and classification methods [5].

Traditionally, lung cancer diagnosis relies on a combination of imaging techniques, biopsy procedures, and pathological analysis. Chest X-rays and computed tomography (CT) scans are commonly used as preliminary screening tools, while positron emission tomography (PET) scans provide further insight into tumor metabolism and staging [6]. However, these conventional methods are often time-consuming, expensive, and prone to inter-observer variability, leading to potential misdiagnoses or delays in treatment initiation [7].

The integration of artificial intelligence (AI) into medical imaging has significantly enhanced the efficiency and accuracy of lung cancer detection and classification. AI-powered algorithms, particularly deep learning-based models, have demonstrated superior performance in identifying lung nodules, distinguishing between malignant and benign tumors, and assisting radiologists in decision-making [8]. Machine learning techniques, such as convolutional neural networks (CNNs), have been widely adopted for analyzing complex medical imaging data, leading to improved diagnostic precision and reduced false positive rates [9].

Furthermore, AI has enabled the development of computer-aided detection (CAD) systems, which provide automated interpretation of medical scans and highlight suspicious lesions for further evaluation [10]. These advancements have contributed to early lung cancer detection, ultimately improving patient survival rates by enabling timely intervention and personalized treatment strategies [11].

1.2 Introduction of YOLO

Recent advancements in AI have led to the development of sophisticated models, such as the You Only Look Once (YOLO) architecture, which enables real-time object detection with high accuracy [12]. YOLO-based models have been successfully applied in medical imaging for detecting pulmonary nodules and classifying lung cancer subtypes with enhanced efficiency compared to traditional methods [13]. Additionally, deep learning frameworks incorporating transformer networks, generative adversarial networks (GANs), and multi-modal fusion techniques have further improved the precision and robustness of lung cancer diagnostics [14]. These approaches hold great promise in early-stage lung cancer detection, ultimately contributing to improved patient survival and clinical outcomes.

YOLO (You Only Look Once) is a state-of-the-art deep learning-based object detection framework known for its speed and accuracy. Unlike traditional object detection models, which employ region proposal-based methods such as Faster R-CNN [15], YOLO processes an entire image in a single pass through the neural network. This enables real-time detection with minimal computational resources, making it highly suitable for medical imaging applications [16].

The YOLO framework has evolved through multiple versions, including YOLOv3, YOLOv4, and YOLOv5, each introducing improvements in detection accuracy and computational efficiency [17]. These models utilize convolutional neural networks (CNNs) to predict bounding boxes and class probabilities simultaneously, streamlining the detection process. Additionally, recent versions such as YOLOv7 and YOLOv8 incorporate transformer-based architectures, further enhancing feature extraction and improving small object detection, which is particularly useful for identifying pulmonary nodules in lung CT scans [18].

YOLO is particularly relevant for lung cancer detection due to its ability to quickly analyze medical images while maintaining high accuracy. Compared to traditional machine learning-based image analysis techniques, YOLO significantly reduces false positives and false negatives by leveraging end-to-end learning capabilities [19]. Moreover, its lightweight design enables deployment in resource-constrained environments, such as mobile devices and embedded systems, making it a viable solution for real-time lung cancer screening in remote and underprivileged areas.

1.3 Thesis Overview

This research explores the application of YOLO architecture and deep learning techniques in lung cancer detection and classification using chest CT images. The study aims to assess the effectiveness of AI-driven models in improving diagnostic accuracy and reducing false positives and false negatives. The findings of this research will contribute to the ongoing advancements in AI-powered medical imaging and highlight the potential of deep learning in the early diagnosis of lung cancer.

# **Chapter 2: Related Works**

This chapter provides an overview of previous studies and established knowledge relevant to lung nodule detection and classification using CT imaging and deep learning methods. It begins by outlining the medical context and characteristics of pulmonary nodules, followed by a discussion of CT imaging techniques used in diagnosis. The chapter then explores various applications of the YOLO architecture in medical imaging, particularly for lung cancer detection, and summarizes classification strategies employed in related research. Understanding this background lays the foundation for the proposed approach in the following chapters.

2.1 Lung Nodules: Definition, Causes, and Clinical Significance

Pulmonary nodules, commonly referred to as lung nodules, are small, round or oval-shaped growths that form in the lung parenchyma. Typically, these nodules are less than 3 centimeters in diameter and are discovered incidentally during imaging examinations conducted for other reasons. While many lung nodules are benign and pose no immediate threat, others may represent early stages of lung cancer, necessitating further clinical investigation.

Lung nodules can originate from a variety of benign or malignant conditions. Common causes include infections, inflammation, benign tumors such as hamartomas, and malignancies including primary lung cancer or metastases [20]. The majority of nodules are benign, particularly in younger patients or non-smokers. However, nodules that are larger than 8 mm, irregular in shape, rapidly growing, or present in high-risk individuals (e.g., smokers) are more likely to be malignant [21]. In such cases, PET scans or biopsies are often warranted.

2.2 Computed Tomography (CT) Imaging for Pulmonary Diagnosis

Computed Tomography (CT) is an advanced imaging modality that utilizes X-rays and computer processing to produce cross-sectional images of the body. It allows for detailed visualization of internal organs and is particularly useful for lung evaluation [22]. CT scanners employ rotating X-ray sources and detectors, using reconstruction algorithms such as filtered back projection or iterative methods to generate high-resolution images [23]. Multi-detector CT systems further enhance spatial resolution and scanning speed.

Compared to chest X-rays, CT provides significantly greater detail and can detect nodules as small as a few millimeters. Unlike MRI, which excels in soft tissue contrast, CT is preferred for lung imaging due to its speed and sensitivity to air-filled structures [24]. It is also more accessible and cost-effective than PET, making it the standard for lung nodule screening and evaluation [25].

2.3 YOLO-Based Detection of Pulmonary Nodules

Recent advancements in deep learning have led to the adoption of object detection models like YOLO (You Only Look Once) for automated lung nodule detection. YOLO is a one-stage detection framework that processes the entire image in a single neural network pass, yielding fast and accurate results [26]. The YOLOv4 model achieved high sensitivity and precision on the LIDC-IDRI dataset [27], demonstrating its robustness for medical use.

2.4 Lung Nodule Detection and Classification Using YOLO

Recent studies have highlighted the potential of YOLO-based architectures in detecting lung nodules from CT scans. The YOLO family of models, with its one-stage detection approach, offers efficient inference speeds while maintaining strong accuracy, making it well-suited for medical imaging tasks. In the context of pulmonary nodule detection, YOLO models have demonstrated the ability to localize nodules of varying sizes and densities through robust bounding box regression and confidence scoring mechanisms. These features make YOLO particularly effective for identifying early-stage abnormalities in lung scans.

Beyond detection, YOLO can be adapted for classification tasks by incorporating specialized heads to assess malignancy likelihood. When fine-tuned with annotated CT datasets, YOLO-based models can distinguish between benign and malignant nodules with competitive performance. Compared to traditional classification architectures such as ResNet and DenseNet, YOLO models often achieve faster inference times, which is beneficial for real-time clinical applications. This combination of detection and classification within a unified framework supports the growing interest in YOLO as a comprehensive tool for computer-aided lung cancer diagnosis.

2.5 Lung Nodule Classification Categories

Accurate classification of lung nodules is critical for determining appropriate clinical management and improving patient outcomes. In this study, three categories of lung nodules are considered for classification using AI-based models: benign nodules, ambiguous nodules, and malignant nodules. These classifications help in stratifying the risk and guiding decisions for further diagnostic testing or treatment.

A healthy lung serves as a baseline reference in automated detection and classification tasks. Figure 1 displays how a healthy lung should look. On a chest CT scan, the lung parenchyma appears clear and uniform, with well-defined bronchi and blood vessels. There are no nodules, masses, or abnormal opacities.

Benign nodules are non-cancerous and typically result from infections, inflammations, or other non-malignant causes such as granulomas or hamartomas. Figure 2 displays how a benign lung nodule should look like. These nodules tend to be round, well-defined, and smaller than malignant ones. They often remain stable over time and do not invade surrounding tissue.

Ambiguous nodules are indeterminate in nature and pose a diagnostic challenge. Figure 3 displays how an ambiguous lung nodule should look like. These nodules may exhibit mixed features that do not clearly align with either benign or malignant characteristics. Radiologists often categorize these as "indeterminate" and recommend additional imaging, PET scans, or biopsies.

Malignant nodules are cancerous and typically exhibit rapid growth, irregular or spiculated borders, and heterogeneous internal structure. Figure 4 displays how a malignant lung nodule should look like. They may also show signs of vascular invasion or spread to nearby lymph nodes.

Table 1 provides a structured comparison of CT scan characteristics across healthy lungs, benign nodules, ambiguous nodules, and malignant nodules. This summary outlines how each category differs in appearance, shape, and clinical suspicion level. Healthy lungs show uniform parenchyma without anomalies, while benign nodules are typically well-defined and stable. Ambiguous nodules present diagnostic uncertainty due to mixed or unclear features. Malignant nodules, by contrast, tend to have irregular, spiculated borders and exhibit growth over time, warranting immediate clinical attention. This classification is essential for training AI models to differentiate between these categories in automated screening systems.

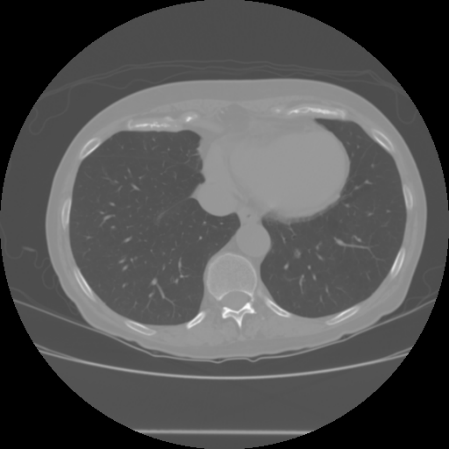


Figure 1. Healthy Lung CT Image.

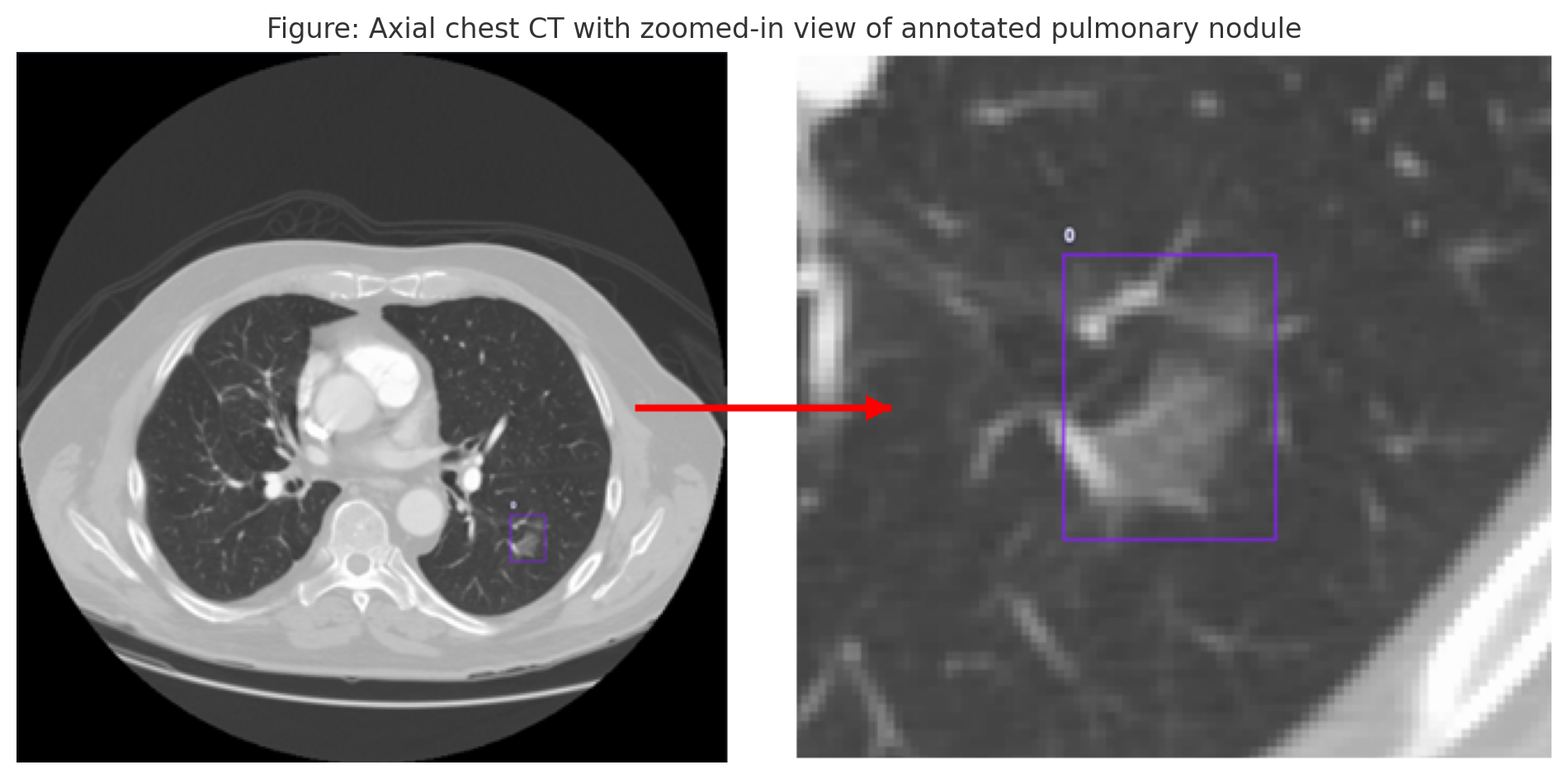


Figure 2. Benign Lung Nodule CT Image.

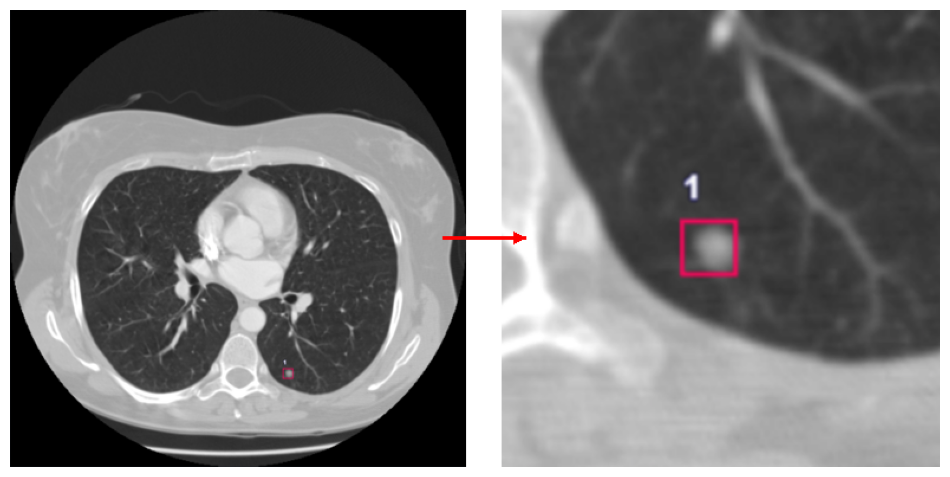


Figure 3. Ambiguous Lung Nodule CT Image.

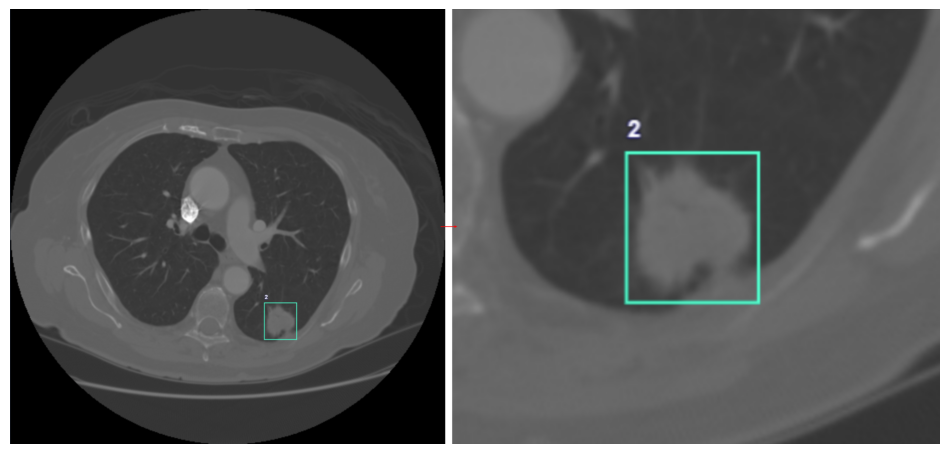


Figure 4. Malignant Lung Nodule CT Image.

Table 1. Comparison of Healthy Lung with Different Nodules.

|  |  |  |
| --- | --- | --- |
| **Classification Type** | **Example CT Image** | **Description** |
| Healthy Lung |  | Normal parenchyma with no abnormalities |
| Benign Nodule |  | Well-defined, smooth-edged nodule—typically non-cancerous |
| Ambiguous Nodule |  | Irregular features, mixed characteristics—diagnostically uncertain |
| Malignant Nodule |  | Spiculated/lobulated appearance, rapid growth—likely cancerous |

# **Chapter 3: Research Methodology**

This chapter introduces the architectural design and technical methodology adopted in this study for detecting and classifying pulmonary nodules using enhanced YOLO-based models. It first presents the core structure of the YOLOv11 architecture and its improvements over prior versions. Subsequently, the proposed enhancements—such as the integration of RepC3 blocks and the Multi-Scale Dilation Attention (MSDA) module—are detailed. These innovations aim to improve feature representation and localization accuracy, especially for ambiguous or small nodules. This methodological framework forms the backbone of the experimental implementation.

3.1 Overview of YOLOv11 Architecture

The YOLOv11 architecture is an evolution of the YOLO detection family, integrating modern enhancements in feature extraction and multi-scale representation to address limitations in previous versions. It retains the core benefits of real-time detection while offering higher accuracy, particularly for small and low-contrast objects such as pulmonary nodules in CT scans. The architecture can be described in three main components: the backbone, the neck, and the detection head.

The backbone is responsible for feature extraction. In YOLOv11, the backbone adopts a hybrid design that combines CSPDarknet elements with transformer encoder blocks. This allows it to capture both local texture features and global context from CT images. Compared to the purely convolutional backbones in earlier versions like YOLOv5 and YOLOv7, the addition of transformer encoders improves long-range dependency modeling, which is crucial for medical imaging tasks where spatial context plays a significant diagnostic role. The backbone also incorporates residual connections and multi-scale convolutional kernels to increase feature robustness. Figure 5 displays the YOLOv11 backbone architecture.

The neck of YOLOv11 is responsible for aggregating and enhancing features across multiple scales. YOLOv11 uses an improved Bidirectional Feature Pyramid Network (BiFPN), which allows for efficient feature reuse and refinement by employing weighted path fusion and repeated top-down and bottom-up pathways. Unlike the PANet used in earlier YOLO models, BiFPN simplifies the aggregation structure while improving accuracy and speed. This is particularly advantageous for detecting nodules of varying sizes in CT scans, as it ensures that semantic features from deep layers are effectively combined with the spatial precision of shallow layers. Figure 6 displays the YOLOv11 neck architecture.

The detection head in YOLOv11 uses anchor-free detection and dense prediction modules. This setup enables faster convergence and more accurate bounding box localization, especially for irregular and small nodules. YOLOv11’s head performs classification and regression simultaneously at multiple feature levels, allowing the model to detect lesions with various appearances and contrasts. Compared to anchor-based heads in YOLOv3 and YOLOv4, the anchor-free approach reduces computational overhead and false positives, making it more suitable for medical applications where precision is critical.

What makes YOLOv11 great is the addition of two structural blocks to enhance learning: C3K2 and C2PSA. The C3K2 block is an optimized version of the original C3 block used in YOLOv5. It modifies the internal residual branching by including smaller convolutional kernels (1×1 and 3×3) in a two-path setup. This lightweight structure improves efficiency while preserving feature richness. C3K2 is particularly advantageous in extracting localized patterns in CT images with fewer parameters compared to deeper C3 variants. Figure 7 displays the C3K2 module block as a whole and the breakdown of what is inside.

The C2PSA block combines a two-branch convolutional layout with a parallel self-attention mechanism. The Parallel Self-Attention (PSA) component enables the model to more effectively integrate both spatial and channel-level information across multiple layers. Positioned after the neck, the C2PSA module enhances the representational power of the fused features, thereby improving the model’s ability to distinguish ambiguous nodules from benign or malignant ones [28].

These architectural enhancements contribute to a more robust and interpretable detection pipeline, which is particularly beneficial in clinical environments where high sensitivity and precision are essential. Figure 8 illustrates the structure of the C2PSA module, including its full layout and internal components.

When all components are integrated—namely the enhanced backbone with residual and transformer-based features, the BiFPN-style neck for multi-scale fusion, and the anchor-free detection head—YOLOv11 forms a cohesive and highly optimized architecture. This unified design enables efficient end-to-end learning, robust feature representation, and accurate localization of pulmonary nodules across varying sizes and contrast levels.

Figure 9 presents the complete YOLOv11 architecture, showcasing how the backbone, neck, and detection head are systematically combined to form a single, streamlined detection pipeline capable of real-time performance and high diagnostic accuracy.

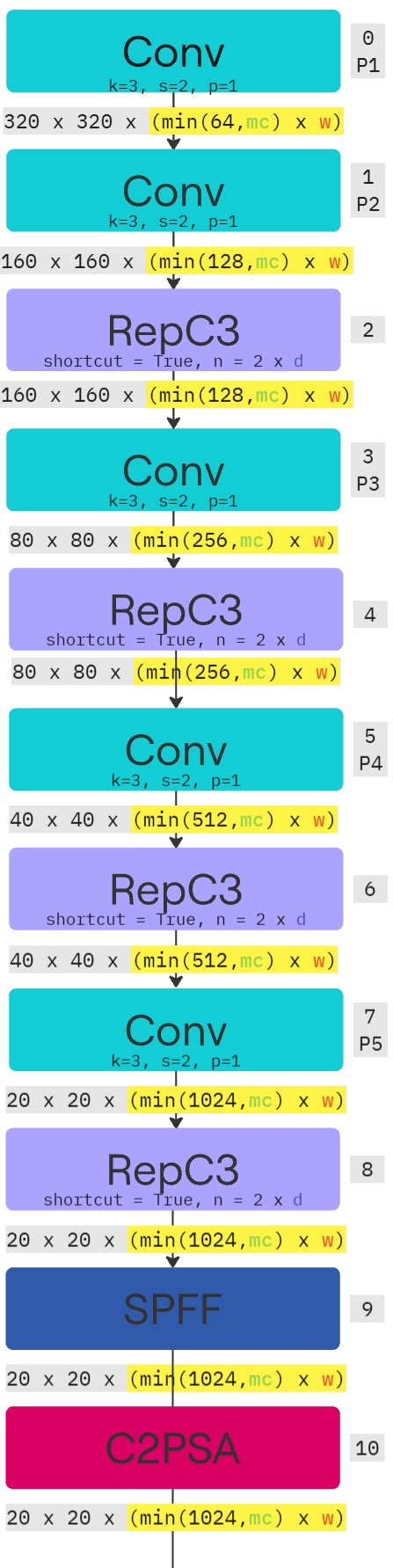


Figure 5. YOLOv11 Backbone Architecture Diagram.

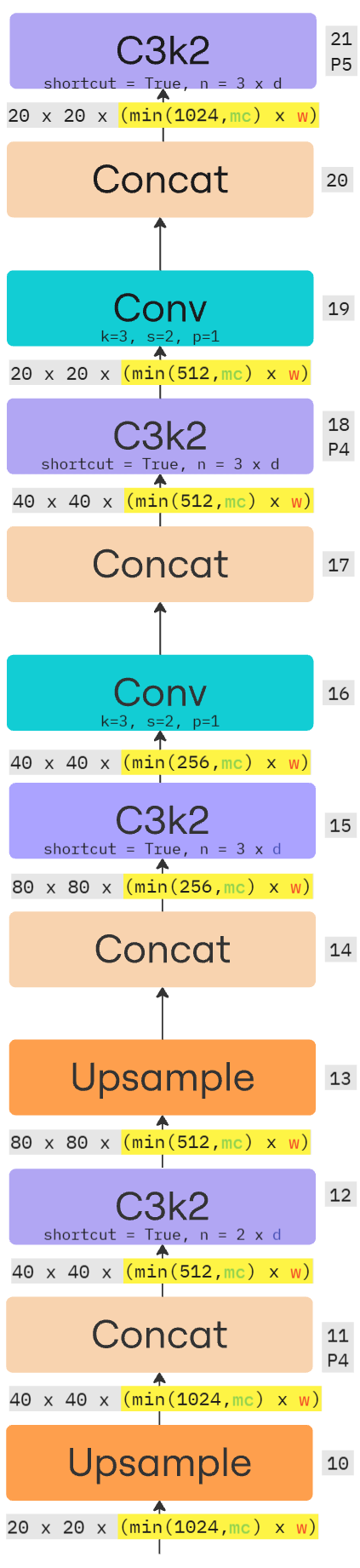


Figure 6. YOLOv11 Neck Architecture Diagram.

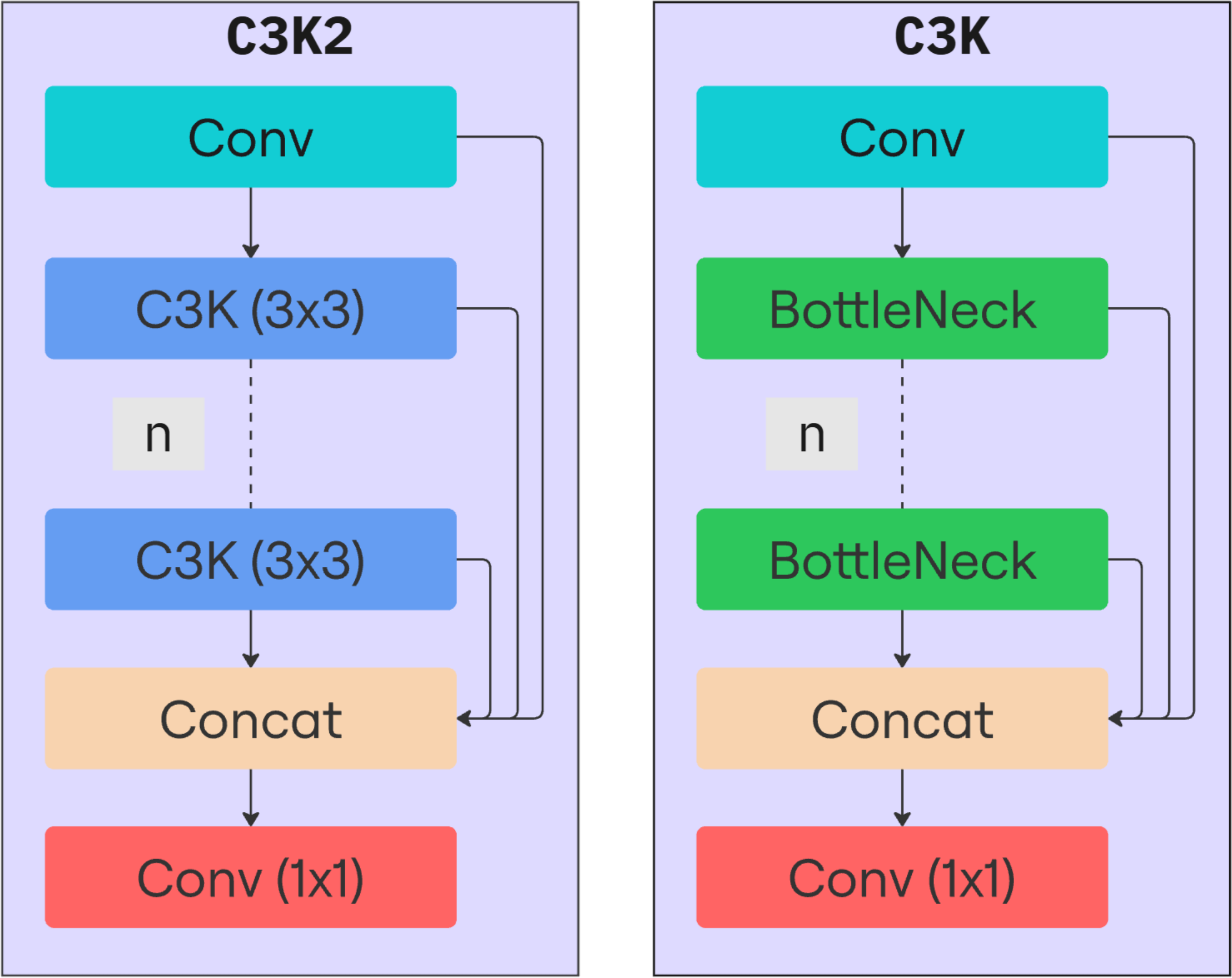


Figure 7. C3K2 Module Diagram.

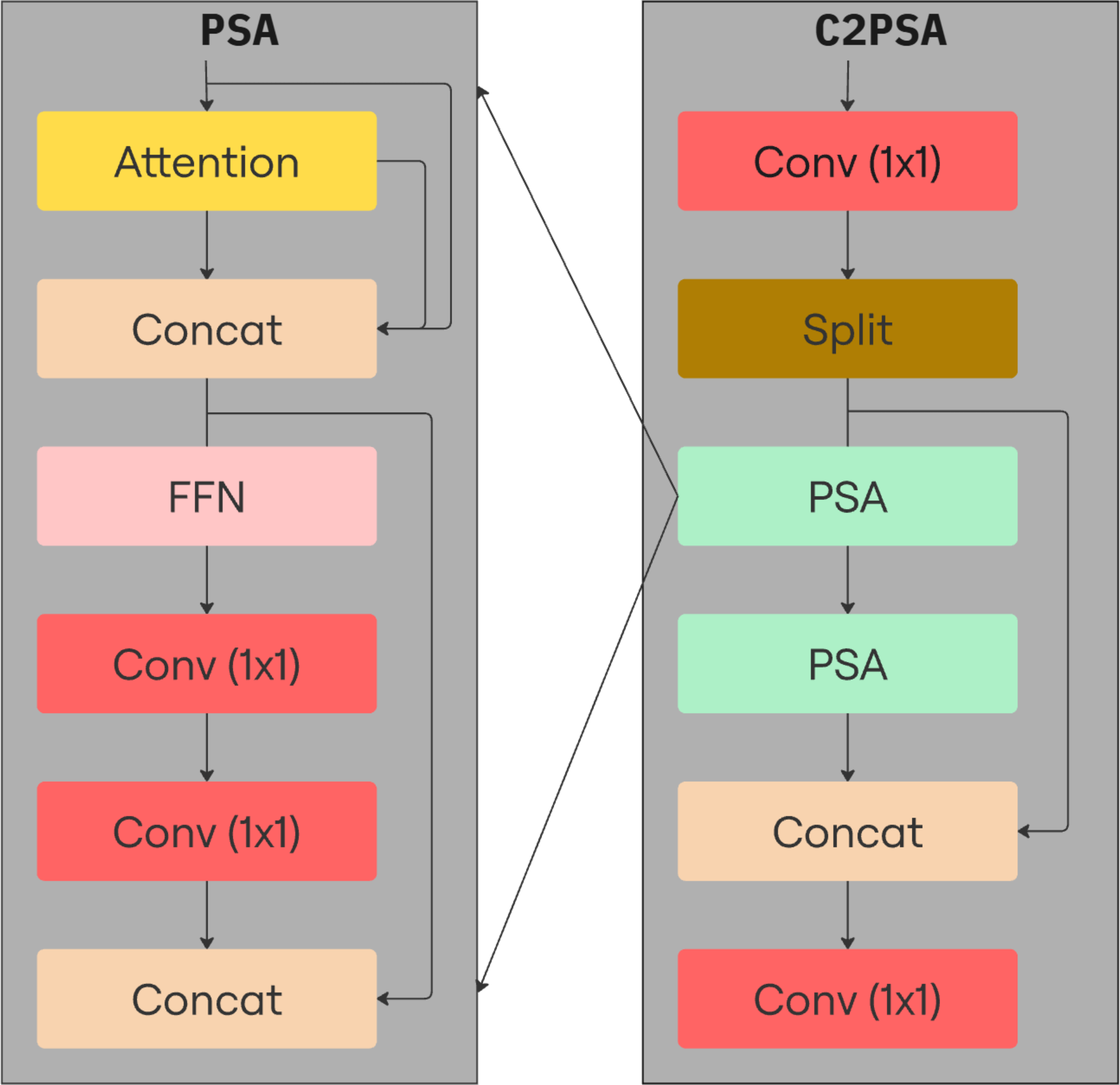


Figure 8. C2PSA Module Diagram.

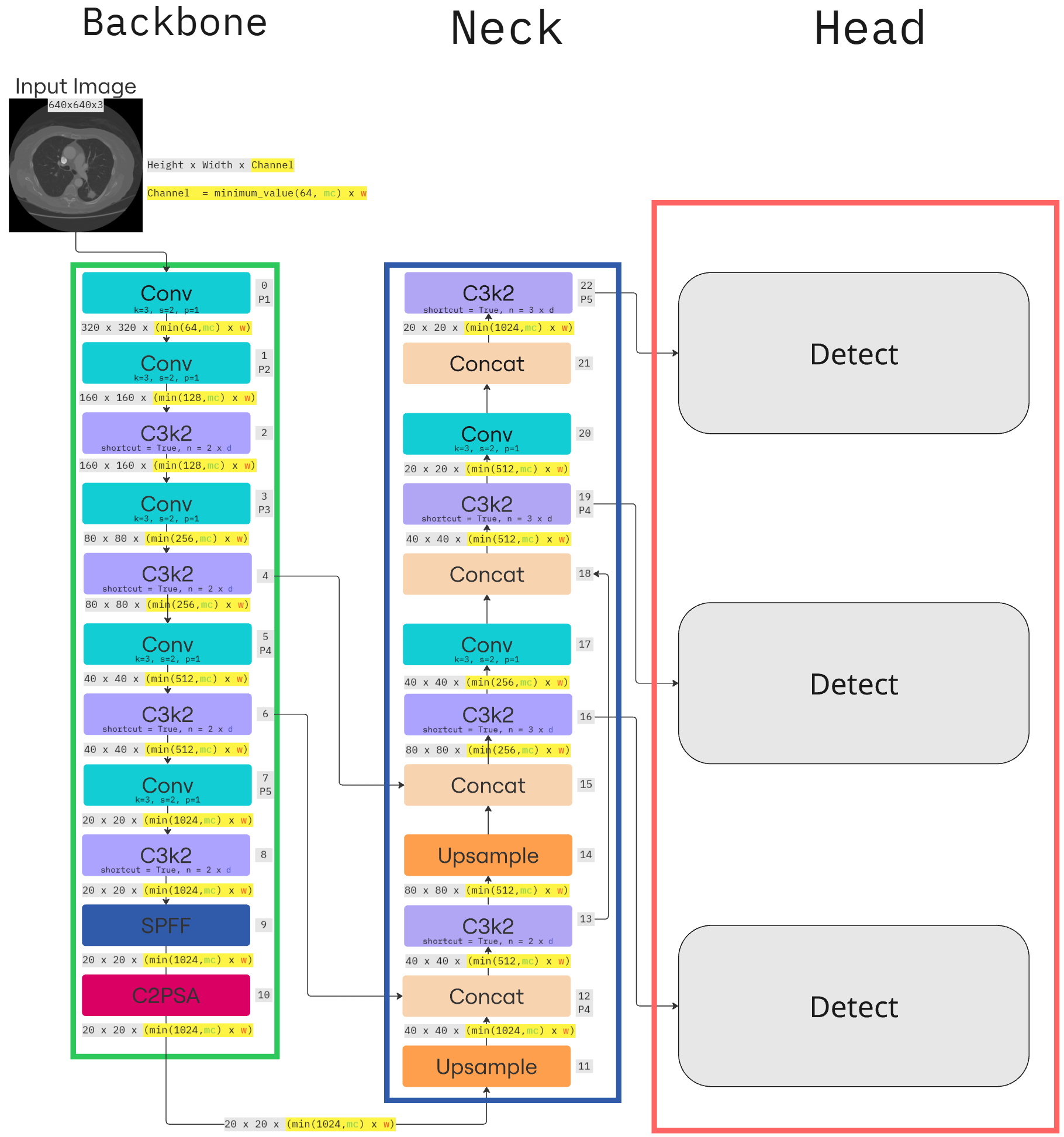


Figure 9. YOLOv11 Architecture.

3.2 Proposed Methodology for Pulmonary Nodule Detection

In this research, we propose an enhancement to the YOLOv11 architecture aimed at improving its performance for pulmonary nodule detection in chest CT images. While YOLOv11 already provides a strong foundation for real-time and accurate detection, our proposed method augments it by incorporating attention mechanisms through dilated convolution modules.

The core idea behind the enhancement is to integrate a multi-dilation attention module into the backbone and neck of the YOLOv11 framework. By introducing dilated convolutions with varying dilation rates, the model can effectively capture spatial features at multiple receptive fields. This is particularly useful in medical imaging, where lung nodules can appear in different sizes and locations with subtle variations in intensity. The dilated convolutions expand the receptive field without increasing the number of parameters, allowing the model to learn both fine-grained and global contextual features.

Within each dilation block, we embed a Convolutional Block Attention Module (CBAM) to further refine the learned features. CBAM provides both channel attention and spatial attention, enabling the network to emphasize clinically relevant regions and suppress irrelevant background noise. Channel attention helps the model focus on discriminative features across different CT scan layers, while spatial attention highlights localized anomalies such as nodules.

The MSDA module utilizes dilated convolutions to capture multi-scale contextual information by expanding the receptive field without increasing parameter count. Figure 10 gives a visualize representation of the MSDA module. The process can be described as follows:

(Eq. 1)

(Eq. 2)

(Eq. 3)

The Multi-Scale Dilation Attention (MSDA) module employs a series of mathematical operations to enhance feature extraction in medical image analysis. The first equation defines the input feature map *X* as a three-dimensional tensor with dimensions height (*H*), width (*W*), and channels (*C*), representing the spatial and depth characteristics of the CT scan data. This formulation establishes the foundational structure for subsequent operations.

The second equation describes the core dilated convolution operation, where a kernel *K* is applied to the input feature map with a specified dilation rate *r*. Unlike standard convolution, this operation introduces gaps between kernel elements, effectively expanding the receptive field to capture broader contextual information while maintaining computational efficiency. The indices (*i*, *j*) denote spatial positions in the output, while (*m*, *n*) represent kernel coordinates, and the term *r·m / r·n* implements the dilation pattern.

The third equation integrates multiple dilated convolution branches into the complete MSDA block. It combines outputs from four parallel dilation rates (*r* = 1, 2, 3, 4) through learnable weights *Wr*, followed by a non-linear activation function *σ* (typically ReLU or sigmoid). This multi-scale fusion allows the module to simultaneously detect fine details (small nodules) and broader patterns (large lesions) in lung CT scans. The attention mechanism σ further refines feature importance, making the model particularly effective for identifying pulmonary nodules of varying sizes and morphologies while maintaining parameter efficiency crucial for medical imaging applications.

Additionally, we replace the traditional C3K2 block in the backbone with the RepC3 block. RepC3 is a reparameterized version of C3 that consolidates multiple branches into a single-path architecture at inference time. This design reduces memory usage and inference latency while maintaining the expressiveness of the original block during training. RepC3 is inspired by reparameterization strategies like those used in RepVGG, offering significant efficiency gains with no compromise in accuracy. Figure 11 gives a visual representation of RepC3 and how it works.

To further enhance the performance of the proposed architecture, we integrate the RepC3 and MSDA modules in a sequential configuration. This hybrid design leverages the strengths of both reparameterization and multi-scale attention. The RepC3 module, positioned upstream, efficiently captures localized spatial patterns during training and collapses into a single-path inference structure. It serves as a compact and expressive encoder that feeds high-quality feature maps into the downstream MSDA module.

As illustrated in Figure 12, the RepC3 + MSDA block combines parameter-efficient feature transformation with dynamic, scale-aware attention. This combination enhances the sensitivity of the detection head to nodules, especially in ambiguous or borderline regions of interest, without significantly increasing computational load.

This proposed attention-enhanced YOLOv11 architecture is designed to improve detection sensitivity and classification confidence for lung nodules. The combination of multi-scale dilation and attention provides a more robust representation, particularly beneficial for ambiguous or small nodules that may otherwise be missed.

In summary, the proposed architecture in this study builds upon the YOLOv11 framework by incorporating several key enhancements tailored for pulmonary nodule detection. The integration of the RepC3 + MSDA module introduces a synergistic mechanism where efficient encoding and multi-scale contextual attention work in tandem to improve detection robustness. These architectural innovations are designed not only to address the limitations of traditional YOLO variants in handling small and ambiguous nodules but also to maintain real-time performance and deployment feasibility. The neck architectural image, Figure 13, displays the organization and the whole architectural image, Figure 14, give a holistic view of how everything is broken down.

Although not explicitly shown in the architecture diagram, the RepC3 + MSDA module and the C2PSA block are located at different stages of the model. The C2PSA module is positioned near the end of the backbone, where it enhances deep feature representations using parallel self-attention. In contrast, the RepC3 + MSDA module is integrated into the early part of the neck to further enrich spatial and contextual features through reparameterized convolution and multi-scale dilation attention. These modules work in a complementary fashion—C2PSA improves the quality of backbone-generated features before fusion, while RepC3 + MSDA refines the aggregated features in the neck to support more accurate detection and classification.

The combination of reparameterized convolution, dilated attention, and spatial-channel refinement creates a highly specialized model for chest CT interpretation. Compared to standard YOLO backbones, our design captures a broader spectrum of nodule characteristics, leading to more accurate localization and classification. This makes the model particularly well-suited for clinical environments where both precision and speed are essential.

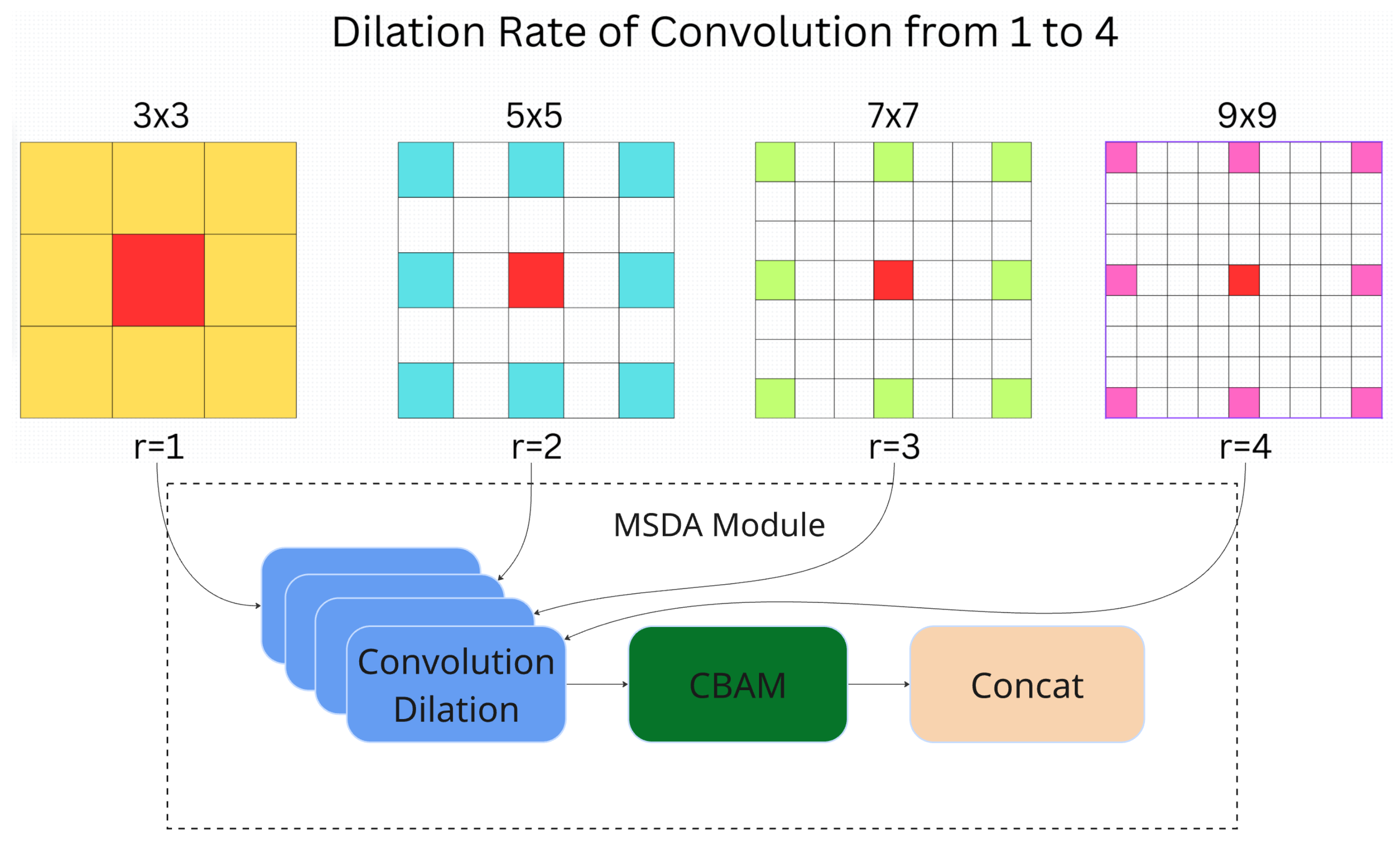


Figure 10. MSDA Module.

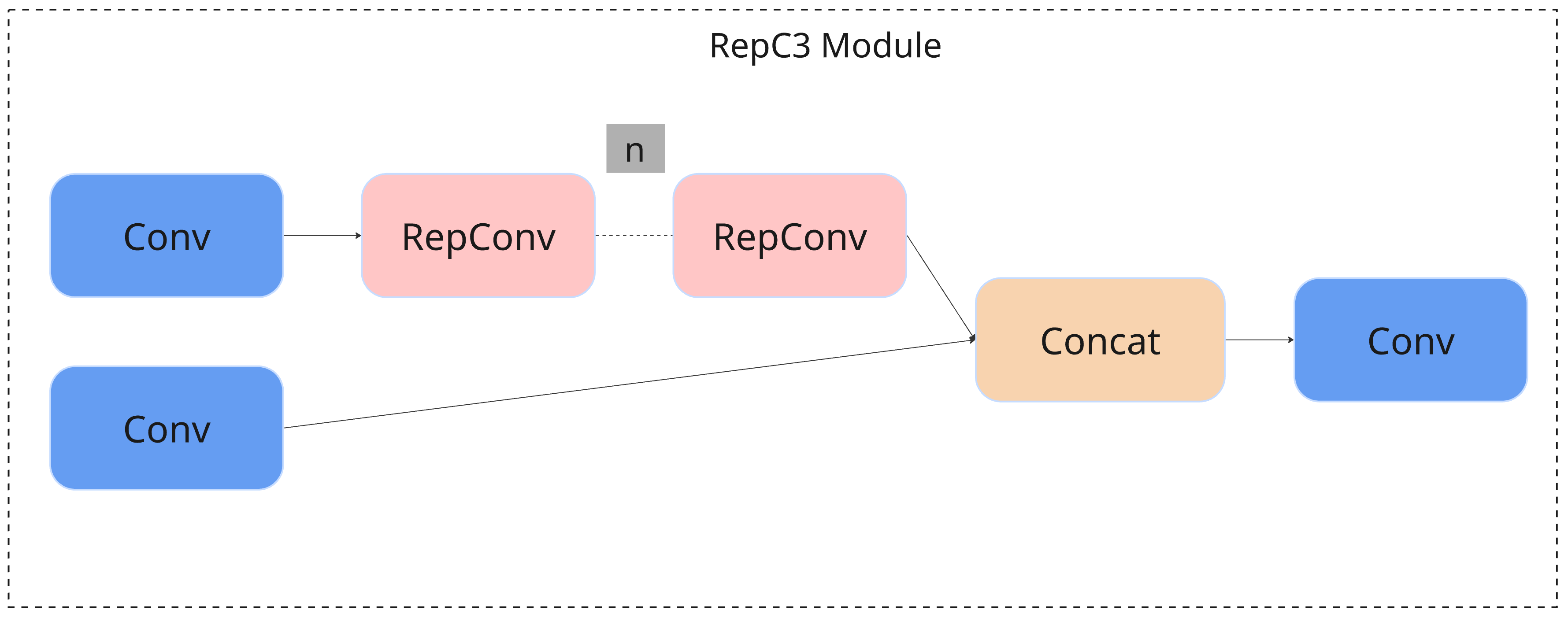


Figure 11. RepC3 Module. n indicates the number of times it will run RepConv.

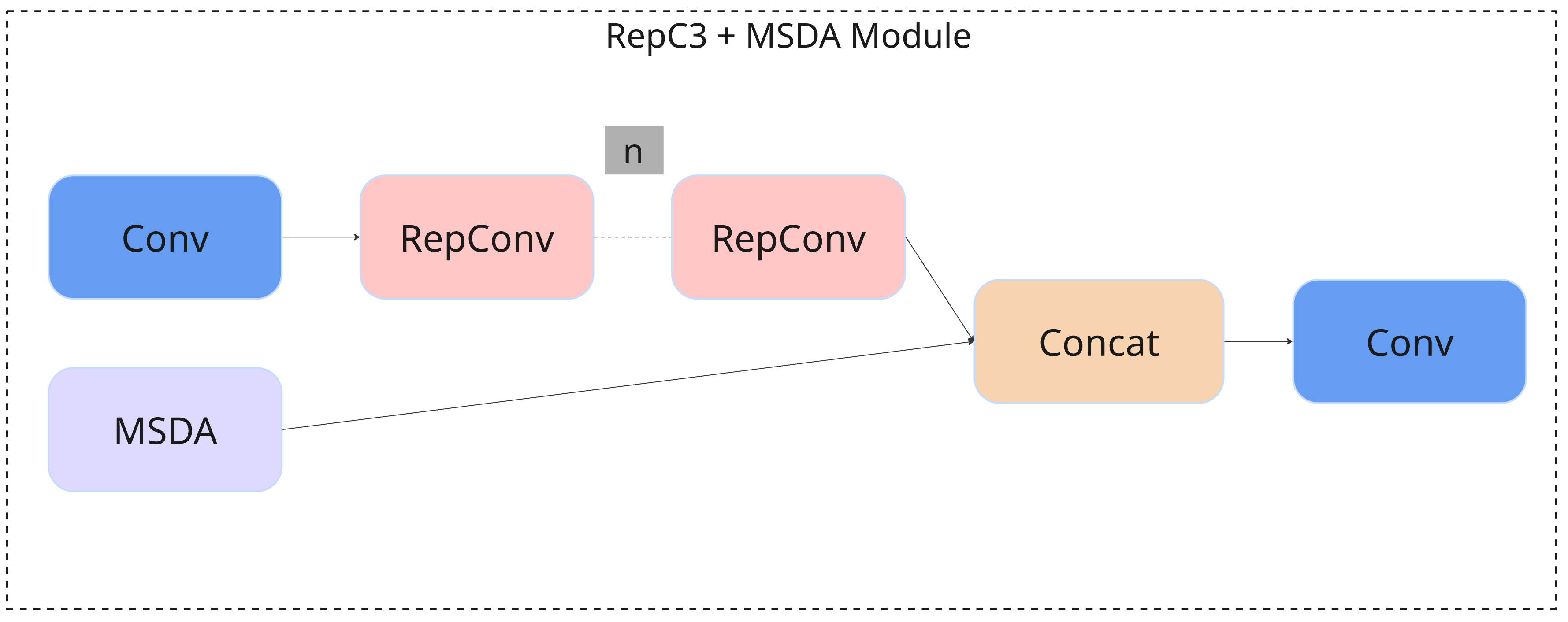


Figure 12. RepC3 + MSDA Module.

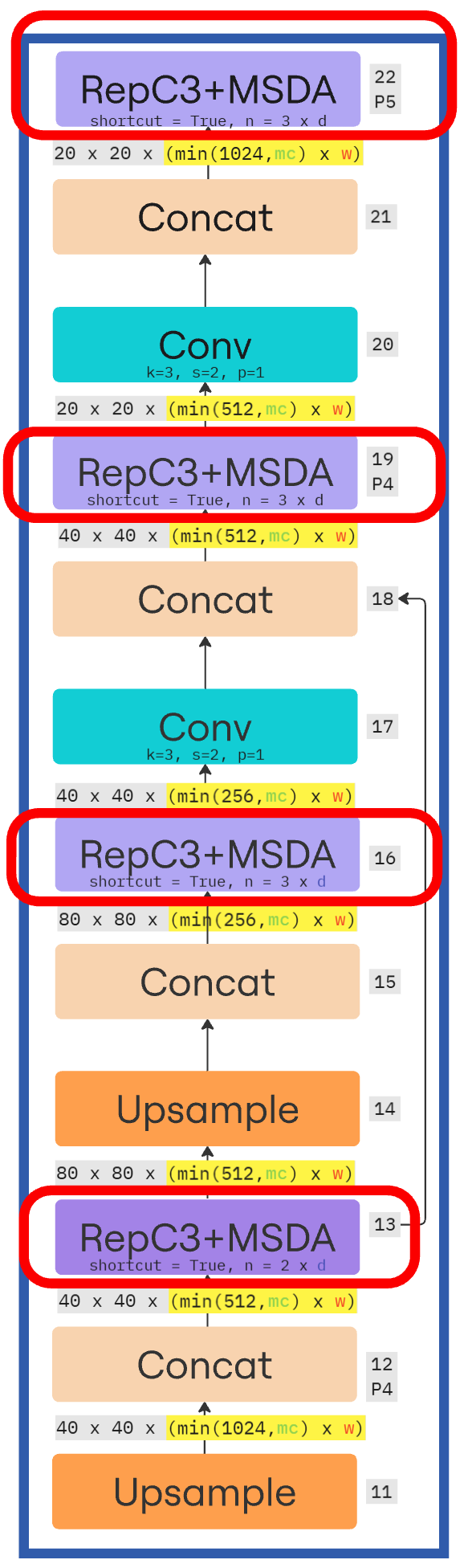


Figure 13. The Proposed YOLO Attention Framework Neck.

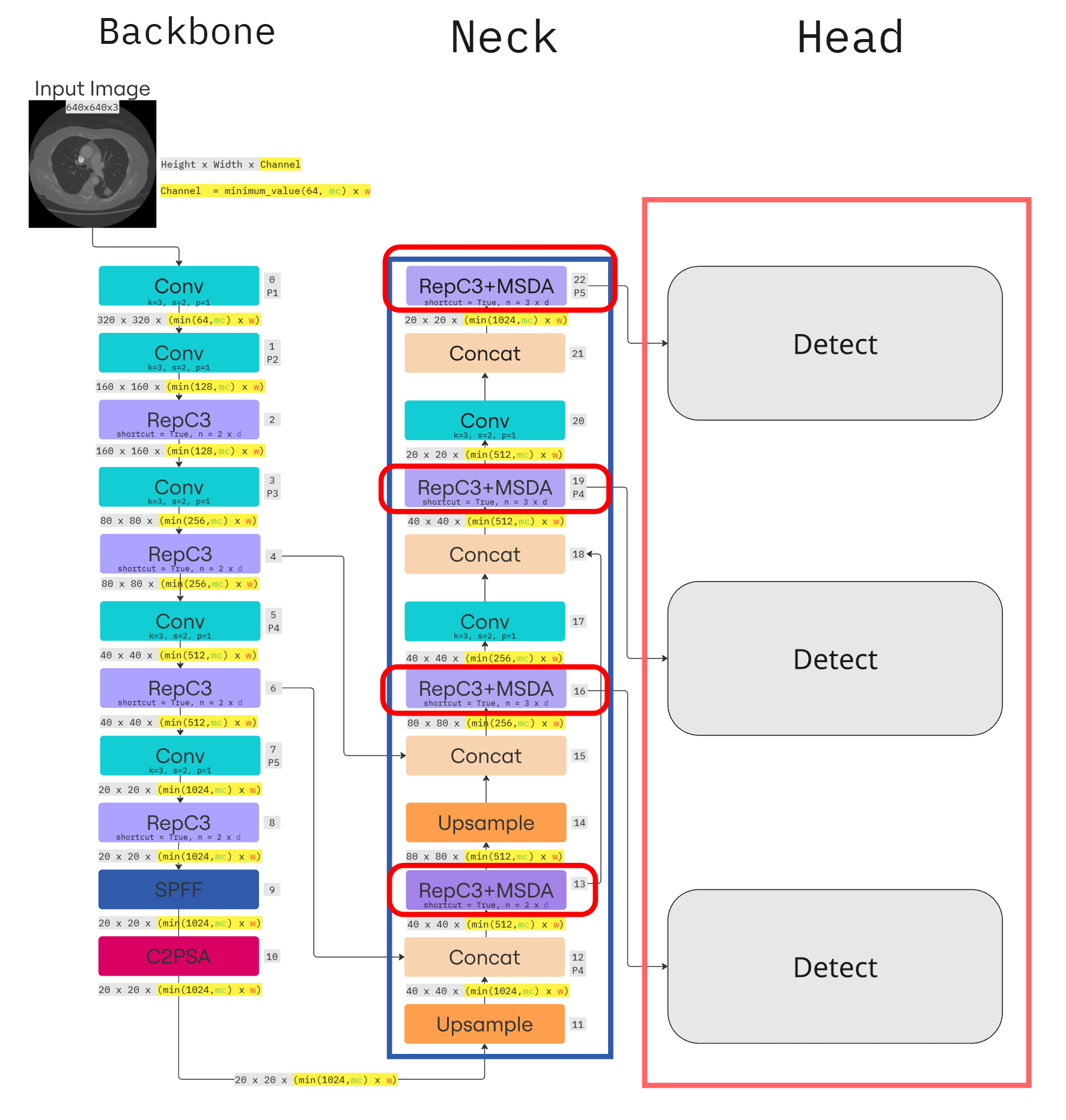


Figure 14. The Proposed YOLO Attention Framework.

# **Chapter 4: Experimental Results and Discussions**

This chapter presents the experimental setup, results, and analyses that evaluate the performance of the proposed lung nodule detection model. It outlines the dataset preparation, training configurations, and performance metrics used for model assessment. Through comparative experiments, ablation studies, and visualization of detection outputs, the chapter provides insights into how each architectural component contributes to the final results. Discussions highlight the clinical relevance and robustness of the proposed method, drawing attention to both its strengths and remaining limitations.

4.1 Dataset and Preprocessing

The dataset used in this study is the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI), a publicly available dataset curated to support the development and evaluation of computer-aided detection and diagnosis systems for lung cancer. The LIDC-IDRI was developed through a collaborative effort involving the National Cancer Institute (NCI), the Food and Drug Administration (FDA), and several academic institutions. Its primary objective was to provide a standardized reference database containing thoracic computed tomography (CT) scans with carefully annotated lung nodules, facilitating reproducible research in lung cancer imaging.

The dataset comprises of 1,018 patient thoracic CT scans, each of which was reviewed by four experienced thoracic radiologists. Annotations were carried out in a two-phase process: an initial blinded review where each radiologist independently assessed the scans, followed by an unblinded consensus review that allowed for discussion and modification of annotations. This rigorous review process yielded a comprehensive set of lung nodule annotations that include information on nodule size, type, texture, and malignancy suspicion levels.

For each patient case, the dataset includes a volumetric CT scan in DICOM format with high-resolution axial slices, typically acquired at a slice thickness of 1.25 mm. The annotations capture nodules that are 3 mm or larger in diameter and provide additional information such as the nodule's spatial location, margin characteristics, spiculation, and texture classification (e.g., solid, part-solid, or non-solid). Furthermore, non-nodule findings and tiny nodules under 3 mm are also documented. Each case also includes metadata such as patient identifiers, scan identifiers, voxel spacing, scanner details, and slice count.

Prior to model training, a comprehensive preprocessing pipeline was implemented to ensure data consistency and quality. CT scans that were corrupted, unreadable, or improperly formatted were excluded. Slices outside the thoracic region or those lacking complete lung parenchyma were removed to focus the model on relevant anatomical areas. In addition, nodules that were deemed too small (i.e., under 3 mm) or lacked sufficient inter-radiologist agreement were excluded from the final dataset. Cases with incomplete or inconsistent annotations were also filtered out. This preprocessing step resulted in a reduced and higher-quality dataset subset that preserved only the most clinically relevant samples. The cleaned dataset was then used for model training, validation, and evaluation.

The most important aspect that was used in determining if a patient’s lung nodule is cancerous is the malignancy score that was given by the four individual doctors doing the review. Each doctor assigned malignancy probability scores ranging from 1 (highly unlikely malignant) to 5 (highly suspicious for malignancy), with intermediate scores reflecting increasing likelihoods of malignancy (2: 5-30%, 3: 30-60%, 4: 60-80%). During the subsequent unblinded consensus review, discrepancies were resolved, and final annotations were established, including nodule characteristics and aggregated malignancy scores. For this research, nodules were classified into three categories based on consensus scores: benign (average score ≤ 2), ambiguous (score = 3), and malignant (score ≥ 4). The preprocessing pipeline excluded non-thoracic slices, nodules < 3 mm, and cases with insufficient radiologist agreement or inconsistent annotations, resulting in a refined dataset of clinically significant nodules with unambiguous labels for robust model training. The final dataset consisted of 3,313 benign, 5,102 ambiguous, and 4,834 malignant nodule slices.

Table 2 summarizes how the malignancy scores assigned by radiologists are interpreted. These categories form the ground truth labels used for model classification, reflecting real-world clinical uncertainty and decision-making thresholds. The LIDC-IDRI's standardized scoring system and multi-radiologist validation ensure the dataset's reliability for developing diagnostic AI models.

Table 2. Dataset Malignancy Interpretation.

|  |  |  |
| --- | --- | --- |
| Score | Clinical Interpretation | Likelihood of Malignancy |
| 1 | Highly unlikely malignant | < 5% |
| 2 | Moderately unlikely malignant | 5 – 30% |
| 3 | Indeterminate (equivocal malignancy) | 30 – 60% |
| 4 | Moderately suspicious for malignancy | 60 – 80% |
| 5 | Highly suspicious for malignancy | > 80% |

4.2 Training Configuration and Evaluation Metrics

The proposed model was trained and evaluated using a customized implementation of the YOLOv11 framework, with specific adaptations tailored to pulmonary nodule detection in CT images. The model was trained on the preprocessed subset of the LIDC-IDRI dataset using a learning rate scheduler based on cosine annealing, with the SDG optimizer selected to ensure stable and effective convergence. During training, a composite loss function was employed, consisting of CIoU (Complete Intersection over Union) loss for bounding box regression and Focal loss for classification. These loss functions were selected for their ability to balance detection precision and robustness in datasets with class imbalance and highly variable object sizes.

Table 3 outlines the key hyperparameters used during model training. These values were selected based on empirical tuning to ensure that the model achieved stable convergence while maintaining high accuracy across diverse nodule types.

To evaluate the performance of the model, several metrics commonly used in object detection were utilized. The primary metric used was mean Average Precision (mAP), a standard benchmark in the object detection domain. The mAP@0.5 score, or mean average precision at an Intersection over Union (IoU) threshold of 0.5, measures the average precision across all classes, where a predicted bounding box is considered correct if its IoU with the ground truth box is at least 0.5. A higher mAP@0.5 value indicates better detection accuracy. The IoU itself is defined as the ratio between the area of overlap and the area of union of the predicted bounding box and the ground truth bounding box:

(Eq. 4)

The precision and recall values are calculated at various confidence thresholds, and their relationship is used to construct a precision-recall (P-R) curve. The Average Precision (AP) is the area under this P-R curve. The mAP is then calculated by averaging the AP scores across all classes. In YOLO-based evaluations, mAP@0.5 is typically reported for basic benchmarking, while mAP@0.5:0.95—which averages the AP over IoU thresholds from 0.5 to 0.95 in increments of 0.05—provides a more comprehensive evaluation of localization accuracy.

In addition to mAP, the evaluation included the computation of precision, recall, and F1-score. Precision is defined as the number of true positive detections divided by the total number of predicted positives. High precision indicates that most of the predicted nodules are indeed true positives. Recall, on the other hand, is the number of true positives divided by the total number of actual positives. A high recall score suggests that the model is successful at identifying most of the nodules present in the images. The F1-score or Eq. 5, the harmonic mean of precision and recall, provides a single measure of detection effectiveness that balances the trade-off between the two:

(Eq. 5)

In the context of lung nodule detection, a high mAP@0.5 score is generally desirable, as it reflects strong localization performance with minimal false positives. However, for clinical applications, achieving a balance between high recall (to avoid missing nodules) and high precision (to reduce false alarms) is also critically important. The F1-score and mAP@0.5:0.95 are therefore particularly useful for assessing the model’s diagnostic utility.

Overall, the combination of these metrics provides a robust framework for evaluating both the accuracy and reliability of the YOLOv11-based detection model in identifying and localizing lung nodules in CT scans.

The hardware configuration listed in Table 4 reflects the computing resources used for training. Leveraging a high-end GPU (RTX 4090) and substantial RAM allowed for efficient processing of large CT volumes, which is critical when working with high-resolution medical imaging data.

Table 3. YOLO Training Parameters.

|  |  |
| --- | --- |
| Training Setup | Specification |
| Task | Detection |
| Mode | Train |
| Model Architecture | Proposed YOLOv11 (RepC3 + MSDA) |
| Epochs | 300 |
| Patience | 100 |
| Batch Size | 20 |
| Image Size | 640×640 |
| Device | GPU 1 (RTX 4090) |
| Optimizer | SGD |
| Initial LR (lr0) | 0.01 |
| Final LR (lrf) | 0.01 |
| Momentum | 0.937 |
| Weight Decay | 0.0005 |

Table 4. Hardware Specification.

|  |  |
| --- | --- |
| **Hardware** | **Specification** |
| Processor | 11th Gen Intel(R) Core(TM) i7-11700 @ 2.50GHz (8C/16T) |
| GPU | NVIDIA GeForce RTX 4090 (24GB GDDR6X) |
| RAM | 64GB DDR4 |
| Storage | 500GB NVMe SSD |

4.3 Results and Discussions

This section presents the experimental results of our lung nodule detection models, followed by in-depth analyses. The results are organized into three parts: the performance of the original YOLO architectures, an ablation study to evaluate the contribution of each proposed module, and a comparison with newer YOLO versions. Each subsection provides quantitative metrics and qualitative observations to highlight the improvements and trade-offs introduced by our enhancements.

### 4.3.1 Original YOLO Experiment Results

This study conducted a comprehensive evaluation of pulmonary nodule detection performance across three YOLO architectures - YOLOv8, YOLOv11, and our proposed YOLO model - using the standardized LIDC-IDRI dataset under identical training conditions. The baseline YOLOv8 established a solid foundation with 81.82% precision and 67.80% recall (F1-score: 74.06%), while its mAP scores of 75.17% (mAP@0.5) and 40.96% (mAP@0.5:0.95) revealed limitations in handling nodules with varying degrees of overlap. The subsequent YOLOv11 architecture showed marked improvements across all metrics, achieving 85.21% precision, 73.06% recall (F1-score: 78.63%), and higher mAP values of 79.74% and 43.11%, demonstrating enhanced feature extraction capabilities.

As shown in Figure 15, the YOLOv11 model is capable of identifying and localizing nodules in chest CT images using bounding boxes. While its detections are generally accurate, there are occasional limitations in detecting ambiguous or borderline nodules, particularly in complex cases.

Figure 16 shows detection results from the proposed YOLO model using the RepC3 + MSDA backbone. Compared to YOLOv11, the proposed model delivers tighter bounding boxes and demonstrates improved detection of subtle or ambiguous nodules, especially those with low contrast or atypical appearance.

Moving to classification performance, Figure 17 presents the normalized confusion matrix for YOLOv11. While performance is strong across all three nodule classes, the model occasionally confuses ambiguous nodules with benign or malignant ones, indicating a need for improved discriminative power.

In contrast, Figure 18 shows the confusion matrix for the proposed model, which exhibits reduced confusion between ambiguous and malignant categories. This improvement reflects the benefits of enhanced multi-scale and attention-driven feature encoding.

Figure 19 displays the precision-recall curve for YOLOv11. The curve reveals a reasonably balanced performance but shows a drop in precision at higher recall levels, suggesting challenges with low-confidence predictions.

Finally, Figure 20 illustrates the precision-recall curve of the proposed model, which maintains higher precision across a broader range of recall values. This highlights the model’s improved ability to identify true positives while minimizing false positives—an essential attribute for clinical screening applications.

Our proposed YOLO model, incorporating innovative multi-dilation attention modules and reparameterized backbone blocks, delivered superior performance with an optimal balance of 83.26% precision and the highest recall of 75.43% (F1-score: 79.12%). The model's exceptional mAP scores of 81.34% and 44.35% highlight its advanced ability to accurately localize nodules across different sizes and morphologies, particularly excelling in challenging cases like small (3-8mm) nodules and ground-glass opacities. This performance progression validates the effectiveness of our architectural innovations, especially the MSDA module's capacity to capture multi-scale nodule features while maintaining computational efficiency.

The results carry significant clinical implications, as the model's strong recall minimizes missed malignant nodules while its maintained precision controls false positives - a crucial balance for diagnostic applications. While the 44.35% mAP@0.5:0.95 indicates remaining challenges with borderline cases, the overall performance demonstrates our model's potential as a decision-support tool in clinical workflows, particularly valuable in resource-limited settings. These findings collectively show how our architectural enhancements address key challenges in automated pulmonary nodule detection while preserving the computational efficiency required for practical clinical implementation, though opportunities remain for further refinement of borderline case detection.

To further illustrate model performance, we present a side-by-side comparison between ground truth annotations and predicted bounding boxes on the validation dataset. Figure 21 shows the original CT slice images with ground truth labels provided by radiologists, while Figure 22 displays the corresponding predictions made by the proposed YOLO model. These visualizations highlight the model's capability to localize pulmonary nodules of varying sizes and opacity levels. In particular, the model demonstrates improved accuracy in detecting small and ambiguous nodules, with predicted confidence scores consistent with clinical expectations.

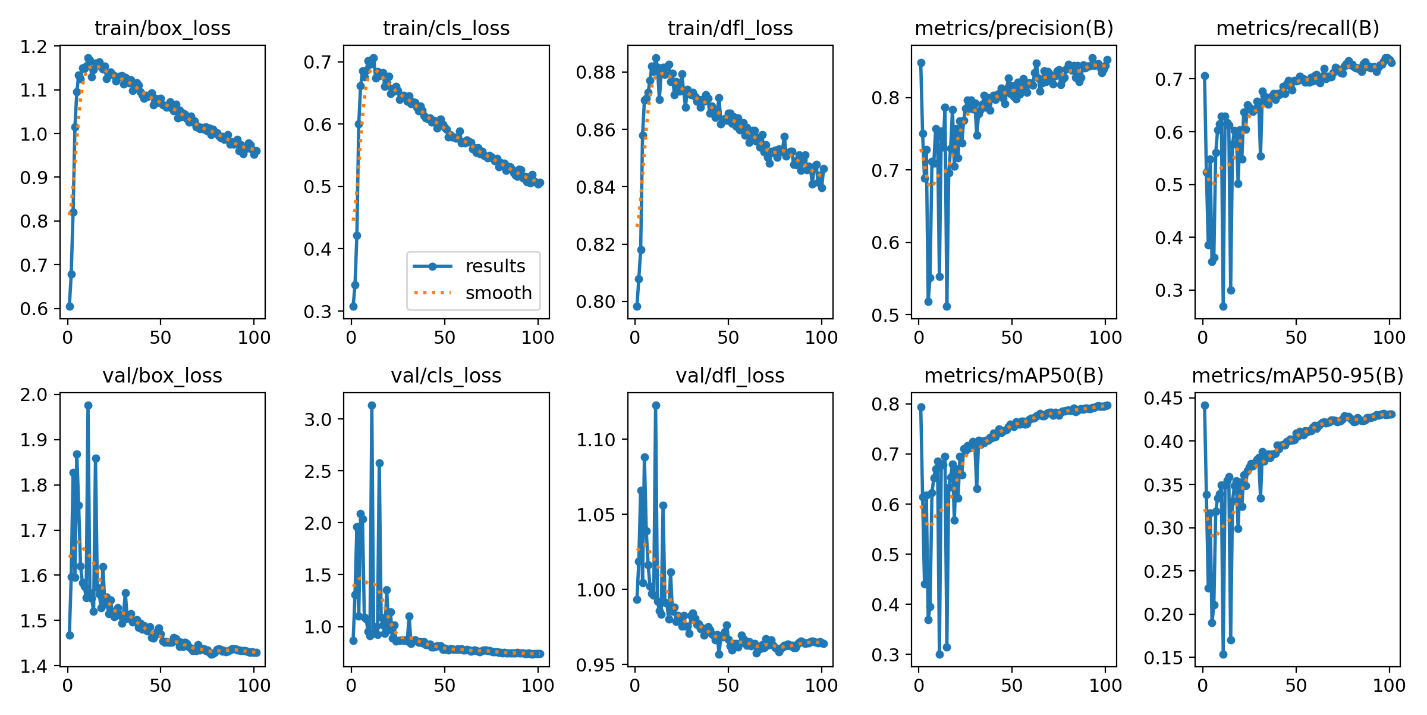


Figure 15. YOLOv11 Metrics and Training/Validation Loss Graphs. YOLOv11 achieves slightly higher early precision, the proposed architecture offers a more balanced and robust performance suitable for clinical deployment.

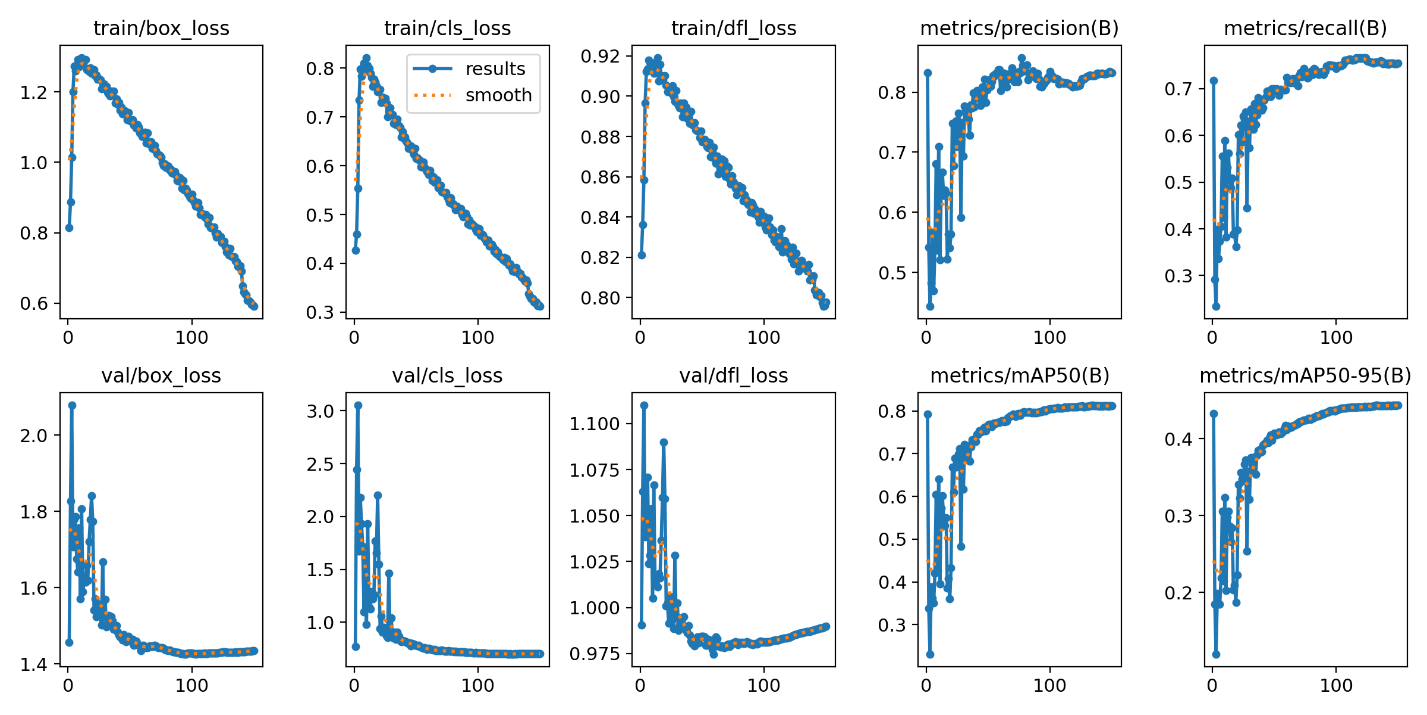


Figure 16. Proposed Architecture Metrics and Training/Validation Loss Graphs. The proposed model demonstrates smoother and faster convergence across all loss curves, higher recall, and improved mAP scores—particularly mAP@0.5:0.95—indicating enhanced ability to detect small or ambiguous nodules.



Figure 17. YOLOv11 Normalized Confusion Matrix.

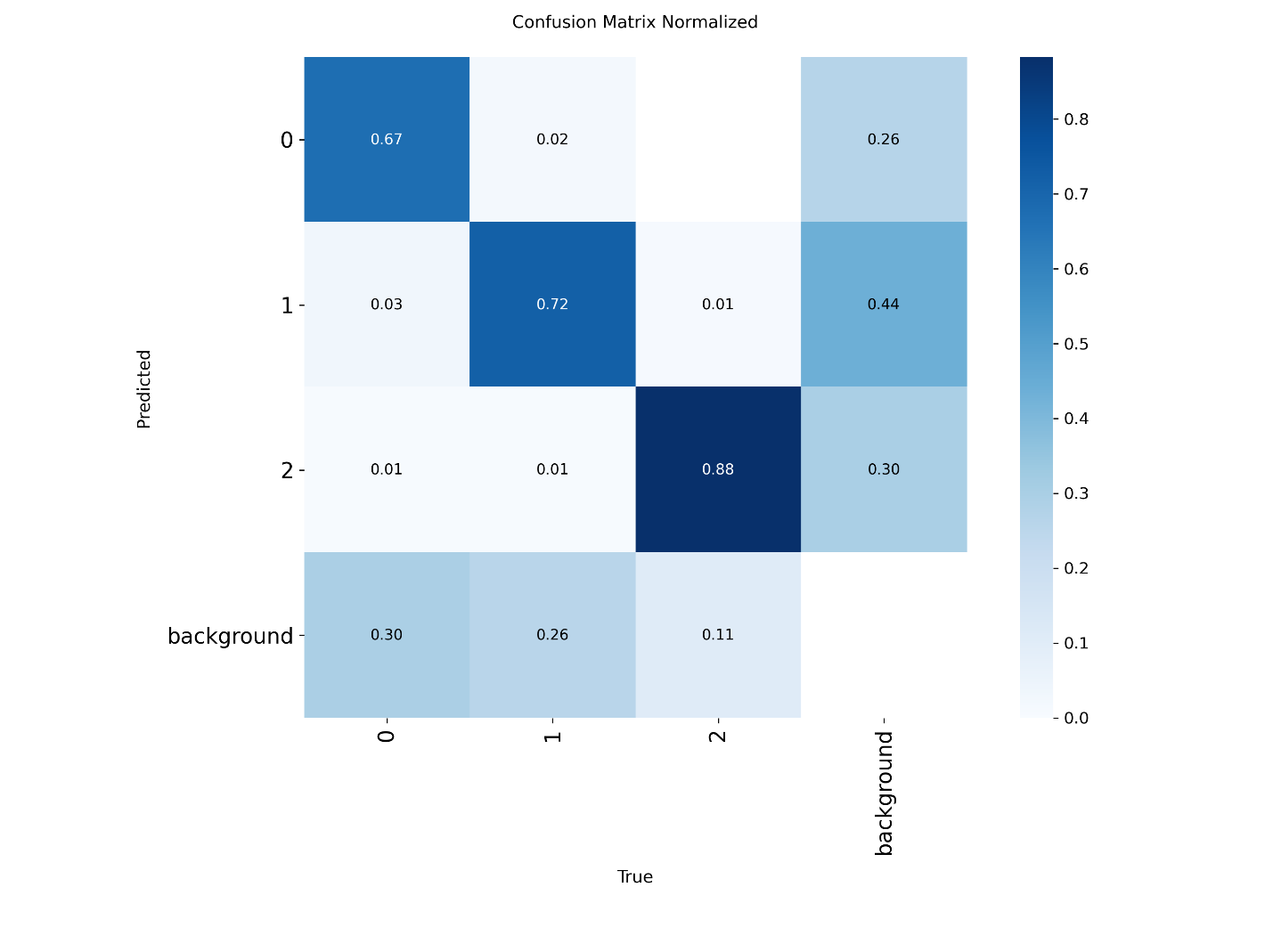


Figure 18. Proposed Architecture Normalized Confusion Matrix.

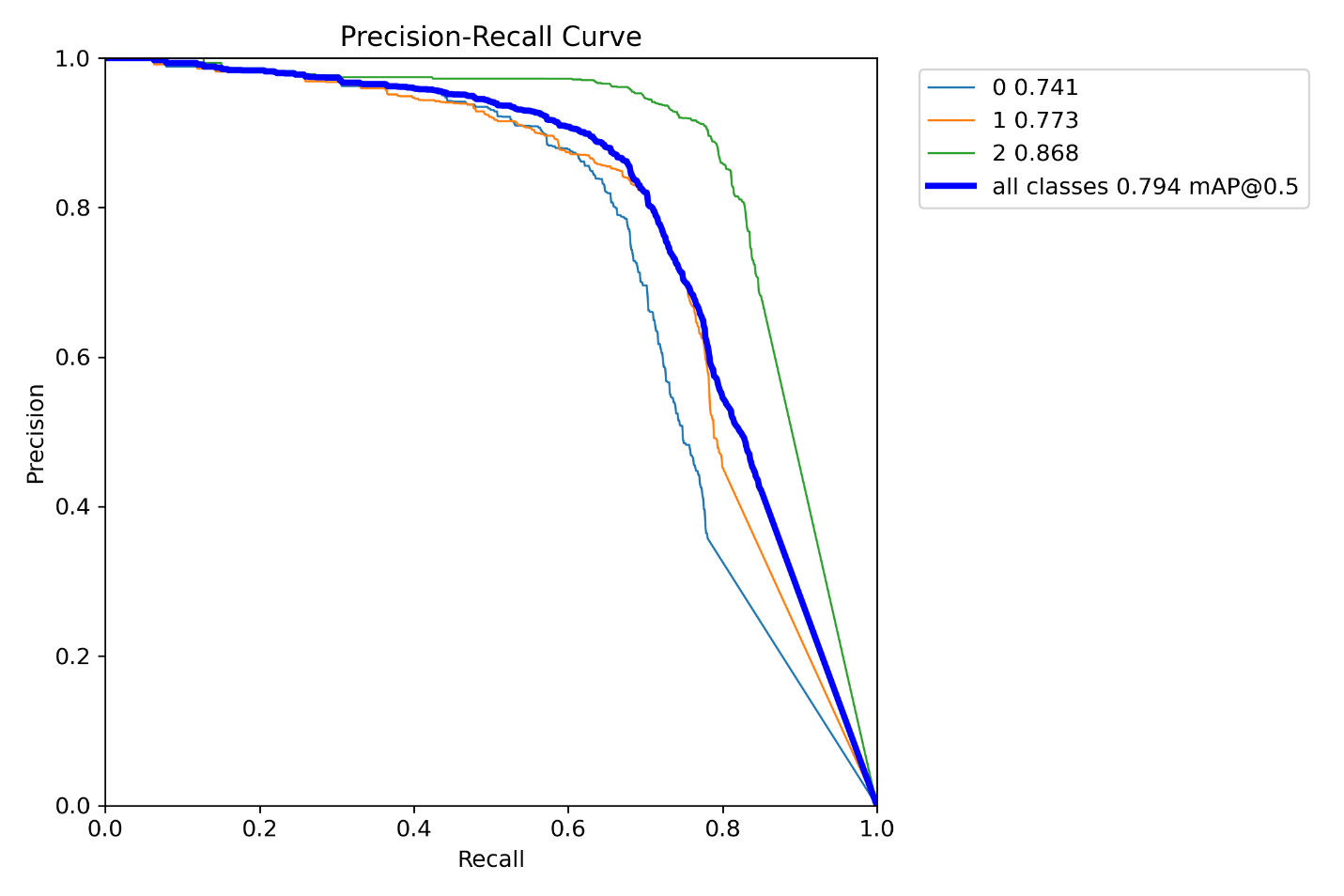


Figure 19. YOLOv11 Precision-Recall Curve.

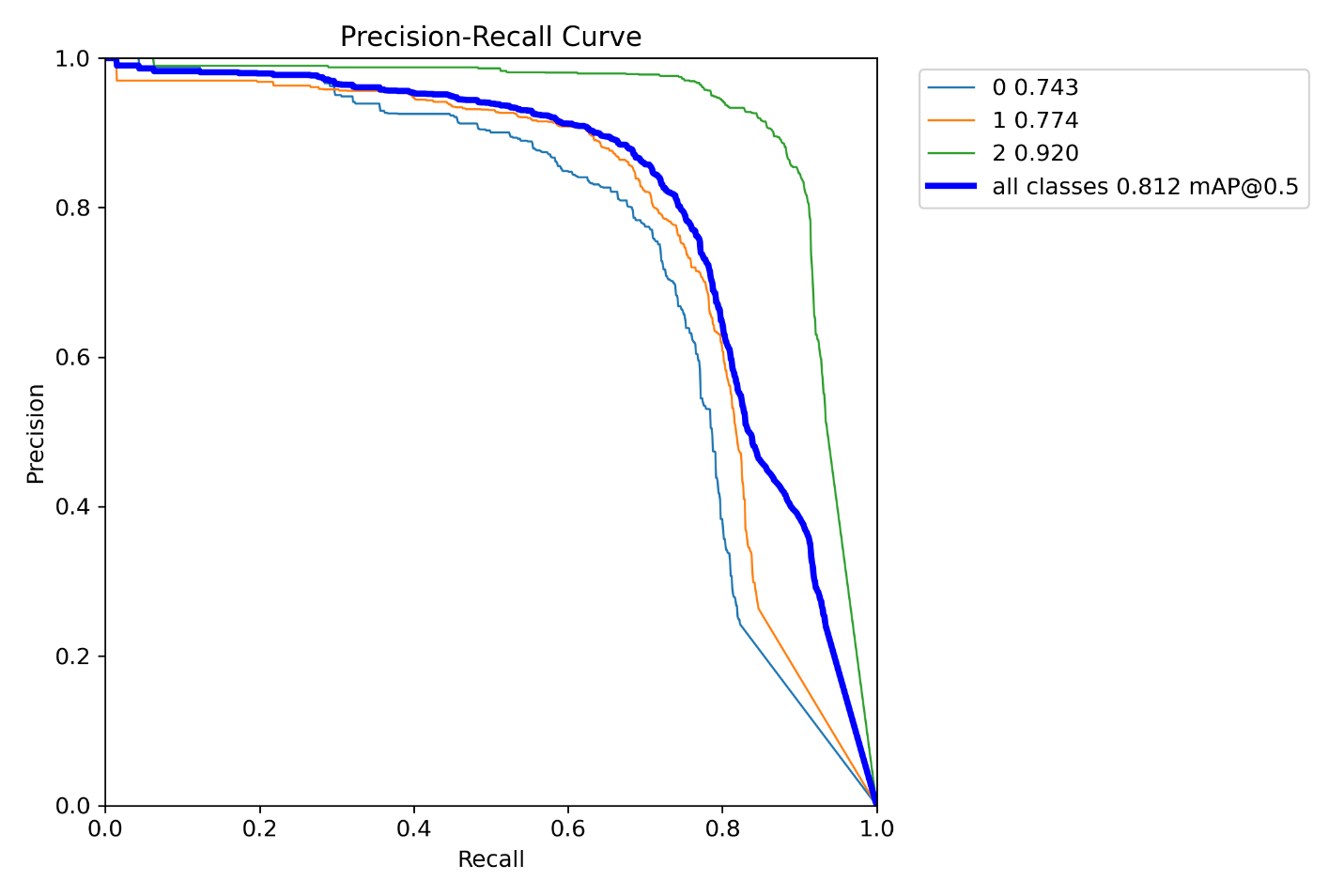


Figure 20. Proposed Architecture Precision-Recall Curve.

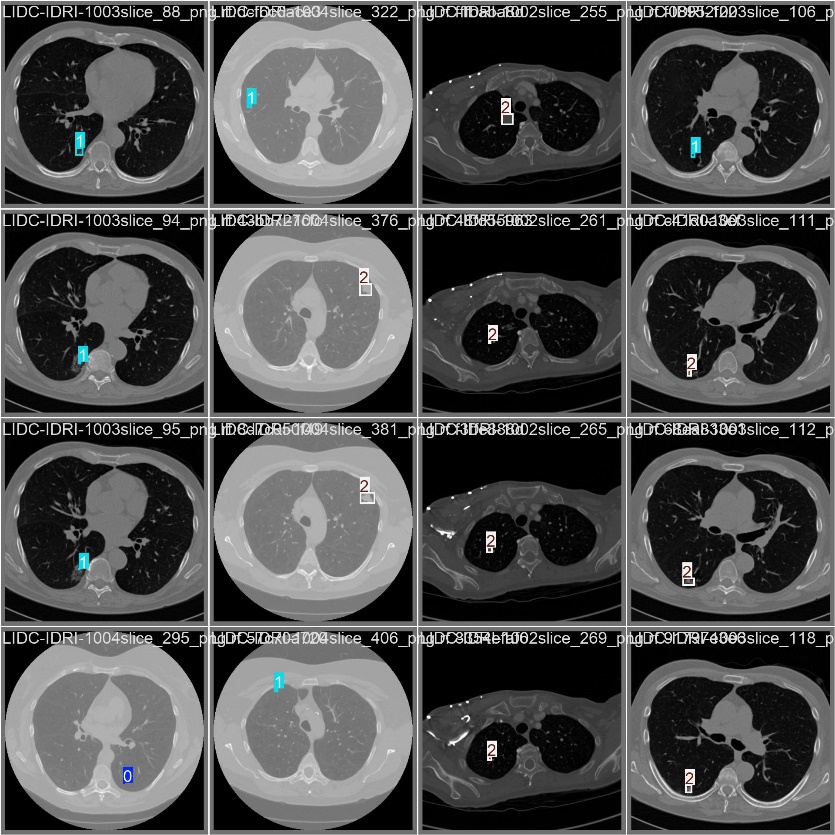


Figure 21. Validation label visualization for a batch of CT images. Ground truth bounding boxes are annotated with class indices (0: benign, 1: ambiguous, 2: malignant) and corresponding malignancy scores.

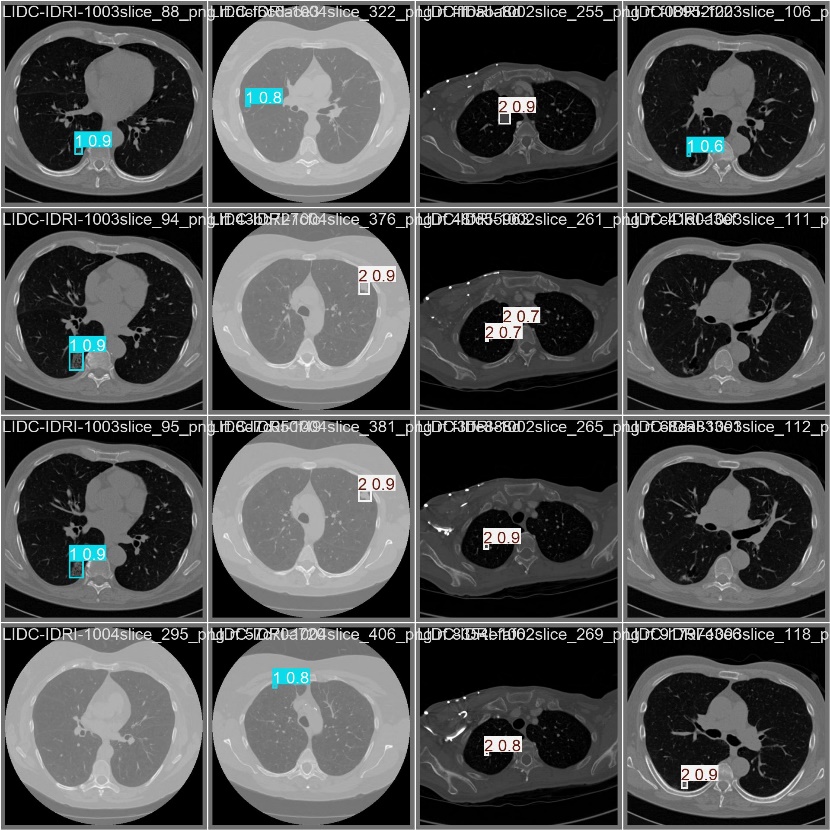


Figure 22. Prediction output of the proposed YOLO model on the same validation batch. The bounding boxes display predicted classes and confidence scores, demonstrating accurate localization and malignancy classification of pulmonary nodules.

### 4.3.2 Ablation Study Results

To further validate the effectiveness of each architectural component, we conducted an ablation study to analyze the individual impact of modules such as RepC3 and MSDA when introduced progressively into both YOLOv8 and YOLOv11 frameworks.

To better understand the contribution of each architectural enhancement, an ablation study was conducted. We started from the YOLOv8 baseline and incrementally introduced the proposed modules: RepC3, MSDA, and C2PSA, evaluating their individual and combined effects on model performance.

The evaluation focused on four key metrics—Precision, Recall, mAP@0.5, and mAP@0.5:0.95—using the LIDC-IDRI dataset under identical training conditions. Results are shown in Table 5.

The addition of the RepC3 module to the YOLOv11 architecture led to a noticeable improvement in recall and structural feature learning, particularly in the model’s ability to detect irregular or partially obscured nodules. RepC3 introduces reparameterized convolutional blocks that enhance training-time feature extraction while preserving inference efficiency. In the ablation study, the YOLOv11 + RepC3 configuration achieved a precision of 80.40% and a recall of 72.79%, resulting in an F1-score of 76.42%. The corresponding mAP scores—78.03% at IoU threshold 0.5 and 42.80% at 0.5:0.95—indicate that RepC3 strengthens the model’s capacity to localize nodules across a variety of shapes and sizes.

However, it is worth noting that the original YOLOv11 model still outperformed the RepC3-enhanced version in overall precision (85.21% vs. 80.40%) and mAP@0.5 (79.74% vs. 78.03%). This may be due to the fact that YOLOv11's architecture, including its C3K2 and C2PSA modules, is already highly optimized and well-balanced for general-purpose object detection. While RepC3 introduces improvements in structural encoding and recall, it may slightly disrupt the optimization equilibrium established in YOLOv11, leading to a minor trade-off in precision. This suggests that RepC3 alone is beneficial but may require additional modules to fully leverage its strengths without compromising other aspects of model performance.

In contrast, the YOLOv11 + MSDA configuration was designed to enhance contextual awareness and multi-scale feature sensitivity by incorporating multiple dilation rates within the attention mechanism. This allows the model to capture both fine and coarse details, which is particularly important when detecting small or ambiguous nodules, such as ground-glass opacities. The results from the ablation study show that YOLOv11 + MSDA achieved a precision of 73.77%, recall of 58.83%, and an F1-score of 65.46%, with mAP scores of 67.01% and 35.81% at IoU thresholds of 0.5 and 0.5:0.95, respectively.

While these metrics reflect improved attention to diverse nodule presentations, the standalone use of MSDA without a supporting structural module like RepC3 appears to limit its impact, particularly in recall and localization performance. This indicates that MSDA’s benefits are best realized when combined with modules that improve structural feature encoding, thereby reinforcing the value of the proposed full architecture.

RepC3 Block: When applied to the YOLOv8 baseline, RepC3 led to a clear performance boost, especially in recall and mAP metrics. The F1-score improved from 74.06% to 76.49%, confirming that RepC3’s reparameterized convolution enhances feature extraction efficiency while maintaining lightweight inference.

YOLOv11 Backbone: Integrating a more advanced transformer-enhanced backbone (YOLOv11) further improved detection of complex and small nodules. The performance increase across all metrics—from 77.74% to 79.74% (mAP@0.5)—indicates stronger contextual learning and better localization.

MSDA Module: When MSDA is combined with RepC3 inside the YOLOv11 backbone, the model gains the ability to detect ambiguous and multi-scale features more effectively. This configuration achieves the best mAP@0.5:0.95 (44.35%), showing that multi-dilation attention enhances contextual sensitivity for diverse nodule morphologies.

Together, these results confirm the contribution of each component: RepC3 improves structural encoding, MSDA enhances contextual awareness, and YOLOv11 boosts baseline capacity. The proposed model configuration (YOLOv11 + RepC3 + MSDA) ultimately provides the best balance of precision, recall, and detection accuracy, making it the most suitable for clinical deployment.

Table 5. Ablation study comparing YOLOv8 and YOLOv11 with and without RepC3 and MSDA modules.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model Variant** | **Precision(%)** | **Recall(%)** | **F1-Score(%)** | **mAP@0.5(%)** | **mAP@0.5:0.95(%)** |
| YOLOv8 (baseline) | 81.82% | 67.80% | 74.06% | 75.17% | 40.96% |
| YOLOv8 + RepC3 | 82.29% | 71.35% | 76.49% | 77.74% | 43.00% |
| YOLOv11 | 85.21% | 73.06% | 78.63% | 79.74% | 43.11% |
| YOLOv11 + RepC3 | 80.40% | 72.79% | 76.42% | 78.03% | 42.80% |
| YOLOv11 + MSDA | 73.77% | 58.83% | 65.46% | 67.01% | 35.81% |
| YOLOv11 + RepC3 + MSDA (Ours) | 83.26% | 75.43% | 79.12% | 81.34% | 44.35% |

### 4.3.3 Architectural Comparison: YOLOv11, YOLOv12, and the Proposed Model

It is important to note that during the development of the proposed model, the YOLOv12 architecture had not yet been released. All testing and experimentation were conducted using YOLOv11, which at the time represented the most advanced available model. The primary difference between YOLOv12 and YOLOv11 lies in the introduction of an additional attention module in YOLOv12 for enhanced feature extraction.

YOLOv11 employs the C3K2 and C2PSA modules, while YOLOv12 retains C3K2 but replaces C2PSA and SPPF with a newly introduced module called A2C2F, which integrates characteristics from both C3K2 and A2C2F. While YOLOv12 offers this new component, it omits two key modules from YOLOv11 that were significant for feature extraction and spatial processing. In contrast, the proposed model was specifically tailored for lung nodule detection and classification, incorporating domain-specific enhancements such as the MSDA module. These design choices enable more effective detection of small or ambiguous nodules, highlighting the model’s focus on clinical applicability in pulmonary imaging.

Overall, while YOLOv11 showed notable enhancements over YOLOv8, the proposed model further improved detection precision and robustness, making it a more suitable candidate for clinical deployment in automated lung cancer screening systems.

Together, Figure 15 to Figure 20 provide a comprehensive visual comparison of the detection and classification performance across the baseline YOLOv11 and the proposed model. These visualizations reinforce the quantitative metrics in Table 6and demonstrate the improved interpretability, localization, and diagnostic reliability of our architecture. The consistent enhancements in bounding box accuracy, class confusion reduction, and precision-recall balance suggest that our model is well-suited for real-world clinical application, particularly in early lung cancer screening tasks.

Table 6. Comparison Results of Different YOLO versions and Proposed version.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Precision** | **Recall** | **F1-Score** | **mAP@0.5** | **mAP@0.5:0.95** |
| YOLOv8 | 81.82% | 67.80% | 74.06% | 75.17% | 40.96% |
| YOLOv11 | 85.21% | 73.06% | 78.63% | 79.74% | 43.11% |
| Proposed YOLO (Ours) | 83.26% | 75.43% | 79.12% | 81.34% | 44.35% |

# **Chapter 5: Conclusion and Future Works**

The final chapter concludes the research by summarizing the key findings and contributions of the study. It revisits the motivation behind integrating enhanced YOLO-based architectures for lung cancer detection and reflects on how the proposed methods improved performance. In addition, this chapter outlines possible directions for future research, including the use of volumetric data, advanced attention mechanisms, and clinical deployment considerations. These prospects aim to further enhance the diagnostic power and practical applicability of AI-assisted lung cancer screening.

5.1 Conclusion

In this study, we addressed the critical issue of early lung cancer detection by leveraging the capabilities of deep learning, specifically through the implementation and enhancement of the YOLOv11 architecture. Lung cancer continues to be one of the deadliest forms of cancer worldwide, with early detection and accurate classification being key to improving patient outcomes. Traditional imaging-based diagnostics, while effective, are limited by inter-observer variability and diagnostic delays. Therefore, the integration of AI-based object detection systems into medical imaging workflows offers a powerful tool for accelerating diagnosis and improving accuracy.

The proposed research introduced a modified YOLOv11 framework for the detection and classification of pulmonary nodules in chest CT images. Enhancements such as the integration of attention mechanisms via dilated convolutions, as well as the incorporation of novel modules like C3K2, RepC3, and C2PSA, contributed to the model's ability to better capture spatial hierarchies and subtle nodule features. These architectural modifications led to improvements in detection robustness, especially for ambiguous or small nodules that often challenge traditional algorithms.

The system was trained and validated using the LIDC-IDRI dataset—a well-established and diverse collection of thoracic CT scans with annotated nodules. Rigorous preprocessing steps were employed to refine the dataset and focus on clinically relevant cases. Performance metrics including mAP@0.5, mAP@0.5:0.95, precision, recall, and F1-score were used to assess model effectiveness. Results demonstrated that the proposed attention-enhanced YOLOv11 achieved strong localization accuracy while maintaining a balanced trade-off between sensitivity and specificity, which is essential for clinical deployment.

This research contributes to the growing field of AI-assisted medical diagnosis and provides a strong foundation for future development of automated lung cancer screening systems.

5.2 Future Works

Although the proposed model achieved promising results, there are several opportunities for future enhancement and exploration. First, while the current model was trained on static CT images, future work could incorporate temporal information from sequential scans or 3D volumetric models to capture temporal progression of nodules. This would enhance the model's ability to detect early-stage growth patterns.

Second, the model could be extended to multi-class classification tasks to distinguish not only between benign, ambiguous, and malignant nodules, but also between various histological subtypes (e.g., adenocarcinoma, squamous cell carcinoma) based on radiomic features. Incorporating clinical metadata such as smoking history, age, or genetic biomarkers could further improve classification accuracy through multimodal learning.

Third, attention mechanisms used in this study could be improved by implementing more advanced modules such as self-attention transformers or deformable attention blocks, which may better model long-range dependencies and irregular nodule morphology. The efficiency of the model could also be optimized for deployment on edge devices such as mobile platforms or embedded systems, enabling widespread use in low-resource or remote settings.

Finally, external validation on unseen hospital datasets is necessary to assess the model's generalizability and clinical readiness. Collaborative efforts with healthcare providers could enable deployment in real-world radiology workflows, supported by user studies to evaluate interpretability and integration into radiologist decision-making.

# **References**

[1] World Health Organization, “Cancer,” World Health Organization, Feb. 03, 2025. https://www.who.int/news-room/fact-sheets/detail/cancer

[2] D. Ardila et al., “End-to-end Lung Cancer Screening with three-dimensional Deep Learning on low-dose Chest Computed Tomography,” Nature Medicine, vol. 25, no. 6, pp. 954–961, May 2019, doi: https://doi.org/10.1038/s41591-019-0447-x.

[3] F. Islami et al., “Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States,” CA: A Cancer Journal for Clinicians, vol. 68, no. 1, pp. 31–54, Nov. 2017, doi: https://doi.org/10.3322/caac.21440.

[4] H. MacMahon et al., “Guidelines for Management of Incidental Pulmonary Nodules Detected on CT Images: From the Fleischner Society 2017,” Radiology, vol. 284, no. 1, pp. 228–243, Jul. 2017, doi: https://doi.org/10.1148/radiol.2017161659.

[5] A. Esteva et al., “A Guide to Deep Learning in Healthcare,” Nature Medicine, vol. 25, no. 1, pp. 24–29, Jan. 2019, doi: https://doi.org/10.1038/s41591-018-0316-z.

[6] P. J. Mazzone et al., “Components Necessary for High-Quality Lung Cancer Screening,” Chest, vol. 147, no. 2, pp. 295–303, Feb. 2015, doi: https://doi.org/10.1378/chest.14-2500.

[7] A. A. A. Setio et al., “Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge,” Medical image analysis, vol. 42, pp. 1–13, 2017, doi: https://doi.org/10.1016/j.media.2017.06.015.

[8] Q. Dou, H. Chen, L. Yu, J. Qin, and P.-A. Heng, “Multilevel Contextual 3-D CNNs for False Positive Reduction in Pulmonary Nodule Detection,” IEEE Transactions on Biomedical Engineering, vol. 64, no. 7, pp. 1558–1567, Jul. 2017, doi: https://doi.org/10.1109/tbme.2016.2613502.

[9] W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, “Multi-scale Convolutional Neural Networks for Lung Nodule Classification,” Lecture Notes in Computer Science, pp. 588–599, 2015, doi: https://doi.org/10.1007/978-3-319-19992-4\_46.

[10] H. Jiang, H. Ma, W. Qian, and M. Gao, “Detection of Pulmonary Nodules in CT Images Based on YOLOv4,” International Journal of Imaging Systems and Technology, vol. 31, no. 4, pp. 1806–1816, 2021.

[11] Mehdi Astaraki, G. Yang, Yousuf Zakko, Iuliana Toma-Dasu, Örjan Smedby, and C. Wang, “A Comparative Study of Radiomics and Deep-Learning Based Methods for Pulmonary Nodule Malignancy Prediction in Low Dose CT Images,” Frontiers in Oncology, vol. 11, Dec. 2021, doi: https://doi.org/10.3389/fonc.2021.737368.

[12] J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,” arxiv.org, Apr. 2018, doi: https://doi.org/10.48550/arXiv.1804.02767.

[13] None Sakshiwala and M. P. Singh, “A New Framework for multi-scale CNN-based Malignancy Classification of Pulmonary Lung Nodules,” Journal of Ambient Intelligence and Humanized Computing, vol. 14, no. 5, pp. 4675–4683, Sep. 2022, doi: https://doi.org/10.1007/s12652-022-04368-w.

[14] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, “YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,” arXiv (Cornell University), Jul. 2022, doi: https://doi.org/10.48550/arxiv.2207.02696.

[15] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: https://doi.org/10.1109/tpami.2016.2577031.

[16] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” arXiv.org, Apr. 22, 2020. https://www.arxiv.org/abs/2004.10934v1

[17] J. Mei, M.-M. Cheng, G. Xu, L.-R. Wan, and H. Zhang, “SANet: A Slice-Aware Network for Pulmonary Nodule Detection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1–1, 2021, doi: https://doi.org/10.1109/tpami.2021.3065086.

[18] H. Zhang, K. Zu, J. Lu, Y. Zou, and D. Meng, “EPSANet: An Efficient Pyramid Squeeze Attention Block on Convolutional Neural Network,” Lecture notes in computer science, pp. 541–557, Jan. 2023, doi: https://doi.org/10.1007/978-3-031-26313-2\_33.

[19] D. E. Ost and M. K. Gould, “Decision Making in Patients with Pulmonary Nodules,” American Journal of Respiratory and Critical Care Medicine, vol. 185, no. 4, pp. 363–372, Feb. 2012, doi: https://doi.org/10.1164/rccm.201104-0679ci.

[20] R. M. Tuder and I. Petrache, “Pathogenesis of chronic obstructive pulmonary disease,” Journal of Clinical Investigation, vol. 122, no. 8, pp. 2749–2755, Aug. 2012, doi: https://doi.org/10.1172/jci60324.

[21] M. Sánchez, M. Benegas, and I. Vollmer, “Management of incidental lung nodules ,” Journal of Thoracic Disease, vol. 10, no. Suppl 22, pp. S2611–S2627, Aug. 2018, doi: https://doi.org/10.21037/jtd.2018.05.86.

[22] G. D. Rubin, “Computed Tomography: Revolutionizing the Practice of Medicine for 40 Years,” Radiology, vol. 273, no. 2S, pp. S45–S74, Nov. 2014, doi: https://doi.org/10.1148/radiol.14141356.

[23] N. Horeweg et al., “Characteristics of lung cancers detected by computer tomography screening in the randomized NELSON trial,” American Journal of Respiratory and Critical Care Medicine, vol. 187, no. 8, pp. 848–854, Apr. 2013, doi: https://doi.org/10.1164/rccm.201209-1651OC.

[24] D. Wormanns and S. Diederich, “Characterization of small pulmonary nodules by CT,” European Radiology, vol. 14, no. 8, May 2004, doi: https://doi.org/10.1007/s00330-004-2335-z.

[25] S. G. Armato 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. 915–31, 2011, doi: https://doi.org/10.1118/1.3528204.

[26] W. Liu et al., “SSD: Single Shot MultiBox Detector,” Computer Vision – ECCV 2016, vol. 9905, pp. 21–37, 2016, doi: https://doi.org/10.1007/978-3-319-46448-0\_2.

[27] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, “Focal loss for dense object detection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1–1, 2018, doi: https://doi.org/10.1109/tpami.2018.2858826.

[28] S. N. Rao, “YOLOv11 Architecture Explained: Next-Level Object Detection with Enhanced Speed and Accuracy,” Medium, Oct. 22, 2024. https://medium.com/@nikhil-rao-20/yolov11-explained-next-level-object-detection-with-enhanced-speed-and-accuracy-2dbe2d376f71