



YOLOv8-BCD: a real-time deep learning framework for pulmonary nodule detection in computed tomography imaging

Wenjun Zhu¹, Xinyue Wang², Jie Xing¹, Xu Steven Xu³, Min Yuan^{1,4}

¹Department of Health Data Science, Anhui Medical University, Hefei, China; ²The Second School of Clinical Medicine, Anhui Medical University, Hefei, China; ³Clinical Pharmacology and Quantitative Science, Genmab Inc., Princeton, NJ, USA; ⁴MOE Key Laboratory of Population Health Across Life Cycle, Hefei, China

Contributions: (I) Conception and design: M Yuan, W Zhu; (II) Administrative support: M Yuan; (III) Provision of study materials or patients: M Yuan; (IV) Collection and assembly of data: W Zhu, X Wang, J Xing; (V) Data analysis and interpretation: W Zhu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Min Yuan, PhD. Department of Health Data Science, Anhui Medical University, 81 Meishan Rd., Hefei 230032, China; MOE Key Laboratory of Population Health Across Life Cycle, 81 Meishan Rd., Hefei 230032, China. Email: myuan@ustc.edu.cn.

Background: Lung cancer remains one of the malignant tumors with the highest global morbidity and mortality rates. Detecting pulmonary nodules in computed tomography (CT) images is essential for early lung cancer screening. However, traditional detection methods often suffer from low accuracy and efficiency, limiting their clinical effectiveness. This study aims to devise an advanced deep-learning framework capable of achieving high-precision, rapid identification of pulmonary nodules in CT imaging, thereby facilitating earlier and more accurate diagnosis of lung cancer.

Methods: To address these issues, this paper proposes an improved deep-learning framework named YOLOv8-BCD, based on YOLOv8 and integrating the BiFormer attention mechanism, Content-Aware ReAssembly of Features (CARAFE) up-sampling method, and Depth-wise Over-Parameterized Depth-wise Convolution (DO-DConv) enhanced convolution. To overcome common challenges such as low resolution, noise, and artifacts in lung CT images, the model employs Super-Resolution Generative Adversarial Network (SRGAN)-based image enhancement during preprocessing. The BiFormer attention mechanism is introduced into the backbone to enhance feature extraction capabilities, particularly for small nodules, while CARAFE and DO-DConv modules are incorporated into the head to optimize feature fusion efficiency and reduce computational complexity.

Results: Experimental comparisons using 550 CT images from the LUNG Nodule Analysis 2016 dataset (LUNA16 dataset) demonstrated that the proposed YOLOv8-BCD achieved detection accuracy and mean average precision (mAP) at an intersection over union (IoU) threshold of 0.5 (mAP_{0.5}) of 86.4% and 88.3%, respectively, surpassing YOLOv8 by 2.2% in accuracy, 4.5% in mAP_{0.5}. Additional evaluation on the external TianChi lung nodule dataset further confirmed the model's generalization capability, achieving an mAP_{0.5} of 83.8% and mAP_{0.5-0.95} of 43.9% with an inference speed of 98 frames per second (FPS).

Conclusions: The YOLOv8-BCD model effectively assists clinicians by significantly reducing interpretation time, improving diagnostic accuracy, and minimizing the risk of missed diagnoses, thereby enhancing patient outcomes.

Keywords: Pulmonary nodule detection; YOLOv8; BiFormer; Content-Aware ReAssembly of Features (CARAFE); Depth-wise Over-Parameterized Depth-wise Convolution (DO-DConv)

Submitted Apr 04, 2025. Accepted for publication Jul 04, 2025. Published online Aug 12, 2025.

doi: 10.21037/qims-2025-824

View this article at: <https://dx.doi.org/10.21037/qims-2025-824>

Introduction

Lung cancer is among the leading causes of cancer-related mortality worldwide (1). Pulmonary nodules, which are small lung abnormalities that may indicate the presence of lung cancer, are commonly identified through routine imaging techniques like computed tomography (CT) scans (2). Detecting these nodules at an early stage is vital for enabling timely interventions and enhancing survival rates (3).

Manual detection of pulmonary nodules is often challenging due to their small size, subtle appearance, and the inherent complexity of medical imaging data, which heavily relies on the clinician's expertise (4). In clinical practice, various detection methods are commonly employed to identify pulmonary nodules in CT images, assisting radiologists in the early diagnosis of lung diseases, particularly lung cancer. Traditional approaches such as thresholding (5) and morphological processing (6) are widely used for their simplicity and ease of implementation; however, they face difficulties in more complex cases where nodules are hard to differentiate from surrounding tissues (7). Machine learning methods, including support vector machines (SVM) (8) and K-nearest neighbors (KNN) (9) classifiers, have been applied to improve detection accuracy by classifying image features (10). However, these methods are often sensitive to noise and necessitate manual feature extraction (11). While effective in certain situations, these classical machine learning techniques are limited by their inability to fully capture the complex patterns in medical images, especially when detecting subtle or irregularly shaped nodules.

Deep learning methods, particularly those based on Convolutional Neural Networks (CNNs) (12) and Transformer Networks (13), have gained significant popularity in medical imaging due to their ability to automatically learn hierarchical features from raw data (14). These models excel at capturing intricate details of complex structures, such as pulmonary nodules, leading to more accurate detection compared to traditional approaches (15). You Only Look Once (YOLO), introduced by Redmon *et al.* in 2016, aimed to achieve efficient and accurate object detection by transforming the problem into a regression task (16). The core concept of YOLO involves dividing the input image into grids and simultaneously predicting bounding boxes and class probabilities within each grid cell, facilitating end-to-end object detection. Among the YOLO series, YOLOv8 has become particularly popular

due to its deeper architecture, improved backbone, refined detection head, and optimized loss functions, all contributing to enhanced accuracy and computational efficiency (17). The choice of YOLOv8 as the baseline over newer variants such as YOLOv9 or YOLOv10 was driven by considerations of practical compatibility, computational efficiency, and the specific requirements of medical imaging applications (18). Although YOLOv9 and YOLOv10 may achieve superior results on general-purpose datasets, their characteristics make them less ideal for clinical tasks requiring interpretability, efficiency, and robustness to small, low-contrast targets. YOLOv9 has higher computational demands, while YOLOv10 performs poorly in detecting small targets, leading to missed detections and localization inaccuracies (19). Nevertheless, challenges such as low resolution, image noise, blurred boundaries, and the diverse morphology of lung nodules persist, limiting the effectiveness of existing models in real-time clinical scenarios (20). These ongoing challenges underscore the necessity for continued innovations in deep learning methodologies, aiming to further enhance detection accuracy, reduce computational requirements, and improve generalization across diverse clinical environments.

In this paper, we present a novel deep learning framework, YOLOv8-BCD, which is built upon the YOLOv8 architecture and incorporates several enhancements to improve both detection accuracy and computational efficiency. First, in the image preprocessing stage, we employ a Super-Resolution Generative Adversarial Network (SRGAN) (21) to enhance the clarity of lung CT images, thereby making the nodule regions more prominent and facilitating more accurate detection (22). Additionally, we integrate the BiFormer attention mechanism (23), which combines transformer and feature pyramid networks, enabling the model to capture both global and local feature information of pulmonary nodules more effectively. To further optimize computational efficiency, we utilize the Content-Aware ReAssembly of Features (CARAFE) up-sampling method (24), which reduces the number of model parameters while preserving critical image details. Furthermore, we introduce the Depth-wise Over-Parameterized Depth-wise Convolution (DO-DConv) layer (25), an extension of depthwise convolutions, to enhance the model's expressive capability without increasing the computational complexity during inference. Rather than introducing entirely new components, this study focuses on the task-specific integration and clinical optimization of these existing modules. The incorporation of BiFormer,

CARAFE, DO-DConv, and SRGAN into the YOLOv8 framework is carefully designed to address the unique challenges of pulmonary nodule detection, such as small nodule size, low contrast, and boundary ambiguity. To the best of our knowledge, this is the first work to propose a unified, clinically oriented framework that integrates these modules for CT-based lung nodule detection. This design enables improved detection performance under real-world imaging conditions, balancing accuracy, robustness, and real-time inference capability. Through extensive experimental evaluation, we demonstrate that YOLOv8-BCD achieves high-precision pulmonary nodule detection with a lightweight architecture, making it well-suited for deployment in resource-constrained environments and advancing the application of artificial intelligence (AI)-based diagnostic tools in clinical practice.

Methods

Datasets and evaluation strategy

Two CT scan datasets were utilized in this study. The LUNG Nodule Analysis 2016 dataset (LUNA16 dataset) (<https://luna16.grand-challenge.org/>), a subset of the publicly available LIDC-IDRI database, was used for ablation and comparative experiments. LUNA16 was specifically created for lung nodule detection and analysis, comprising CT scans from approximately 1,000 patients and thousands of annotated pulmonary nodules. To ensure diversity and representativeness, we selected 550 CT images containing nodules of varying sizes across different anatomical regions for analysis. These images enabled a systematic evaluation of individual module contributions within the YOLOv8-BCD framework and a comprehensive assessment of detection performance.

To further evaluate the model's generalization beyond the LUNA16 dataset, the TianChi lung nodule competition dataset (<https://tianchi.aliyun.com/competition/entrance/231601/information>) was introduced as an independent external validation set. The TianChi dataset includes 1,971 annotated CT scans collected from multiple hospitals across China, providing greater diversity in nodule characteristics, imaging protocols, and patient populations. This dual-dataset evaluation strategy enables a robust assessment of the model's performance under varied clinical conditions, including differences in image quality and acquisition parameters. This study was conducted in accordance with the Declaration of Helsinki and its subsequent amendments.

CT image augmentation

To improve the detection accuracy of pulmonary nodules in lung CT images, we introduce the SRGAN model for data augmentation (21). SRGAN is an advanced image super-resolution technique capable of generating high-resolution, high-quality images from low-resolution inputs. The SRGAN model consists of two networks: the generator and the discriminator. The generator creates high-resolution images from low-resolution ones, while the discriminator distinguishes between the generated images and real high-resolution images. Through adversarial training, the generator learns the feature distribution of lung CT images and produces high-resolution, high-quality enhanced images. These enhanced images offer clearer details of the nodules and more detailed texture information, facilitating better feature extraction and recognition of nodules by the YOLOv8-BCD model. The effects of data augmentation are illustrated using images from four randomly selected patients (Figure S1).

Image labelling

In this study, LabelImg is utilized to manually annotate the locations of lung nodules in the selected images from the LUNA16 dataset both before and after SRGAN enhancement, with the annotations stored in YOLO format. Rectangular bounding boxes are drawn to completely enclose the outline of each nodule. If a nodule is located at the edge of the lung, the bounding box is drawn based on expert judgment. During the annotation process, pulmonary blood vessels and nodules that are difficult to distinguish are excluded from the annotations to ensure the accuracy and reliability of the lung nodule detection. The dataset contains 550 images, which are randomly divided into training and validation sets with a 4:1 ratio, resulting in 440 images for training and 110 images for validation.

BiFormer

BiFormer, a transformer-based variant, incorporates the Bidirectional Routing Attention (BRA) module to efficiently implement self-attention calculations. The BRA module consists of three main components: region partitioning and input projection, region-to-region routing with a directed graph, and token-to-token self-attention (23).

The attention is given in Eq. [1]:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{N}}\right)V \quad [1]$$

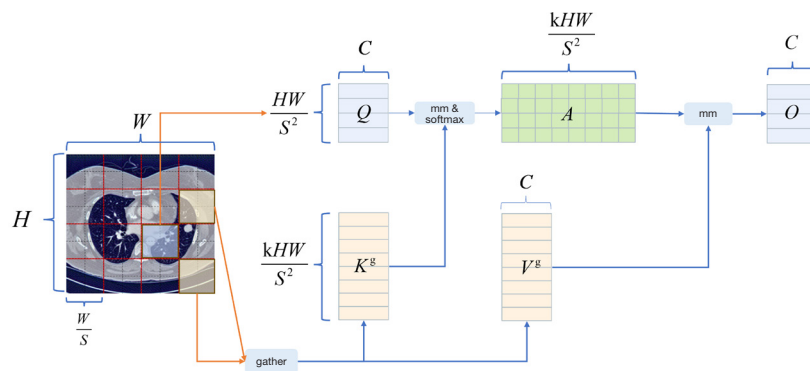


Figure 1 Schematic illustration of the BiFormer attention mechanism.

where query $Q \in \mathbb{R}^{N_q \times C}$, key $K \in \mathbb{R}^{N_k \times C}$ and value vector $V \in \mathbb{R}^{N_v \times C}$. Softmax maps inputs into a (0, 1) range, and $\sqrt{N}$ is a scalar representing the dimension of key vectors. In the region partitioning and input projection component, a two-dimensional (2D) input feature map $X \in \mathbb{R}^{H \times W \times C}$, where H , W and C denote height, width, and channels, respectively, is partitioned into $S \times S$ non-overlapping regions, each containing $\frac{HW}{S^2}$ feature vectors and X is reshaped to $X^r \in \mathbb{R}^{\frac{S^2 \times HW}{S^2} \times C}$. The reshaped feature map yields query, key, and value vectors according to Eq. [2]:

$$Q = X^r W^q, K = X^r W^k, V = X^r W^v \quad [2]$$

where $Q, K, V \in \mathbb{R}^{\frac{S^2 \times HW}{S^2} \times C}$; W^q, W^k and $W^v \in \mathbb{R}^{C \times C}$ are the weights of each linear projection. The second component, region-to-region routing with a directed graph, computes region relationships using a weighted directed graph derived from the input feature map. The average values of Q and K within nodule partitions, denoted as Q^r and $k^r \in \mathbb{R}^{S^2 \times C}$, respectively, are utilized to compute a semantic similarity adjacency matrix as $A^r = Q^r (k^r)^T$. To reduce computational redundancy, BRA retains indices of regions with the highest relevance $I^r = \text{topkIndex}(A^r)$. In the third component, token-to-token attention integrates keys and queries with tokens selected from the index matrix for efficient GPU-based computation, $K^g = g(K, I^r), V^g = g(V, I^r)$ where $g(\cdot)$ is the function used to select the most relevant token from the index matrix. The final output is obtained by $O = \text{Attention}(Q, K^g, V^g) + \text{LCE}(V)$ where local contrast enhancement (LCE) refers to the local enhancement operation on multi-scale token aggregation performed by the deep convolutional network.

In pulmonary nodule detection, nodules often exhibit characteristics such as low resolution, blurry local regions, and small-scale structures, which present significant challenges for traditional object detection algorithms (26).

Furthermore, due to the high variability in shape and size, detection models are easily influenced by noise, resulting in misclassification or missed detections. To address these challenges and improve detection accuracy in complex scenarios, this study integrates the BiFormer attention mechanism into the model. BiFormer enhances the model's capability to focus adaptively on detailed texture features and small-scale structures, effectively reducing false positives and improving detection robustness (27). A schematic illustration of the BiFormer attention mechanism is presented in Figure 1.

DO-DConv

In object detection and image segmentation tasks, traditional depth-wise convolution is commonly used in lightweight networks due to its computational efficiency (28). However, standard depth-wise convolution is limited in its ability to capture complex structural information because of the relatively small number of parameters, resulting in reduced feature representation capacity (29). To overcome this limitation, we introduce DO-DConv into the YOLOv8-BCD framework. The core concept of DO-DConv is to enhance the expressiveness of depth-wise convolution by introducing additional learnable parameters through an over-parameterization strategy, thereby significantly improving the feature extraction capability without increasing computational complexity during inference (25). While preserving the original sampling positions and receptive fields of standard depth-wise convolution kernels, DO-DConv's optimized parameter structure enables the model to learn richer feature representations, which is particularly advantageous in challenging object detection scenarios.

The architecture of DO-DConv is conceptualized as a sequence of two depth-wise convolution operations. Initially, the input feature map undergoes a convolution with the first depth-wise convolution kernel to produce an intermediate feature map. This intermediate map is then processed with a second depth-wise convolution kernel to generate the final output feature map. The spatial dimensions of the convolution kernel ($M \times N$) and the number of input channels (C_{in}) determine the transformation from the input feature map ($\mathbb{P} \in \mathbb{R}^{(M \times N) \times C_{in}}$) to the output of DO-DConv ($\mathbb{O} \in \mathbb{R}^{C_{out}}$). This transformation can be computed using two mathematically equivalent methods. The first computational approach, known as feature combination (feature composition), is represented by $\mathbb{O} = \mathbb{W} \circ (\mathbb{D} \circ \mathbb{P})$ where $\circ$ represents the depth-wise convolutions. $\mathbb{D} \in \mathbb{R}^{(M \times N) \times D_{mul} \times C_{in}}$ and $\mathbb{W} \in \mathbb{R}^{C_{out} \times D_{mul} \times C_{in}}$ represent the operations using the kernels of the first and second depthwise convolutions, respectively, and D_{mul} and C_{out} are the depth multiplier and number of output channels. The second method, referred to as kernel combination (kernel composition), involves combining the kernels before the convolution process, as illustrated by $\mathbb{O} = (\mathbb{D}^T \circ \mathbb{W}^T)^T \circ \mathbb{P}$ where $\mathbb{D}^T$ and $\mathbb{W}^T$ are the transpose of $\mathbb{D}$ and $\mathbb{W}$.

CARAFE

CARAFE is a content-aware feature reassembly operator that adaptively predicts and applies convolution kernels to local regions, thereby preserving critical details during up-sampling (24). This is especially beneficial in lung nodule detection, where nodules tend to be small, blurry, or located in low-resolution areas. By adaptively adjusting the reconstruction process based on image content, CARAFE more accurately recovers nodule details, even under challenging conditions such as small nodules or poor image quality, thereby improving detection accuracy (30). Furthermore, CARAFE effectively fuses multi-scale information, which is crucial for capturing the varied morphological characteristics of lung nodules across different scales. Compared to traditional up-sampling methods, CARAFE not only reduces information loss but also enhances positioning accuracy and robustness, making it particularly advantageous in complex backgrounds (31).

Assuming that the input feature map $\mathcal{X}$ has the number of channels C and the spatial size $H \times W$, and the up-sampling factor is set to σ , the size of the output feature map $\mathcal{X}'$ is $C \times \sigma H \times \sigma W$. For any target position $l' = (i', j')$ in the output feature map, its corresponding position in the input

feature map is $l = (i, j)$, where $i = \lfloor i' / \sigma \rfloor$, $j = \lfloor j' / \sigma \rfloor$. Define $\mathcal{X}_l$ as the $k \times k$ local area of the input feature map $\mathcal{X}$ at position l , that is, the neighborhood of this position is $N(\mathcal{X}_l, k)$.

The whole process can be divided into two main steps: kernel prediction and content-aware reassembly. The convolution kernel prediction module of CARAFE is responsible for generating content-based reorganized convolution kernels. For each position l on the input feature map, there are σ^2 target positions on the output feature map. Each target position requires a $k_{up} \times k_{up}$ reorganized convolution kernel. The convolution kernel size output by this module is $C_{up} \times H \times W$, $C_{up} = \sigma^2 k_{up}^2$. The convolution kernel prediction module consists of three submodules. The first submodule is the channel compressor, which uses 1×1 convolution to reduce the dimension of the input feature channel from C to C_m , reducing computational overhead and improving the efficiency of CARAFE. The second submodule is the content encoder. A convolution operation with a kernel size of $k_{encoder}$ is used to generate a content-based reorganized convolution kernel, whose parameters are $k_{encoder} \times k_{encoder} \times C_m \times C_{up}$. To strike a balance between computational complexity and performance, the empirical formula sets $k_{encoder} = k_{up} - 2$. The third submodule is the kernel normalizer. Through the Softmax normalization operation, the sum of the weights of the convolution kernels in the local area is made 1, ensuring that the mean of the feature map remains unchanged. The specific calculation formula is $W_l = \psi(N(\mathcal{X}_l, k_{encoder}))$. After predicting the reorganized convolution kernel, CARAFE reorganizes the feature map. The reorganization process uses a weighted summation form $\mathcal{X}'_l = \sum_{n=-r}^r \sum_{m=-r}^r W_{l'(n,m)} \cdot \mathcal{X}_{(i+n, j+m)}$ where $r = \lfloor k_{up} / 2 \rfloor$. This process ensures that the features of adjacent regions can be effectively aggregated, making the semantic expression of the up sampled features stronger. The computational complexity of CARAFE mainly depends on the convolution kernel sizes $k_{encoder}$ and k_{up} . For the content encoder, its computational complexity is approximately $\mathcal{O}(H \times W \times k_{encoder}^2 \times C_m \times C_{up})$, while the computational complexity of the content-aware reconstruction module is $\mathcal{O}(H \times W \times k_{up}^2 \times C)$. Since $k_{encoder} = k_{up} - 2$, a good balance between computational complexity and effect can be achieved by choosing a suitable k_{up} . The overall framework of CARAFE is illustrated in Figure 2.

The overview of YOLOv8-BCD

The YOLOv8 algorithms have demonstrated remarkable effectiveness in object detection tasks, gaining widespread

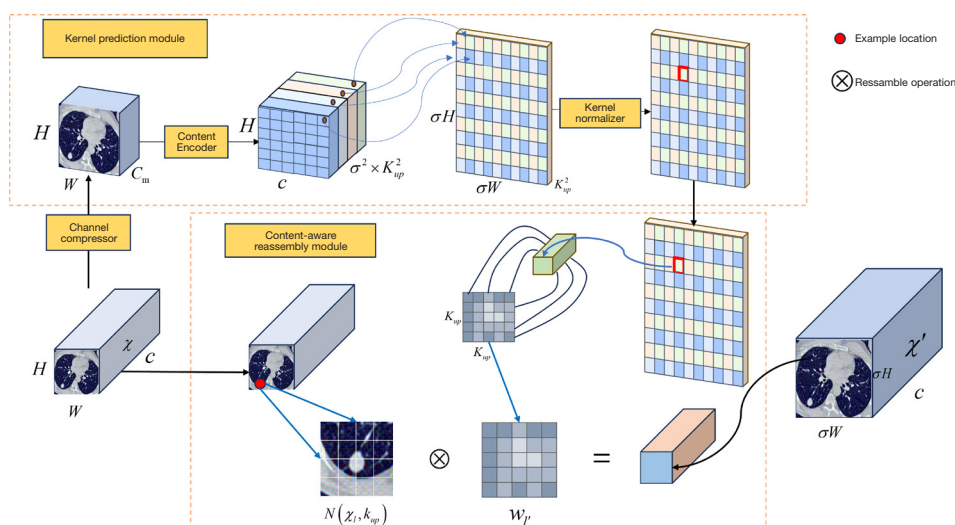


Figure 2 The overall framework of CARAFE. CARAFE, Content-Aware ReAssembly of Features.

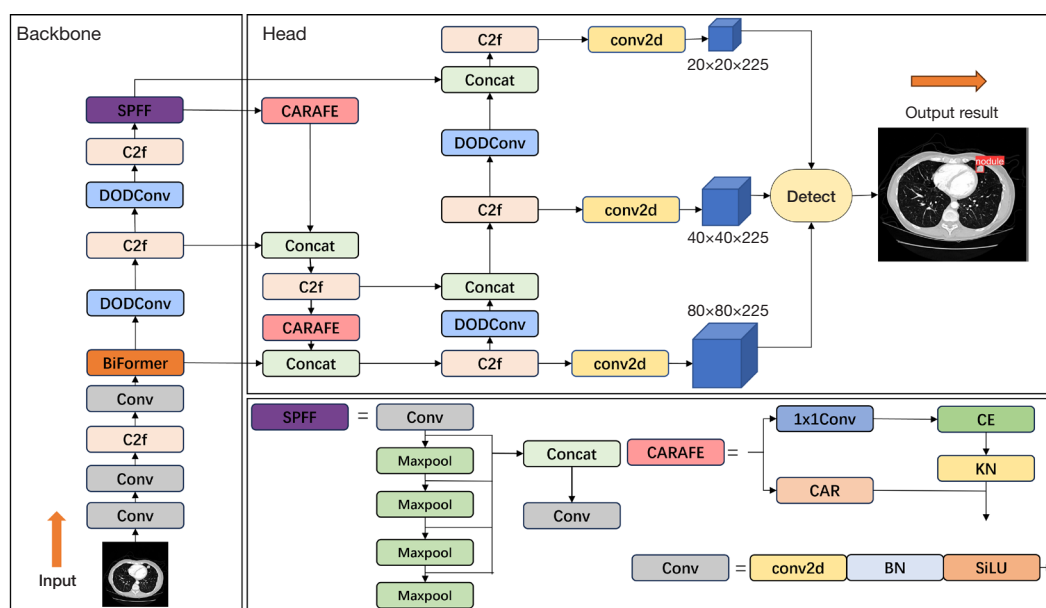


Figure 3 The overall architecture of the YOLOv8-BCD model.

adoption across various domains due to their excellent accuracy and computational efficiency (32). To further improve YOLOv8 performance for pulmonary nodule detection in lung CT imaging, we propose an optimized architecture, termed YOLOv8-BCD. This enhanced model integrates several key modules designed specifically to address the challenges associated with lung nodule detection. Briefly, this study introduces targeted enhancements to the YOLOv8 architecture by integrating the BiFormer

and DO-DConv modules into the backbone, and the CARAFE module into the head. The BiFormer module employs bidirectional self-attention to capture global and local features effectively, while DO-DConv enhances detailed feature extraction without significantly increasing computational load. The CARAFE module improves feature fusion and small nodule detection by optimizing the up-sampling process in the head. The detailed architecture of YOLOv8-BCD is presented in *Figure 3*.

In the challenging task of lung nodule detection, the varied morphology of nodules and their often-blurred boundaries, similar in appearance to adjacent lung tissues like blood vessels, significantly complicate detection efforts. DO-DConv enhances the model's ability to model local details by improving the depth convolution's feature learning capacity, thereby facilitating the accurate capture of essential features of lung nodules, particularly in low contrast or complex background settings. This capability makes DO-DConv particularly effective in enhancing detection accuracy. Integrated into the YOLOv8 detection framework, DO-DConv significantly boosts the model's local feature extraction prowess, thereby enhancing detection precision while maintaining fast inference speeds. Specifically, DO-DConv refines the feature representation capabilities of depth-wise convolutions within the YOLOv8 backbone, enabling the network to discern finer local details across various scales and thus improving the detection outcomes for small nodules. Importantly, the parameter optimization strategy of DO-DConv achieves these enhancements without imposing additional computational loads, allowing the model to sustain high-efficiency inference while achieving better accuracy. In lung nodule detection, DO-DConv particularly aids in accurately identifying small and variably shaped nodules in CT images. By optimizing local feature extraction, DO-DConv also improves the model's mean average precision (mAP) and overall detection accuracy, demonstrating its value in enhancing diagnostic imaging tools.

Evaluation criteria

Average precision (AP), defined as the area under the precision-recall (PR) curve, is a key metric for evaluating model performance. In multi-class classification tasks, the mAP, which is the average AP across all classes, serves as a commonly used benchmark. In this study, $\text{mAP}_{0.5}$ is employed as the primary evaluation indicator, reflecting the model's detection performance when the intersection over union (IoU) threshold is 0.5.

Beyond accuracy metrics, this study further evaluates the model's computational complexity and structural efficiency. Specifically, the number of parameters indicates the model's capacity, billion floating-point operations (GFLOPs) represent the computational cost, the weight file size reflects storage requirements, and the layer count denotes the network depth. Notably, convolutional computational complexity largely depends on kernel size, the number

of output channels, and the spatial dimensions of the output feature map. The formulas for calculating model parameters and GFLOPs are $\text{Params} = [i \times (k \times k) \times C_{\text{out}}] + C_{\text{out}}$, $\text{GFLOPs} = H \times W \times \text{Params}$, respectively. The mAP at a 50% IoU threshold served as the primary performance metric.

Ablation experiment design and implementation

The experiments systematically evaluate the contributions of three components—BiFormer, CARAFE, and DO-DConv by testing their individual and combined effects on YOLOv8. Starting with the baseline YOLOv8 (no components), single-component variants (Ablation 1–3) isolate the impact of each module, while pairwise combinations (Ablation 4–6) assess synergistic interactions (e.g., BiFormer + CARAFE). The final model YOLOv8-BCD integrates all three components to validate their collective necessity. To further assess the impact of image quality on model behavior, the complete ablation study was conducted under two settings: with and without SRGAN-based image enhancement. Performance is measured using accuracy (mAP), speed [frames per second (FPS)], efficiency (parameters/GFLOPs), and clinical metrics (e.g., false positives rate).

All experiments were conducted on a Windows 11 operating system, utilizing an AMD Ryzen 7945HX processor (with integrated Radeon graphics) and an NVIDIA GeForce RTX 4060 laptop GPU. The experimental environment was built on Python 3.8 and PyTorch 2.1.0 (supporting CUDA 11.8). Model training spanned 150 epochs with a batch size of 4, using consistent hyperparameter settings across all experiments.

Results

The YOLOv8-BCD model exhibits outstanding performance in pulmonary nodule detection, as demonstrated through rigorous experimental analysis on the LUNA16 dataset. *Figure 4A* presents the confusion matrix of classification predictions, highlighting the model's high true positive rate (TPR) and low false positive and false negative rates, achieving an accuracy of 0.88. While occasional false negatives arise due to factors such as extremely small nodule sizes (<3 mm) or imaging noise, the model maintains remarkable robustness and precision in detecting pulmonary nodules. *Figure 4B* illustrates the PR curve across varying recall levels, with an $\text{mAP}_{0.5}$ of 0.883. The curve demonstrates that precision remains consistently

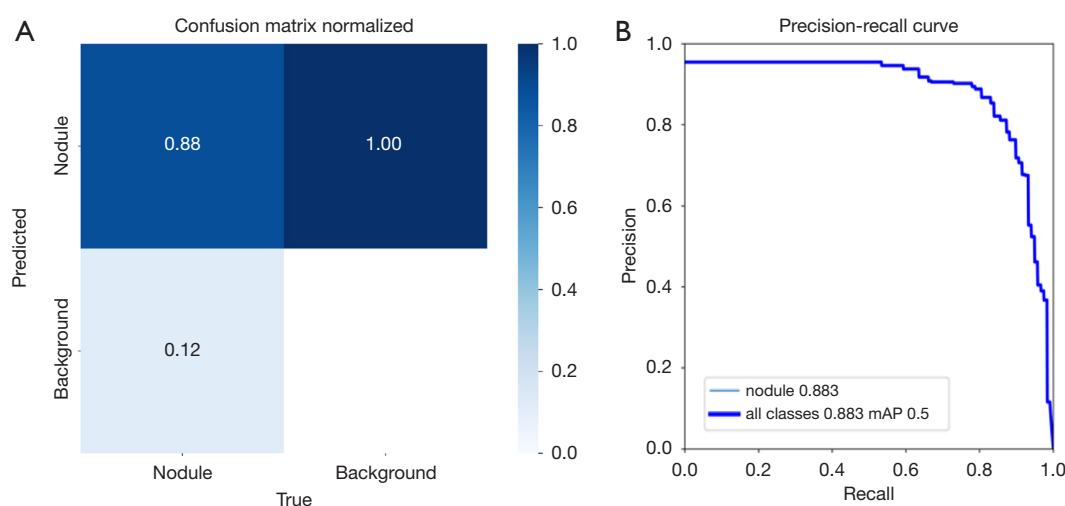


Figure 4 Performance analysis. (A) Confusion matrix: distribution of true/false positives and negatives. (B) Precision-recall curve: relationship between precision and recall across detection thresholds. mAP, measured using accuracy.

high even as recall increases, underscoring the model's ability to maintain diagnostic precision while effectively suppressing false positives.

Ablation experiment results

Ablation experiments were conducted to systematically evaluate the individual contributions of the BiFormer, CARAFE, and DO-DConv modules to the proposed model's performance. The ablation analysis was first conducted on CT images enhanced by SRGAN preprocessing. This approach allowed for a clearer evaluation of each module's contribution under improved image quality conditions. The ablation experiments indicate that BiFormer, CARAFE, and DO-DConv each contribute distinct advantages to the YOLOv8-BCD framework for pulmonary nodule detection. Integrating these three modules into YOLOv8-BCD yields the best overall performance across precision, recall, and mAP, demonstrating both the effectiveness and the soundness of this method in complex target detection tasks. *Table 1* summarizes the experimental results and compares accuracy, recall, $\text{mAP}_{0.5}$, $\text{mAP}_{0.5-0.95}$, parameter count, and computational complexity for different configurations.

When BiFormer alone was introduced to the baseline YOLOv8 model (Experiment 1), the recall rose from 79.7% to 85.4%, $\text{mAP}_{0.5}$ increased from 83.8% to 84.7%, and $\text{mAP}_{0.5-0.95}$ also showed an improvement, although accuracy declined slightly from 84.2% to 83.4%. The substantial gain in recall suggests that BiFormer's bidirectional self-

attention mechanism effectively captures global features, thereby reducing missed detections, whereas the slight decrease in accuracy may be attributed to the noise introduced by focusing on broader contextual information.

Replacing BiFormer with CARAFE upsampling (Experiment 2) produced an increase in recall to 85.7%, pushed $\text{mAP}_{0.5}$ to 85.2%, and maintained $\text{mAP}_{0.5-0.95}$ at 40.2%, while accuracy underwent a slight drop. The gains in recall and mAP can be attributed to CARAFE's dynamic feature reconstruction process, which aids in recovering details of small nodules. Its content-aware upsampling strategy mitigates the blurring and artifact problems commonly found in conventional upsampling approaches, leading to more accurate retention of local features.

When only DO-DConv was integrated into YOLOv8 (Experiment 3), accuracy improved markedly to 90.3%, while recall dipped marginally, with $\text{mAP}_{0.5}$ remaining at 83.8% and $\text{mAP}_{0.5-0.95}$ at 39.7%. This notable improvement in accuracy can be ascribed to DO-DConv's over-parameterized depthwise convolutions, enabling finer modeling of subtle texture features in nodules. However, the modest decline in recall indicates that DO-DConv may be less sensitive to nodules with indistinct boundaries, thereby causing occasional missed detections.

Combining BiFormer and CARAFE (Experiment 4) led to gains in both precision and recall, with $\text{mAP}_{0.5-0.95}$ reaching 42.6%. This outcome illustrates how BiFormer's global context modeling and CARAFE's content-aware upsampling complement each other. BiFormer strengthens

Table 1 YOLOv8-BCD ablation experiment results with or without SRGAN enhancement

Model	BiFormer	CARAFE	DO-DConv	Precision (%)	Recall (%)	mAP _{0.5} (%)	mAP _{0.5-0.95} (%)	Params (M)	GFLOPs
With SRGAN									
YOLOv8				84.2	79.7	83.8	39.7	3.01	8.1
Ablation1	√			83.4	85.4	84.7	40.6	11.36	32.3
Ablation2		√		80.2	85.7	85.2	40.2	3.14	8.6
Ablation3			√	90.3	75.4	83.8	39.2	9.07	26.7
Ablation4	√		√	84.2	81.1	84.9	42.6	9.70	29.5
Ablation5	√	√		85.4	79.4	85.5	39.0	11.83	34.2
Ablation6		√	√	77.5	83.1	86.2	39.7	9.54	28.5
YOLOv8-BCD	√	√	√	86.4	80.6	88.3	41.0	9.75	31.3
Without SRGAN									
YOLOv8				82.0	77.2	82.6	37.1	3.01	8.2
Ablation1	√			79.6	80.9	85.4	38.5	11.36	32.5
Ablation2		√		82.7	79.4	84.5	41.6	3.15	8.7
Ablation3			√	87.1	76.1	83.3	38.8	9.10	26.2
Ablation4	√		√	83.9	78.1	85.7	39.5	9.71	29.7
Ablation5	√	√		84.2	78.9	85.9	40.2	11.83	34.4
Ablation6		√	√	75.2	82.3	86.0	39.4	9.84	28.2
YOLOv8-BCD	√	√	√	85.4	80.6	87.1	40.4	9.76	31.5

mAP_{0.5} (%), mean average precision at IoU =0.5. mAP_{0.5-0.95} (%), mean average precision from IoU =0.5–0.95. CARAFE, Content-Aware ReAssembly of Features; DO-DConv, Depth-wise Over-Parameterized Depth-wise Convolution; GFLOPs, Giga Floating Point Operations; IoU, intersection over union; Params (M), parameters in millions; SRGAN, Super-Resolution Generative Adversarial Network; YOLO, You Only Look Once.

the model's capability to address complex or low-resolution targets, while CARAFE refines the recovery of fine-grained details, particularly in small nodule regions. Consequently, their synergy substantially bolsters multi-scale detection performance.

Adding DO-DConv in conjunction with BiFormer (Experiment 5) further increased accuracy to 85.4%, although recall decreased slightly and mAP_{0.5} remained at 85.5%. DO-DConv enhances local feature extraction and benefits from BiFormer's improved global perception. However, the additional parameters introduced during DO-DConv training may contribute to a slight reduction in recall. Nonetheless, the improved accuracy underscores DO-DConv's efficacy in detecting objects characterized by complex structures or diffuse boundaries.

When CARAFE was paired with DO-DConv (Experiment 6), accuracy declined to 77.5%, likely because the emphasis on local details from DO-DConv can render

the model more susceptible to background noise, while CARAFE's enhanced upsampling can also lead to additional false positives. Nonetheless, recall and mAP both improved, signifying better coverage of small nodules and more comprehensive detection of intricate edge features.

Finally, integrating all three modules (BiFormer, CARAFE, and DO-DConv) yielded the best overall performance (Experiment 7). The accuracy reached 86.4%, recall 80.6%, and mAP_{0.5} improved to 88.3%, with mAP_{0.5-0.95} at 41%. These results, achieved with minimal additional computational overhead, highlight the effectiveness of each component and illustrate their combined advantages. BiFormer furnishes superior global feature modeling, CARAFE preserves fine-grained detail through content-aware upsampling, and DO-DConv refines local feature learning by introducing over-parameterized depthwise convolution. By jointly leveraging these capabilities, YOLOv8-BCD delivers robust detection

Table 2 Comparative analysis of YOLOv8-BCD's detection accuracy and speed against other leading target detection methods on the LUNA16 dataset for the task of lung nodule detection

Model	Precision (%)	Recall (%)	mAP _{0.5} (%)	mAP _{0.5-0.95} (%)	Params (M)	GFLOPs	FPS
SSD	62.1	35.5	59.6	35.6	26.29	62.75	39.65
Faster R-CNN	30.7	28.7	34.6	30.1	137.10	370.21	12.13
DETR	83.2	44.3	82.0	42.8	36.76	114.24	22.61
PP-YOLOE	85.6	72.9	82.6	35.8	8.35	13.9	81.97
YOLOv8	84.2	79.7	83.8	39.7	3.01	8.1	94.34
YOLOv9	82.1	88.1	86.7	39.7	60.79	266.1	22.32
YOLOv10	83.1	74.8	81.8	37.6	2.69	8.2	79.37
YOLOv8-BCD	86.4	81.5	88.3	41.0	9.75	31.3	71.94

mAP_{0.5} (%), mean average precision at IoU =0.5. mAP_{0.5-0.95} (%), mean average precision from IoU =0.5–0.95. DETR, Detection TRansformer; Faster R-CNN, faster region-based convolutional neural network; FPS, frames per second; GFLOPs, Giga Floating Point Operations; IoU, intersection over union; Params (M), parameters in millions; LUNA16, LUNg Nodule Analysis 2016; SSD, single shot multibox detector; YOLO, You Only Look Once.

performance for small, morphologically diverse nodules and shows excellent potential for a wide range of automated detection tasks in medical image analysis.

To further assess the impact of SRGAN-based resolution enhancement on model performance, we conducted a comprehensive comparison between configurations with and without SRGAN preprocessing. As shown in *Table 1* (with SRGAN or without SRGAN), incorporating SRGAN led to notable improvements in detection precision and localization accuracy, particularly for small and low-contrast nodules. After adding SRGAN, YOLOv8-BCD's mAP_{0.5} increased from 87.1% to 88.3%, mAP_{0.5-0.95} improved from 40.4% to 41.0%, and precision rose from 85.4% to 86.4%. These gains indicate that the fine-detail enhancement provided by SRGAN improves the model's ability to identify nodule boundaries, thereby enhancing localization accuracy. Recall also showed a modest increase, suggesting that SRGAN primarily enhances discriminative power without substantially altering detection coverage.

The impact of SRGAN varied across modules. The BiFormer module demonstrated the greatest recall improvement (from 80.9% to 85.4%) and an mAP_{0.5-0.95} gain of 2.1%, indicating strong synergy between global feature modeling and high-resolution imagery. The CARAFE module also benefited, with recall rising from 79.4% to 85.7%, though precision slightly decreased, possibly due to overfitting to high-frequency background details. The DO-DConv module showed more modest gains, with a slight increase in precision and a minor decrease in recall,

suggesting less dependence on boundary clarity and potential misjudgment introduced by sharpening.

In summary, incorporating SRGAN enhances detection performance, especially for small and low-contrast nodules. When combined with BiFormer and CARAFE, the benefits are particularly pronounced, underscoring the complementary relationship between image enhancement and feature extraction. These findings highlight the important role of image preprocessing in improving the accuracy and robustness of medical image analysis workflows.

Performance comparison

The comparative analysis presented in *Table 2* is based on experiments conducted using the LUNA16 dataset and shows that YOLOv8-BCD achieves superior overall performance in lung nodule detection tasks compared to several advanced detection models. Specifically, YOLOv8-BCD exhibits the highest mAP at 0.5 IoU (mAP_{0.5} =88.3%) and a competitive precision (86.4%) and recall (80.6%). Additionally, it achieves the highest mAP across the stricter IoU range (0.5–0.9) at 41.0%, surpassing all other evaluated methods. Regarding efficiency, YOLOv8-BCD maintains a balanced computational cost with moderate model parameters (9.75 M) and GFLOPs (31.3), resulting in a solid inference speed (71.94 FPS). Although its FPS is slightly lower than YOLOv8 (94.34 FPS) and PP-YOLOE (81.97 FPS), YOLOv8-BCD significantly outperforms both in terms of detection accuracy metrics, making it a superior



Figure 5 Lung nodule detection performed using eight different models: SSD, Faster R-CNN, DETR, PP-YOLOE, YOLOv8, YOLOv9, YOLOv10, and YOLOv8-BCD (columns 2 through 9). The first column represents the ground truth, with nodules labeled by an experienced radiologist. DETR, DEtection TRansformer; Faster R-CNN, faster region-based convolutional neural network; SSD, single shot multibox detector; YOLO, You Only Look Once.

choice for achieving high-accuracy, efficient lung nodule detection in clinical settings.

While YOLOv8-BCD's recall (80.6%) is slightly lower than that of YOLOv9 (88.1%), this is offset by its higher overall precision and mAP values. In clinical practice, excessive false positives resulting from overly high recall can overwhelm radiologists and compromise diagnostic reliability (33). Therefore, a balanced precision-recall trade-off is preferred. YOLOv8-BCD is designed to minimize missed detections of clinically relevant nodules while maintaining a low false positive rate, ensuring reliable and

efficient detection performance in real-world settings.

Additionally, despite its higher parameter count and computational complexity compared to the baseline YOLOv8, YOLOv8-BCD maintains real-time inference capability, achieving 71.94 FPS, which is well above the 30 FPS threshold commonly required for clinical diagnostic applications (34). This trade-off between accuracy and efficiency is acceptable in medical imaging, where diagnostic performance, reliability, and interpretability often outweigh the marginal increase in computational cost.

Figure 5 displays lung nodule detection results from eight

Table 3 Comparative analysis of YOLOv8-BCD's detection accuracy and speed against other leading target detection methods TianChi lung nodule dataset for the task of lung nodule detection

Model	Precision (%)	Recall (%)	mAP _{0.5} (%)	mAP _{0.5-0.95} (%)	Params (M)	GFLOPs	FPS
SSD	55.3	41.2	70.7	32.0	26.29	62.74	51.52
Faster R-CNN	41.7	34.1	32.5	21.2	137.10	370.21	13.08
DETR	56.2	40.4	64.1	30.9	36.76	114.24	24.19
PP-YOLOE	87.8	72.9	81.6	36.1	8.35	13.9	70.7
YOLOv8	84.9	76.4	80.3	41.7	3.15	8.2	132.7
YOLOv9	75.9	80.2	75.2	33.5	60.79	266.1	28.3
YOLOv10	78.8	77.5	79.6	41.5	2.71	8.4	119.4
YOLOv8-BCD	84.4	78.9	83.8	43.9	9.75	31.3	98

mAP_{0.5} (%), mean average precision at IoU =0.5. mAP_{0.5-0.95} (%), mean average precision from IoU =0.5–0.95. DETR, DEtection TRansformer; Faster R-CNN, Faster region-based convolutional neural network; IoU, intersection over union; Params (M), parameters in millions; YOLO, You Only Look Once.

randomly selected CT images (columns). We randomly selected 8 CT images from all 550 images. The first column shows nodules labeled by experienced radiologists, while the remaining columns display nodules detected by various deep learning methods. As illustrated in *Figure 5*, YOLOv8-BCD consistently identifies lung nodules across multiple slices, demonstrating robust performance in diverse anatomical contexts. In contrast, single shot multibox detector (SSD) failed to detect almost all nodules, and faster region-based convolutional neural network (Faster R-CNN) and DEtection TRansformer (DETR) tended to produce false-positive detections. PP-YOLOE, YOLOv8, and YOLOv10 exhibited comparable performance but occasionally missed smaller nodules. Notably, YOLOv8-BCD successfully detected all labeled nodules, particularly excelling at identifying small nodules (e.g., as shown in the second row). We have also developed an online Shiny application (App) that allows users to upload CT scans and obtain detection results. You can access the App at <https://yolov8-bcd.streamlit.app/>.

To assess performance in more diverse clinical scenarios, we conducted additional experiments using the publicly available TianChi dataset. This dataset, comprising 1,971 annotated CT scans, features significant variability in imaging protocols, patient demographics, and nodule types, making it a valuable benchmark for external validation. As shown in *Table 3*, YOLOv8-BCD achieved the highest detection accuracy among all evaluated models, with an mAP_{0.5} of 83.8% and an mAP_{0.5-0.95} of 43.9%, demonstrating strong capability in localizing and recognizing pulmonary

nodules across varying IoU thresholds. The model also achieved a precision of 84.4% and a recall of 78.9%, indicating robust detection of true nodules while effectively suppressing false positives.

YOLOv8-BCD maintained high computational efficiency, achieving 98 FPS despite its increased complexity (9.75 M parameters, 31.3 GFLOPs), meeting clinical real-time requirements. Notably, it performed well even on challenging cases in the TianChi dataset, such as small nodules, low-contrast lesions, and nodules overlapping vessels. In comparison, YOLOv8 offered faster inference but lower detection accuracy and recall, particularly for small and low-contrast nodules. YOLOv9 showed relatively high recall (80.2%) but at the cost of significantly greater model complexity and lower inference speed (28.3 FPS). YOLOv10 achieved high speed but lagged behind YOLOv8-BCD in detection accuracy. Overall, the TianChi results demonstrate that YOLOv8-BCD balances detection accuracy and inference speed, with strong generalization to complex, clinically relevant datasets.

K-fold cross-validation experiments

To further validate model robustness, we conducted 5-fold cross-validation on the LUNA16 dataset (550 annotated CT images). This approach allowed for the assessment of performance stability and generalization across different data partitions. As shown in *Table 4*, YOLOv8-BCD demonstrated consistent and reliable results, with minimal variation across folds. The model maintained high precision

Table 4 K-fold cross-validation results on the LUNA16 dataset

K-Fold	Precision (%)	Recall (%)	mAP _{0.5} (%)	mAP _{0.5-0.95} (%)
Fold-1	80.1	85.1	84.6	42.8
Fold-2	90.7	84.7	93.7	49.4
Fold-3	80.9	87.1	87.5	45.3
Fold-4	89.1	87.0	95.8	51.7
Fold-5	87.1	91.5	94.1	53.6

mAP_{0.5} (%), mean average precision at IoU =0.5. mAP_{0.5-0.95} (%), mean average precision from IoU =0.5–0.95. IoU, intersection over union; LUNA16, LUNG Nodule Analysis 2016.

and recall across all folds, underscoring its stability and insensitivity to random data splits. Specifically, mAP_{0.5} ranged from 84.6% to 95.8%, and mAP_{0.5-0.95} ranged from 42.8% to 53.6%, reflecting strong detection performance across different nodule types, sizes, and contrasts. Notably, recall remained above 84% in all folds, highlighting the model's ability to detect true positive nodules, which is essential for early-stage lung cancer screening.

Discussion

In this paper, we present an enhanced YOLOv8-BCD model tailored for lung nodule detection, which builds upon the original YOLOv8 by integrating the BiFormer attention mechanism, CARAFE upsampling, and DO-DConv advanced convolution techniques. These modifications aim to bolster feature extraction, improve local information recovery, and increase the number of learnable parameters, thereby boosting both the detection accuracy and robustness of the model. Additionally, the use of SRGAN for image enhancement significantly refines the details and contrast in lung CT images, enhancing nodule detection performance.

Key findings from this research include: (I) YOLOv8-BCD demonstrated significant improvements across key evaluation metrics. On the LUNA16 dataset, it achieved a detection accuracy of 86.4%, a recall of 80.6%, an mAP_{0.5} of 88.3%, and an mAP_{0.5-0.95} of 41.0%. On the more diverse TianChi dataset, YOLOv8-BCD achieved an mAP_{0.5-0.95} of 83.8% and an mAP_{0.5-0.95} of 43.9%. These results outperform baseline models such as YOLOv8, YOLOv9, and YOLOv10, validating the effectiveness of the proposed architectural enhancements in improving both accuracy and generalizability. (II) Ablation studies confirm the synergistic effects of the BiFormer attention mechanism, CARAFE

upsampling, and DO-DConv convolution module, with the combined use of these features allowing the model to excel in terms of precision, recall, and mAP. (III) Comparative analysis with other models reveals that YOLOv8-BCD not only excels in precision and recall but also maintains a reasonable computational complexity. It fulfills real-time detection requirements while ensuring high performance, making it a superior option for clinical applications.

The image detail enhancement provided by SRGAN preprocessing, the global feature modeling capability of BiFormer, the fine-grained structural reconstruction of CARAFE, and the enriched convolutional representation from DO-DConv collectively led to significant improvements in both detection accuracy and computational efficiency, without introducing substantial computational overhead. Experimental results show that, compared to the original YOLOv8, the YOLOv8-BCD model achieved a 4.5% increase in mAP_{0.5} on the LUNA16 dataset. Additionally, it attained an mAP_{0.5} of 83.8% and an mAP_{0.5-0.95} of 43.9% on the TianChi dataset.

The YOLOv8-BCD model offers high accuracy, efficiency, and cost-effectiveness, effectively aiding doctors in the preliminary evaluation of lung CT nodules. This can enhance detection efficiency, improve diagnostic accuracy, reduce the risk of missed and misdiagnoses, and provide robust support for early lung cancer screening and diagnosis. Its deployment in medical facilities can enhance the efficiency of lung nodule detection and reduce labor costs. Real-time model deployment can expedite the diagnostic process and enhance accuracy, offering doctors valuable tools and enhancing patient care. As the model's performance continues to improve, it is poised to have broad clinical application prospects and contribute significantly to the fields of medical image analysis and precision medicine.

Conclusions

The YOLOv8-BCD model demonstrates high detection accuracy and robust performance in identifying pulmonary nodules in CT images, outperforming baseline models including YOLOv8, YOLOv9, and YOLOv10. The integration of BiFormer attention, CARAFE upsampling, and DO-DConv enhances feature representation and refines the extraction of localized structural information essential for small nodule detection, while SRGAN-based preprocessing improves image quality and nodule visibility. Notably, YOLOv8-BCD achieves superior mAP scores on both LUNA16 and TianChi datasets, validating its

generalizability across diverse clinical imaging data. The model maintains real-time detection capability with limited computational overhead, offering practical value in clinical workflows. These findings support the use of YOLOv8-BCD as an efficient and accurate tool for early lung cancer screening and computer-aided diagnosis in clinical practice.

Acknowledgments

None.

Footnote

Funding: This work was partially supported by Natural Science Research Project of Anhui Educational Committee (No. 2024AH050693), the Population Health and Eugenics Anhui Provincial Key Laboratory Project (No. JKYS20233) and the National Natural Science Foundation of China (No. 82073578).

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://qims.amegroups.com/article/view/10.21037/qims-2025-824/coif>). X.S.X. reports being a full-time employee of Genmab Inc., a for-profit biotechnology company. He receives no funding or support from the company for the present study. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study was conducted in accordance with the Declaration of Helsinki and its subsequent amendments.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Smeltzer M, Liao W, Faris NR, Fehnel C, Goss J, Shepherd CJ, Qureshi T, Matthews AT, Ray M, Osarogiagbon RU. Early detection of lung cancer through programmatic management of pulmonary nodules in the Mississippi Delta. *JCO Oncology Practice* 2023;19:126.
2. Prosper AE, Kammer MN, Maldonado F, Aberle DR, Hsu W. Expanding Role of Advanced Image Analysis in CT-detected Indeterminate Pulmonary Nodules and Early Lung Cancer Characterization. *Radiology* 2023;309:e222904.
3. Mazzone PJ, Lam L. Evaluating the Patient With a Pulmonary Nodule: A Review. *JAMA* 2022;327:264-73.
4. Adams SJ, Stone E, Baldwin DR, Vliegenthart R, Lee P, Fintelmann FJ. Lung cancer screening. *Lancet* 2023;401:390-408.
5. Kumar A, Choudhry MS. Optimum thresholding for nodule segmentation of lung CT images. 2022 IEEE World Conference on Applied Intelligence and Computing (AIC). IEEE; 2022:272-6. Available online: <https://ieeexplore.ieee.org/abstract/document/9848878/>
6. Halder A, Chatterjee S, Dey D. Adaptive morphology aided 2-pathway convolutional neural network for lung nodule classification. *Biomedical Signal Processing and Control* 2022;72:103347.
7. Ma X, Song H, Jia X, Wang Z. An improved V-Net lung nodule segmentation model based on pixel threshold separation and attention mechanism. *Sci Rep* 2024;14:4743.
8. Mhaske MM, Manza RR, Pradhan PK. Detection of Lung Cancer Using SVM with Lung Nodule Segmentation. *PENSEE International Journal* 2021;7:23-9.
9. Saikia T, Hansdah M, Singh KK, Bajpai MK. Classification of lung nodules based on transfer learning with K-Nearest Neighbor (KNN). 2022 IEEE international conference on imaging systems and techniques (IST). IEEE; 2022:1-6. Available online: <https://ieeexplore.ieee.org/abstract/document/9827668/>
10. Bhatt SD, Soni HB, Kher HR, Pawar TD. Automated system for lung nodule classification based on resnet50 and svm. 2022 3rd International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT). IEEE; 2022:1-5. Available online: <https://ieeexplore.ieee.org/abstract/document/10064515/>
11. Sharifani K, Amini M. Machine learning and deep learning: A review of methods and applications. *World Information Technology and Engineering Journal* 2023;10:3897-904.
12. Yar H, Abbas N, Sadad T, Iqbal S. Lung nodule detection

- and classification using 2D and 3D convolution neural networks (CNNs). In: Goyal LM, Saba T, Krishnan A, Larabi-Marie-Sainte S, editors. Artificial Intelligence and Internet of Things 2021;365-86.
13. Zhao X, Li J, Qi M, Chen X, Chen W, Li Y, Liu Q, Tang J, Han Z, Zhang C. MSTD: A Multi-scale Transformer-based Method to Diagnose Benign and Malignant Lung Nodules. *IEEE Access* 2025. Available online: <https://ieeexplore.ieee.org/abstract/document/10844082/>
 14. Fauzya SP, Ardiyanto I, Nugroho HA. A Comparative Study on Lung Nodule Detection: 3D CNN vs Vision Transformer. 2024 8th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE). *IEEE*; 2024:417-22. Available online: <https://ieeexplore.ieee.org/abstract/document/10729900/>
 15. Heuvelmans MA, van Ooijen PMA, Ather S, Silva CF, Han D, Heussel CP, Hickes W, Kauczor HU, Novotny P, Peschl H, Rook M, Rubtsov R, von Stackelberg O, Tsakok MT, Arteta C, Declerck J, Kadir T, Pickup L, Gleeson F, Oudkerk M. Lung cancer prediction by Deep Learning to identify benign lung nodules. *Lung Cancer* 2021;154:1-4.
 16. Redmon J, Divvala S, Girshick R, Farhadi A. You Only Look Once: Unified, Real-Time Object Detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: *IEEE*; 2016:779-88. Available online: <http://ieeexplore.ieee.org/document/7780460/>
 17. Varghese R, Sambath M. Yolov8: A novel object detection algorithm with enhanced performance and robustness. 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS). *IEEE*; 2024:1-6. Available online: <https://ieeexplore.ieee.org/abstract/document/10533619/>
 18. Widayani A, Putra AM, Maghriebe AR, Adi DZC, Ridho MHF. Review of Application YOLOv8 in Medical Imaging. *Physics Letters* 2024. [cited 2025 Mar 27]. Available online: <https://e-journal.unair.ac.id/IAPL/article/download/57001/28746>
 19. Elnady N, Adel A, Badawy W. Harnessing YOLOv9 for Enhanced Detection of Lung Cancer: A Deep Learning Approach. 2024 Intelligent Methods, Systems, and Applications (IMSA). *IEEE*; 2024:518-3. Available online: <https://ieeexplore.ieee.org/abstract/document/10652879/>
 20. Wu X, Zhang H, Sun J, Wang S, Zhang Y. YOLO-MSRF for lung nodule detection. *Biomedical Signal Processing and Control* 2024;94:106318.
 21. Ledig C, Theis L, Huszar F, Caballero J, Cunningham A, Acosta A, Aitken A, Tejani A, Totz J, Wang Z, Shi W. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. *arXiv:1609.04802* [Preprint]. 2017 [cited 2025 Mar 27]. Available online: <http://arxiv.org/abs/1609.04802>
 22. Zhu H, Han G, Peng Y, Zhang W, Lin C, Zhao H. Functional-realistic CT image super-resolution for early-stage pulmonary nodule detection. *Future Generation Computer Systems* 2021;115:475-85.
 23. Zhu L, Wang X, Ke Z, Zhang W, Lau R. BiFormer: Vision Transformer with Bi-Level Routing Attention. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, BC, Canada: *IEEE*; 2023:10323-33. Available online: <https://ieeexplore.ieee.org/document/10203555/>
 24. Wang J, Chen K, Xu R, Liu Z, Loy CC, Lin D. CARAFE: Content-Aware ReAssembly of FEatures. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South): *IEEE*; 2019:3007-16. Available online: <https://ieeexplore.ieee.org/document/9010830/>
 25. Cao J, Li Y, Sun M, Chen Y, Lischinski D, Cohen-Or D, Chen B, Tu C. DO-Conv: Depthwise Over-Parameterized Convolutional Layer. *IEEE Trans Image Process* 2022;31:3726-36.
 26. Jiang B, Li N, Shi X, Zhang S, Li J, de Bock GH, Vliegenthart R, Xie X. Deep Learning Reconstruction Shows Better Lung Nodule Detection for Ultra-Low-Dose Chest CT. *Radiology* 2022;303:202-12.
 27. Zheng Y, Wang M, Zhang B, Shi X, Chang Q. GB-CD-YOLO: A high-precision and real-time lightweight model for wood defect detection. *IEEE Access* 2024;12:12853-68.
 28. Patel H, Prajapati K, Sarvaiya A, Upla K, Raja K, Ramachandra R, Busch C. Depthwise convolution for compact object detector in nighttime images. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. 2022:379-89. Available online: http://openaccess.thecvf.com/content/CVPR2022W/PBVS/html/Patel_Depthwise_Convolution_for_Compact_Object_Detector_in_Nighttime_Images_CVPRW_2022_paper.html
 29. Singh KK, Singh A. Diagnosis of COVID-19 from chest X-ray images using wavelets-based depthwise convolution network. *Big Data Mining and Analytics* 2021;4:84-93.
 30. Lv K, Wu R, Chen S, Lan P. CCI-YOLOv8n: Enhanced

- Fire Detection with CARAFE and Context-Guided Modules. arXiv:2411.11011 [Preprint]. 2024 [cited 2025 Mar 28]. Available online: <http://arxiv.org/abs/2411.11011>
31. Yuan M, Zhou Y, Ren X, Zhi H, Zhang J, Chen H. YOLO-HMC: An improved method for PCB surface defect detection. *IEEE Transactions on Instrumentation and Measurement* 2024;73:1-11.
 32. Diwan T, Anirudh G, Tembhurne JV. Object detection using YOLO: challenges, architectural successors, datasets and applications. *Multimed Tools Appl* 2023;82:9243-75.
 33. Hicks SA, Strümke I, Thambawita V, Hammou M, Riegler MA, Halvorsen P, Parasa S. On evaluation metrics for medical applications of artificial intelligence. *Sci Rep* 2022;12:5979.
 34. Lee H, Ko H, Chung H, Nam Y, Hong S, Lee J. Real-time realizable mobile imaging photoplethysmography. *Sci Rep* 2022;12:7141.

Cite this article as: Zhu W, Wang X, Xing J, Xu XS, Yuan M. YOLOv8-BCD: a real-time deep learning framework for pulmonary nodule detection in computed tomography imaging. *Quant Imaging Med Surg* 2025;15(9):8189-8204. doi: 10.21037/qims-2025-824