Optimized YOLOv8 Model for the Automatic Detection of Pulmonary Nodules

Zichao Liu, Lili Wei, Tingqiang Song

Abstract — **Pulmonary nodule detection faces significant challenges in balancing accuracy and efficiency, particularly when identifying subtle nodules of varying sizes within complex thoracic backgrounds. These limitations can adversely affect clinical decision-making, leading to missed or false detections. To address these issues, we propose an optimized YOLOv8-based model that incorporates four key advancements: 1) VanillaNet as a streamlined backbone to reduce computational complexity while maintaining robust feature extraction; 2) Dynamic Large Kernel Attention v4 (DLKAv4), a novel multi-scale attention mechanism that enhances sensitivity to faint nodule patterns; 3) Detect-DBB, a redesigned detection head that utilizes diverse branch convolutions for precise localization; and 4) a dual-loss framework that combines Quality Focal Loss (QFL) for the alignment of classification and localization tasks with Multi-scale Proportional Dynamic IoU (MPDIoU) for geometric optimization—effectively bridging the gap between detection confidence and localization accuracy. Comprehensive evaluation on the LUNA16 dataset demonstrated significant improvements: a 3.39% increase in precision, a 6.86% gain in recall, a 3.4% improvement in mAP50, and a 4.67% enhancement in mAP50-95, all achieved with a reduction in parameters. Importantly, validation on a self-constructed dataset confirmed the model's effectiveness in real-world scenarios, achieving reliable detection of sub-centimeter and ground-glass nodules under diverse imaging conditions. These advancements not only enhance computational efficiency but also address critical clinical applications by reducing diagnostic uncertainties, thereby offering a balanced solution for integration into time-sensitive clinical workflows. This optimized framework demonstrates substantial potential to support radiologists in the assessment of pulmonary nodules while maintaining practical computational demands.**

Index Terms—**Pulmonary nodule detection; Clinical decision-making; YOLOv8; Attention; Clinical applications**

I. Introduction

LUNG cancer is a leading cause of cancer-related deaths worldwide, with an increasing incidence rate. Early detection, accurate diagnosis, and timely treatment are critical for improving survival rates [1]. Manually detecting nodules in CT scans is both time-consuming and prone to errors, particularly when identifying smaller, early-stage nodules. Automated deep learning-based systems present a promising solution to this challenge. Recent advancements in artificial intelligence (AI) for chest imaging have significantly enhanced both the speed and accuracy of lung nodule detection, benefiting healthcare providers and patients alike [2-3]. Early detection is essential for improving patient prognosis, which has driven the development of numerous computer-aided diagnostic (CAD) tools designed to assist in the detection and classification of nodules [4]. Over the past two decades, research in image processing has made substantial progress in enhancing nodule detection, segmentation, and classification in CT scans [5]. AI has not only improved diagnostic efficiency and accuracy but also supported better patient management and treatment strategies [6]. Cutting-edge advancements in AI and machine learning (ML) demonstrate considerable potential for addressing the challenges of lung cancer screening and promoting health equity [7]. CAD systems help alleviate the workload for radiologists, and deep learning-based tools have shown high accuracy in detecting lung nodules without significantly increasing false-positive rates [8].

The YOLO object detection series is highly regarded for its real-time capabilities, effectively balancing accuracy with computational efficiency. YOLOv8 [9] enhances earlier models through architectural improvements and optimized loss functions, resulting in superior performance. However, detecting pulmonary nodules remains a challenge due to their small size, irregular shapes, and the complex tissues surrounding them, necessitating specialized optimizations for medical imaging. This research advances deep learning in medical image analysis by demonstrating how customized detection models can enhance both accuracy and efficiency in the early diagnosis of lung cancer, which is critical for reducing mortality rates. The primary contributions of this study are as follows:

(1) The original backbone was replaced with VanillaNet, significantly enhancing feature extraction while maintaining low computational complexity. This change improved both detection accuracy and processing efficiency for pulmonary nodules. Additionally, DLKAv4, which integrates DCNv4 with DLKA, was introduced to enhance the model's ability to capture fine-grained features and increase recall rates.

(2) Detect-DBB was implemented as the new detection head, optimizing feature representation and enhancing the differentiation of nodule types and sizes. QFL was employed

Manuscript received December 31, 2024; revised May 23, 2025.

This work was supported by the Chinese Nursing Association Research Program under Grant ZHKY202118 and Shandong Nursing Association Research Program under Grant SDHLKT202209.

Z. C. Liu is a postgraduate student of College of Information Science and Technology, Qingdao University of Science and Technology, Qingdao, 266061, China.(e-mail: 848811928@qq.com).

L. L. Wei is a president of Office of the Dean, The Affiliated Hospital of Oingdao University, Oingdao, 266000, China.(e-mail: 13573828157@163.com).

T. Q. Song is a vice-president of College of Information Science and Technology, Qingdao University of Science and Technology, Qingdao, 266061, China. (corresponding author: phone: 18661909298; e-mail: songtq@qust.edu.cn).

for classification, while MPDIoU was utilized for bounding box loss, resulting in improved localization accuracy.

(3) Extensive experiments conducted on the LUNA16 dataset, which included comparative and ablation studies, resulted in significant improvements: precision increased by 3.39%, recall by 6.86%, mAP50 by 3.4%, and mAP50-95 by 4.67%. Further validation on a self-constructed dataset confirmed the model's robustness and applicability across various scenarios, highlighting its innovative nature and practical impact.

II. Related research

Automated lung nodule diagnosis is critical in both research and clinical contexts; however, several challenges persist. Recent advances in object detection have led to effective solutions. For instance, Reference [10] improved detection accuracy over traditional Support Vector Machines (SVM) and K-Nearest Neighbors (KNN) by optimizing the activation functions and convolutional layers of Convolutional Neural Networks (CNN) using the Improved Moth Flame Optimization (IMFO) algorithm. Similarly, Reference [11] introduced the deep learning-based Synthetic Bone-Suppressed (DLBS) algorithm, which enables radiologists to detect nodules from X-rays without the need for additional equipment or radiation exposure. The Computer-Aided Diagnosis (CAD) systems discussed in References [12–14] employ methods such as the 3D Multi-Scale Vision Transformer (3D-MSViT), a 3D U-shaped CNN with channel attention, and a Bayesian-optimized Vision Transformer to enhance feature extraction and predictive accuracy. Reference [15] implements the Stochastic Gradient Descent with Adaptive Learning Rate (SGDA) module to improve network generalization, while Reference [16] utilizes 3D Faster R-CNN and 3D CNNs to enhance feature recognition and minimize false positives. In contrast, Reference [17] adopted an innovative approach using an all-optical Diffractive Deep Neural Network (D2NN) for CT scan detection. Reference [18] presented a hybrid Hierarchical Deep Ensemble Neural Network (HDE-NN) with Channel-Based Selective Optimization (CBSO) for tumor classification, whereas Reference [19] integrated spatiotemporal attention with an enhanced MobileNetV3 and GhostNet for nodule detection. Reference [20] introduced a 3D Multi-Attention Enhanced Detection (MAED) network with self-attention and region-attention modules to reduce false positives, and Reference [21] developed DBPNDNet, which focuses on automated nodule detection and segmentation.

AI-based systems have significantly improved detection rates; however, clinical validation remains essential [22]. Despite notable advancements in accuracy and sensitivity, some methods continue to face challenges with specific types of nodules, real-time processing, and extensive training requirements. YOLO models, such as BiRPN-YOLOvX and SwiF-YOLO, have enhanced diagnostic accuracy by utilizing Adaptive Spatial Feature Fusion (ASFF) and Swin transformers [23, 24]. YOLOv7 has improved the detection of small nodules through components like Small Object Detection Layer (SODL), Multi-Scale Residual Fusion (MSRF), and Enhanced Object Detection Convolution (EODConv) [25]. Reference [26] proposed a model that incorporates a deformable attention module to retain critical features and reduce noise, while the Weighted Intersection over Union (WIoU) loss function addresses the impact of low-quality samples on gradient calculations. The Lung-YOLO algorithm [27] employed a Multi-scale Dual-branch Attention (MSDA) mechanism and a Cross-layer Aggregation Module (CLAM) to enhance the detection of small targets. To overcome the limitations of YOLO-based systems, a modified version of YOLOv5 [28] was proposed for automatic lung nodule detection in CT scans. This version features improved feature extraction layers and customized anchor boxes for small nodules, highlighting the model's potential for efficient lung cancer screening. This aligns with the objective of this manuscript to advance YOLOv5-based detection techniques.

This study presents an optimized YOLOv8 model specifically designed for the automatic detection of pulmonary nodules. The modifications focus on enhancing feature extraction and classification accuracy while ensuring computational efficiency. When evaluated on the LUNA16 dataset, a standard benchmark for lung nodule detection, the proposed enhancements to YOLOv8 resulted in a significant improvement in detection accuracy, particularly for challenging cases involving small or irregular nodules.

III. Improved YOLOv8 network

Owing to its impressive speed and performance, YOLOv8 has become a pivotal tool in modern object detection, enhancing both network efficiency and inference, particularly in the early detection of lung nodules. However, it still faces challenges, such as detecting small objects and providing detailed classifications [29]. Furthermore, research specifically focused on lung nodule detection remains limited, despite its clinical significance. To address these limitations, we developed an optimized version of YOLOv8 specifically designed for lung nodule detection. Our model retains the speed of YOLOv8 while significantly improving detection accuracy, as demonstrated by our experimental results. This approach offers a faster and more precise method for early diagnosis and treatment. Fig. 1 illustrates the architecture of the enhanced YOLOv8.

A. VanillaNet for backbone improvements

VanillaNet [30] is an efficient neural network architecture designed for real-time detection, emphasizing both simplicity and performance. Initially, it enhances feature representation and nonlinearity by incorporating additional layers and complex structures, which are subsequently optimized through techniques such as layer merging. This optimization reduces computational demands while maintaining performance. This balance between complexity and efficiency facilitates faster and more accurate inference. To improve its nonlinear capabilities, VanillaNet employs parallel activation functions, allowing it to capture a broader range of features, particularly in the detection of small targets, such as lung nodules. As a result, it achieves higher accuracy compared to traditional networks that rely on a single activation function.

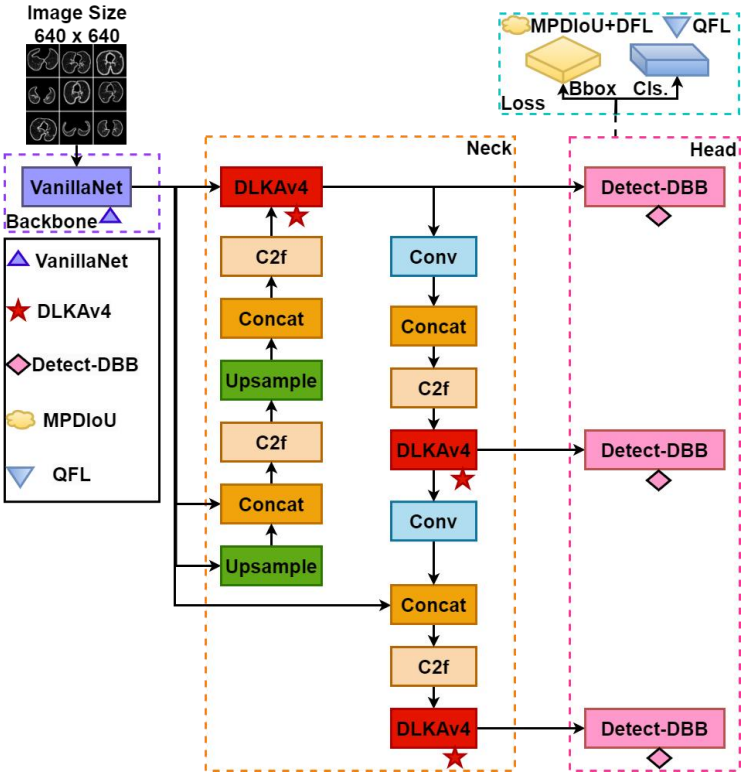


Fig. 1. Improved network structure

When integrated as the backbone of YOLOv8, VanillaNet optimizes the inference architecture without compromising detection accuracy, particularly in applications such as lung nodule detection. This integration not only accelerates inference but also enhances overall performance, making it highly effective in real-world scenarios.

B. The new attention mechanism DLKAv4

The DLKA framework [31] addresses the high computational cost associated with self-attention mechanisms by utilizing large convolutional kernels to model global receptive fields. This approach enhances feature extraction by incorporating extensive contextual information, thereby improving the model's ability to capture both large-scale and fine details. In the context of lung nodule detection, DLKA excels at identifying subtle features while considering the surrounding context due to its expanded receptive field. DCNv4 [32], an enhancement over DCNv3 [33], refines deformable convolution by eliminating softmax normalization in spatial aggregation. This modification allows for more flexible adaptation of sampling positions to image features, thereby improving memory efficiency and reducing computational redundancy. Consequently, it accelerates processing, particularly for high-resolution data, and enhances model convergence.

The DLKAv4 module integrates DLKA and DCNv4 into a cohesive attention mechanism. The large convolutional kernels in DLKAv4 extract comprehensive contextual information from lung images, providing global coverage similar to that of self-attention mechanisms. Meanwhile, the deformable convolution in DCNv4 enhances flexibility by adapting the sampling grid to the features of the image, which is crucial for accurate lung nodule detection. By combining these techniques, DLKAv4 minimizes computational overhead, accelerates convergence, and leverages the memory optimizations offered by DCNv4. This integration significantly improves efficiency in processing large-scale datasets, making DLKAv4 particularly effective for medical imaging tasks such as lung nodule detection. Fig. 2 illustrates the structure of DLKAv4.

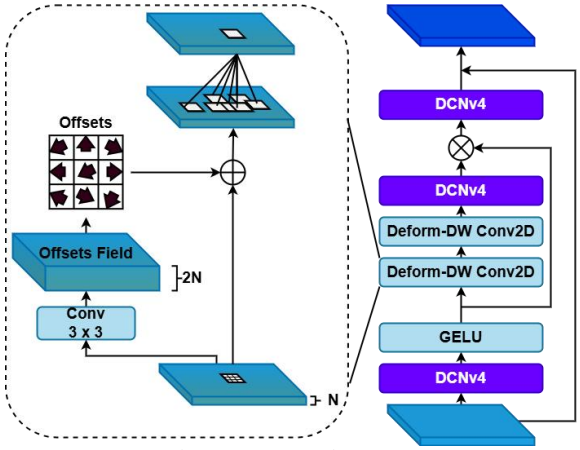


Fig. 2. Structure of DLKAv4

C. The novel head Detect-DBB

The DBB[34] notably increases the complexity of convolutional layers during training, thereby enhancing the network's ability to capture diverse features through the incorporation of branches with varying sizes and configurations. This method combines convolutional kernels of different dimensions with average pooling to improve the model's representational capacity. What makes the DBB particularly effective is its two-phase approach: during training, it employs a complex branching structure, while for inference, these branches are fused into a single convolutional layer, ensuring efficiency without compromising performance. Furthermore, the DBB can be seamlessly integrated into existing architectures, replacing

standard convolutional layers without altering the overall network design.

As illustrated in Fig. 3, during the training phase (a), DBB operates with parallel branches that consist of convolutional and pooling layers of varying sizes, which subsequently merge their outputs. After training (b), this intricate structure simplifies into a single convolutional layer, ensuring that microstructural complexity is preserved during training while maintaining operational efficiency during inference. This approach enhances feature representation without compromising the model's runtime performance.

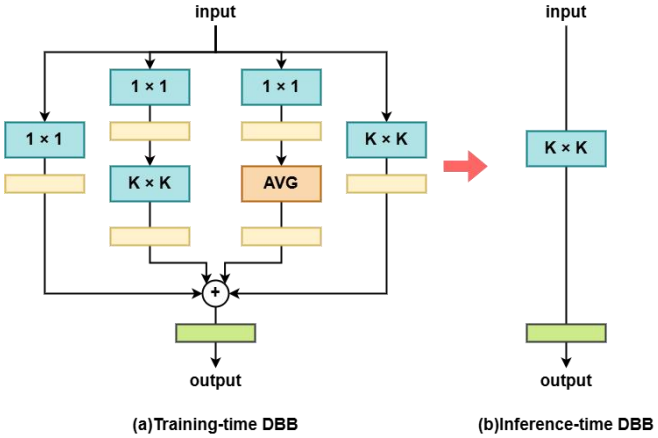


Fig. 3. DBB Training and Reasoning Process

In this study, we introduce Detect-DBB, a novel detection head that integrates DBB directly in front of the original YOLOv8 detection head. This integration enhances the feature extraction process by enabling the model to capture more refined and contextually rich information, thereby improving object detection performance. By incorporating DBB, Detect-DBB maintains inference efficiency while introducing microstructural complexity during training, representing a significant innovation in detection head design. Fig. 4 illustrates the structure of Detect-DBB.

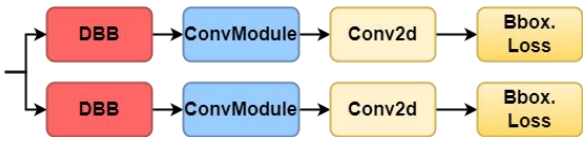


Fig. 4. Structure of Detect-DBB

D. The new Optimization of the loss function

The loss function of YOLOv8 comprises two components: classification and regression. For the classification loss, Binary Cross-Entropy Loss (BCEL) is utilized, while the regression component employs Distribution Focus Loss (DFL) and Bounding Box Regression Loss (BBRL). The overall loss function can be expressed as follows:

$$f_{Loss} = \lambda_1 f_{BCEL} + \lambda_2 f_{DFL} + \lambda_3 f_{BBRL} \tag{1}$$

The MPDIoU [35] was utilized as the bounding box loss function to enhance performance by addressing the limitations of traditional regression loss functions, such as GIoU, DIoU, CIoU, and EIoU. These conventional methods often struggle with bounding boxes that have similar aspect ratios but differ in size or location. MPDIoU improves this process by calculating bounding box similarity through the minimum keypoint distance of a horizontal rectangle, while also incorporating the overlap area, centroid distance, and width-height deviation. This approach provides a more precise evaluation of keypoint distance, resulting in a more accurate representation of the differences between the predicted and actual bounding boxes. Such precision significantly enhances detection accuracy, particularly for subtle variations in size or position among bounding boxes with similar aspect ratios. The application of MPDIoU improves YOLOv8's capability to detect objects with complex shapes and positional variations, as demonstrated in lung nodule detection. Fig. 5 illustrates the components of the LMPDIoU loss function, highlighting its role in enhancing detection accuracy.

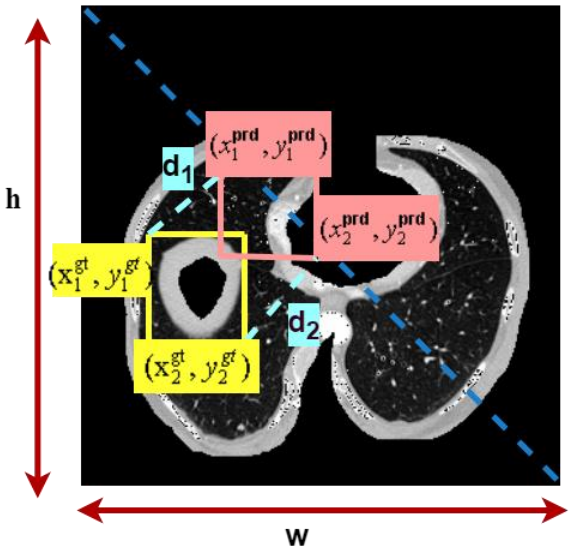


Fig. 5. Factors of the LMPDIoU loss function

Specifically, the bounding box for each prediction is as follows:

$$B_{prd} = \begin{bmatrix} x_{prd} \\ y_{prd} \\ w_{prd} \\ h_{prd} \end{bmatrix} \tag{2}$$

The true bounding box is approximated by minimizing the following loss function:

$$B_{gt} = \left[x_{gt}, y_{gt}, w_{gt}, h_{gt}\right]^T \tag{3}$$

$$L = \min_{\theta} L(B_{gt}, B_{prd} | \theta) \tag{4}$$

Where Bgt is the set of real bounding boxes and θ is the parameter of the regression depth model. The LMPDIoU loss function formula is as follows:

$$LMPDIoU = 1 - MPDIoU \tag{5}$$

QFL[36] addresses the challenge of inconsistent classification and localization in object detection by fostering a stronger synergy between these tasks. Traditional methods often treat classification and localization as separate processes, which diminishes their interdependence. QFL resolves this issue by incorporating localization quality metrics (e.g., CIoU, EIoU, and SIoU scores) into the classification loss, thereby enhancing the interaction between tasks during training.

By dynamically adjusting the weight of the localization quality score within the classification process, the QFL ensures that the classification loss more accurately reflects the complexity of the localization task. This adjustment emphasizes challenging examples, thereby enhancing both classification and localization accuracy. Unlike standard Focal Loss, which employs binary labels (0 and 1), QFL utilizes continuous labels ranging from 0 to 1, providing a more nuanced representation of localization precision. By prioritizing difficult objects, QFL minimizes errors and significantly improves the detection performance of models such as YOLOv8. This method offers a practical approach by integrating localization accuracy into classification, thereby enhancing detection and aligning model training with real-world challenges, which leads to more reliable outcomes. The optimized YOLOv8 loss function is presented in Equation (6).

$$f_{Loss} = \lambda_1 f_{QFL} + \lambda_2 f_{DFL} + \lambda_3 f_{MPDIoU} \tag{6}$$

IV. Results

A. Experimental environment

The implemented system requires substantial computational processing during operation, necessitating specific hardware requirements. Table I provides a comprehensive overview of our system's configuration.

TABLE I
EXPERIMENTAL CONFIGURATION

Project	Specific information
Operating system	Linux
CPU	Intel(R) Xeon(R) E5-2683 v4
GPU	NVIDIA GeForce RTX 4060 Ti
Memory	32GB
Language	Python 3.11
Development platform	PyTorch

To ensure the rigor and fairness of the experiment, we standardized the experimental parameter settings, as detailed in Table II.

TABLE II
EXPERIMENTAL HYPERPARAMETERS

Hyperparameters	Value
Learning Rate	0.0001
Image Size	640 × 640
Momentum	0.937
Optimizer	SGD
Batch Size	8
Epoch	300
Weight Decay	0.0005

B. Dataset and evaluation index

For this study, we utilized the publicly available ISBI Lung Nodule Analysis 2016 (LUNA16) dataset, which comprises 888 lung CT scans. These images encompass a diverse array of nodule characteristics, including various shapes and sizes, accurately reflecting the complexity of lung anatomy. To ensure precise annotation, each scan was independently reviewed by four experienced radiologists, resulting in the identification of 1,186 labeled lung nodules. The variations in nodule dimensions, forms, and densities make this dataset particularly well-suited for evaluating different detection techniques. To facilitate comprehensive model assessment, we partitioned the dataset into training, validation, and test sets using an 8:1:1 ratio. This approach enhances the reliability of our findings. Additionally, we implemented data augmentation strategies to further improve model robustness. Figure 6 illustrates an example from the LUNA16 dataset.

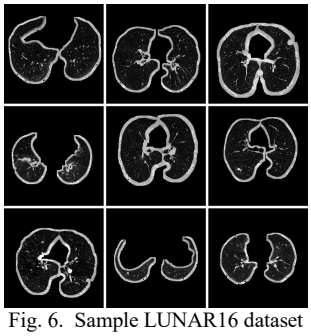


Fig. 6. Sample LUNAR16 dataset

The evaluation metrics employed are Precision, Recall, and mean Average Precision (mAP). Recall indicates the detection status, while mAP integrates Precision and Recall to offer a more comprehensive assessment of overall performance.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{7}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{8}$$

$$\text{AP} = \int_0^1 \text{P(R)dR} \tag{9}$$

$$\text{mAP} = \frac{\sum_{i=1}^{K} \text{AP}_i}{\text{K}} \tag{10}$$

Here, TP (True Positives) refers to pulmonary nodules that are correctly identified, TN (True Negatives) indicates non-pulmonary nodules that are not misclassified as pulmonary nodules, FP (False Positives) denotes non-pulmonary nodules that are incorrectly identified as pulmonary nodules, and FN (False Negatives) represents pulmonary nodules that are not accurately identified.

C. Comparison Experiment

Bounding box loss function comparison experiment

To emphasize the significance of introducing MPDIoU, we conducted a comprehensive comparison with other commonly used loss functions, including CIoU, EIoU, SIoU, and WIoU. Each of these loss functions was evaluated based on their effectiveness in optimizing object detection models under identical experimental conditions. Table III presents the experimental results of the comparison of the bounding box loss functions.

TABLE III
LOSS FUNCTION COMPARISON EXPERIMENTAL RESULTS

loss function	mAP50/%	mAP50-95/%
CIoU	89.3	46.1
EIoU	89.6	46.3
SIoU	89.7	46.6
WIoU	90.2	47.1
MPDIoU	91.4	49.5

As demonstrated in Table III, the MPDIoU loss function consistently outperforms other loss functions across both metrics, achieving a mAP50 of 91.4% and an mAP50-95 of 49.5%. This underscores the superior precision and detection accuracy of MPDIoU, particularly across multiple IoU thresholds, thereby validating its effectiveness in enhancing model performance. The comparison emphasizes the critical importance of selecting the appropriate loss function to optimize object detection models for improved accuracy.

Comparison of different backbone networks

The detection head plays a crucial role in influencing the performance of object detection. Our study compared five backbone networks — MobileNetV2, MobileNetV3, GhostNetV2, ShuffleNetV2, and VanillaNet — to evaluate their impact on pulmonary nodule detection. The results, summarized in Table IV, illustrate the performance of each backbone in terms of mAP50, mAP50-95, and model size. VanillaNet achieved the highest accuracy, with an mAP50 of 88.63% and a mAP50-95 of 47.4%, while maintaining the smallest model size at 4.87 MB. ShuffleNetV2 followed closely but had a larger weight. GhostNetV2 and MobileNetV3 demonstrated competitive accuracy butrequired significantly more parameters. MobileNetV2 exhibited the lowest performance.

These results indicate that VanillaNet optimally balances detection accuracy and computational efficiency. Its streamlined architecture enhances feature extraction while minimizing redundancy, making it ideal for real-time clinical applications. By integrating VanillaNet into our optimized YOLOv8 model, we achieve superior performance with reduced computational overhead, thereby improving the feasibility of automated pulmonary nodule detection in medical imaging.

TABLE IV
DETECTION HEAD COMPARISON RESULTS

backbone networks	mAP50/%	mAP50-95/%	Weights/MB
MobileNetV2	85.3	43.9	10.26
MobileNetV3	87.7	46.4	13.51
GhostNetV2	88.1	46.8	13.74
ShuffleNetV2	88.6	47.1	6.32
VanillaNet	88.6	47.4	4.87

Model Comparison Experiment

To thoroughly evaluate the performance of our enhanced YOLOv8 model in lung nodule detection, we conducted comparative experiments against state-of-the-art models, including Faster R-CNN, SSD, GhostNetV2, RT-DETR, YOLOv5, YOLOv6, YOLOv7, YOLOv9, and the standard YOLOv8. To ensure a fair assessment, all models were trained and tested under identical conditions using the LUNA16 dataset. A summary of the experimental findings is presented in Table V. The results demonstrate that our model achieves superior precision, recall, and mAP, all while maintaining a streamlined architecture with fewer parameters and robust image-processing capabilities.

Compared to YOLOv9, which achieves a mAP50 of 91.85%, our model enhances accuracy while maintaining a significantly smaller size and achieving a higher inference speed. Furthermore, our model outperforms RT-DETR and GhostNetV2, which, despite their competitive accuracy, incur increased computational overhead due to their larger model weights.

The results demonstrate that our optimized YOLOv8 model effectively balances accuracy, efficiency, and speed, making it highly suitable for real-time pulmonary nodule detection. Its lightweight design ensures feasibility for deployment in clinical applications while maintaining state-of-the-art detection performance.

D. Ablation Experiment

To analyze the influence of each component in the enhanced YOLOv8 model, we conducted a series of ablation studies utilizing the LUNA16 dataset. This investigation aimed to determine the individual contributions of four major modifications: (A) replacing the backbone with VanillaNet, (B) incorporating the DLKAv4 attention module into the neck, (C) introducing the Detect-DBB structure as a new detection head, and (D) refining the loss functions by substituting the bounding box loss with MPDIoU and the classification loss with QFL. The findings, summarized in Table VI, were evaluated based on three key performance indicators: precision, recall, and mAP.

The first enhancement, A (VanillaNet replacing the backbone), resulted in a significant improvement in both precision and recall. Precision increased by 1.21%, rising from 87.42% to 88.63%, while recall improved by 2.38%, increasing from 83.87% to 86.25%. This indicates that VanillaNet enhances feature extraction.

The incorporation of the B (DLKAv4 attention module) significantly enhanced performance, particularly in recall, which increased by 2.35%, rising from 86.25% to 88.60%. Precision also improved by 1.23%, increasing from 88.63% to 89.86%, and mAP50 rose from 90.10% to 91.26%. The DLKAv4 module strengthens the representation of spatial and scale features, thereby enhancing detection accuracy for nodules of varying sizes and shapes.

The third enhancement, C (Detect-DBB replacing the detection head), led to further improvements, particularly in mAP50, which increased by 0.67%, rising from 91.26% to 91.93%. Additionally, precision saw a slight increase from 89.86% to 90.07%. This new detection head significantly enhances both localization and classification accuracy.

TABLE V
COMPARISON OF EXPERIMENTAL RESULTS

Model	Precision/%	Recall/%	mAP50/%	mAP50-95/%	FPS/(f/s)	Weights/MB
Faster R-CNN	84.88	82.48	88.25	43.59	90.4	27.4
SSD	83.79	80.08	86.28	41.88	98.7	23.9
Ghostnetv2	88.28	85.25	90.23	47.26	123.1	5.13
RT-DETR	89.68	88.92	90.83	47.74	91.5	63.1
YOLOv5	84.65	82.29	88.05	43.12	128.3	3.65
YOLOv6	85.83	83.14	88.72	44.69	114.3	9.94
YOLOv7	84.35	81.21	87.28	42.74	96.6	71.3
YOLOv8	87.42	83.87	89.34	46.12	119.2	5.76
YOLOv9	90.79	90.25	91.85	50.13	69.1	124.2
ours	90.81	90.73	92.74	50.79	121.7	3.71

Finally, the implementation of D (which involved replacing traditional loss functions with MPDIoU and QFL) resulted in the most substantial overall improvementsgains, with mAP50 improvingby 0.81%, rising from 91.93% to 92.74%, while precision improved by 0.74%, increasing from 90.07% to 90.81%. These innovative loss functions enhanced both localization and classification, significantly elevating the overall performance of the model.

TABLE VI
COMPARISON OF EXPERIMENTAL RESULTS

Model	Precision/%	Recall/%	mAP50/%	mAP50-95/%
YOLOv8	87.42	83.87	89.34	46.12
+A	88.63	86.25	90.10	47.43
+A+B	89.86	88.60	91.26	48.51
+A+B+C	90.07	89.12	91.93	49.22
+A+B+C+D	90.81	90.73	92.74	50.79

E. Self-constructed dataset experiments

To enhance the stability and robustness of our model, we curated a custom dataset comprising 500 images specifically designed to meet our experimental requirements. This dataset includes variations in lighting conditions, object sizes, and occlusions, effectively addressing key challenges in lung nodule detection. To ensure consistency, all CT images underwent identical preprocessing steps, rendering the dataset highly suitable for experimentation and serving as an excellent benchmark for evaluating object detection models (see Fig. 7).

For a comprehensive assessment, we trained and tested our model alongside several state-of-the-art architectures — including YOLOv5, YOLOv6, YOLOv7, YOLOv8, YOLOv9, and RT-DETR—utilizing the same dataset. The detection performance was meticulously analyzed using key evaluation metrics, including precision, recall, mAP50, and mAP50-95.

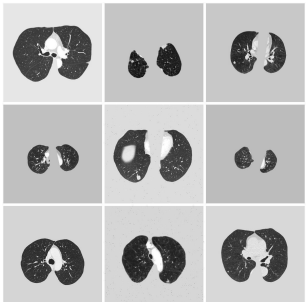


Fig. 7. Sample self-built dataset

The experimental results presented in Table VII demonstrate the comprehensive advantages of our model over existing methods in lung nodule detection tasks. Notably, our approach achieves a precision of 90.8% and a recall of 88.3%, surpassing even the latest YOLOv9 (89.7% precision, 87.5% recall) and RT-DETR (88.6% precision, 86.2% recall). This performance gap becomes even more pronounced in the mAP50 metric (91.7% compared to YOLOv9's 90.8%), indicating superior detection accuracy under standard evaluation conditions.

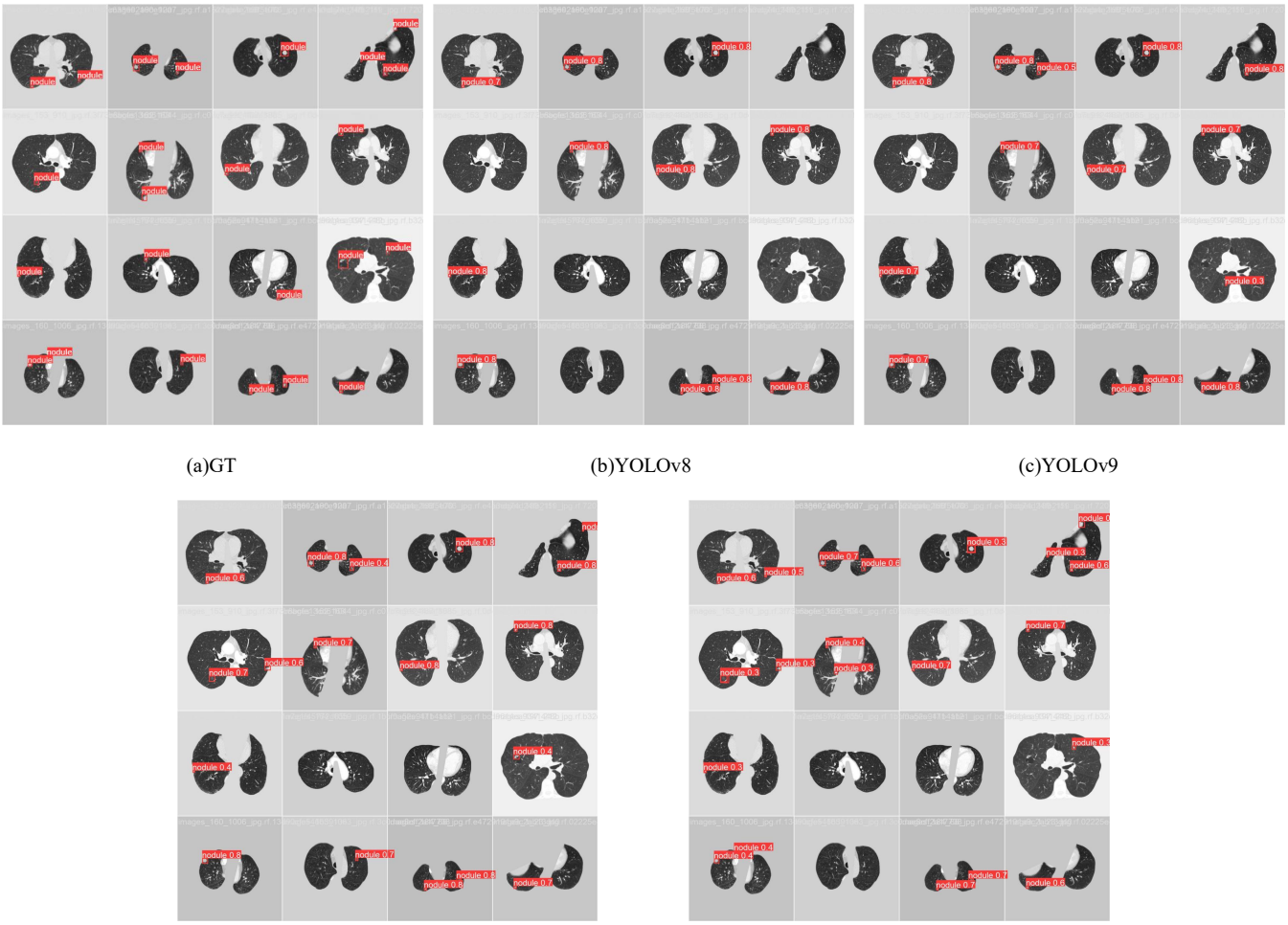
The advantage remains evident in the more rigorous mAP50-95 assessment, where our model achieves 49.7%, compared to YOLOv9's 48.3%. This suggests an enhanced capability in addressing multi-scale detection challenges. These improvements are a result of our optimized network architecture and training strategy, which are specifically designed to accommodate the unique characteristics of medical imaging, particularly in addressing the high similarity between nodules and surrounding tissues.

Significantly, the experiments demonstrate that models trained on our custom medical dataset consistently outperform those utilizing generic datasets, thereby confirming the necessity of domain-specific data curation. The performance gaps of 2.1% to 6.3% across various YOLO versions underscore that architectural improvements alone cannot compensate for mismatches between datasets and tasks. This finding offers essential guidance for medical computer vision research, advocating for a balanced approach to both algorithmic innovation and dataset specialization.

TABLE VII
EXPERIMENTAL RESULTS ON SELF-BUILT DATASET

Model	Precision/%	Recall/%	mAP50/%	mAP50-95/%
RT-DETR	88.6	86.2	89.3	47.4
YOLOv5	86.8	84.2	87.7	45.6
YOLOv6	86.9	84.4	87.8	45.9
YOLOv7	84.5	82.6	85.8	44.7
YOLOv8	87.1	85.5	88.5	46.2
YOLOv9	89.7	87.5	90.8	48.3
ours	90.8	88.3	91.7	49.7

Fig. 8 provides a visual comparison of detection results between ground truth (GT) annotations and four state-of-the-art models: YOLOv8, YOLOv9, RT-DETR, and our proposed Nodule-YOLOv8. This visualization highlights the distinct performance characteristics of each model, particularly in challenging scenarios involving small nodules, irregularly shaped lesions, and nodules with low contrast against surrounding tissues. While YOLOv8, YOLOv9, and RT-DETR demonstrate competent detection capabilities, they occasionally overlook subtle nodules or generate false positives in areas with vascular overlap.



(a)GT (b)YOLOv8 (c)YOLOv9

(d)RT-DETR (e)Nodule-YOLOv8

Fig. 8. Comparison of model prediction results

V. Conclusion

This study systematically optimized the YOLOv8 architecture and validated its effectiveness for pulmonary nodule detection using the LUNA16 dataset. Component-level ablation studies were conducted to evaluate four strategic enhancements: the replacement of the VanillaNet backbone, the integration of the DLKAv4 attention module in feature fusion pathways, the implementation of the Detect-DBB detection head, and joint optimization utilizing MPDIoU regression loss in conjunction with QFL classification loss. Experimental results demonstrated significant improvements in sensitivity, specificity, and multi-threshold average precision metrics, thereby establishing an optimized framework for CT-based pulmonary nodule screening.

Future research will adopt a dual focus on algorithmic refinement and clinical integration. Technologically, efforts will concentrate on validating the generalizability of models across low-dose CT protocols and multi-vendor equipment, while developing cross-modal detection frameworks compatible with X-ray and MRI imaging. Clinically, we will prioritize embedded deployment solutions for real-time inference, augmented by interpretable heatmap visualizations to enhance diagnostic confidence. Multi-center prospective trials will assess the compatibility of clinical workflows, ultimately establishing an integrated management system that incorporates intelligent alert mechanisms, malignancy grading, and growth prediction models. This initiative aims to standardize AI-radiology collaboration paradigms, advancing intelligent solutions for pulmonary nodule management through harmonized computational and clinical approaches.

References

[1] Di Yang,Yafei Miao,Changjiang Liu,et al.(2024).Advances in artificial intelligence applications in the field of lung cancer.Frontiers in Oncology1449068-1449068.

[2] Chassagnon G, De Margerie-Mellon C, Vakalopoulou M, et al. Artificial intelligence in lung cancer: current applications and perspectives. Jpn J Radiol 2023;41(3):235-244.

[3] Abadia AF, Yacoub B, Stringer N, et al. Diagnostic Accuracy and Performance of Artificial Intelligence in Detecting Lung Nodules in Patients With Complex Lung Disease: A Noninferiority Study. J Thorac Imaging 2022;37(3):154-161.

[4] Li R, Xiao C, Huang Y, Hassan H, Huang B. Deep Learning Applications in Computed Tomography Images for Pulmonary Nodule Detection and Diagnosis: A Review. Diagnostics (Basel) 2022;12(2):298.

[5] Gu D, Liu G, Xue Z. On the performance of lung nodule detection, segmentation and classification. Comput Med Imaging Graph 2021;89:101886.

[6] Di Yang,Yafei Miao,Changjiang Liu,et al.(2024).Advances in artificial intelligence applications in the field of lung cancer.Frontiers in Oncology1449068-1449068.

[7] Adams Scott J,Mikhael Peter,Wohlwend Jeremy,Barzilay Regina,Sequist Lecia V & Fintelmann Florian J.(2023).Artificial Intelligence and Machine Learning in Lung Cancer Screening..Thoracic surgery clinics(4),401-409.

[8] Chen Jingwen,Cao Rong,Jiao Shengyin,et al.(2023).Application value of a computer-aided diagnosis and management system for the

detection of lung nodules..Quantitative imaging in medicine and surgery(10),6929-6941.
[9] Swathi Y, Challa M. YOLOv8: Advancements and Innovations in Object Detection.ADICS:Chennai, India;2024.
[10] Sebastian AE, Dua D. Lung Nodule Detection via Optimized Convolutional Neural Network: Impact of Improved Moth Flame Algorithm. Sens Imaging 2023;24(1):11.
[11] Kim H, Lee KH, Han K, et al. Development and Validation of a Deep Learning-Based Synthetic Bone-Suppressed Model for Pulmonary Nodule Detection in Chest Radiographs. JAMA Netw Open 2023;6(1):e2253820.
[12] Mkindu H, Wu L, Zhao Y. 3D multi-scale vision transformer for lung nodule detection in chest CT images. SIViP 2023;17:2473–2480.
[13] Mkindu H, Wu L, Zhao Y. Lung nodule detection of CT images based on combining 3D-CNN and squeeze-and-excitation networks. Multimed Tools Appl 2023;82:25747–25760.
[14] Mkindu H,Wu L,Zhao Y.Lung nodule detection in chest CT images based on vision transformer network with Bayesian optimization.Biomedical Signal Processing and Control 2023;85:104866.
[15] Xu R, Liu Z, Luo Y, et al. SGDA: Towards 3-D Universal Pulmonary Nodule Detection via Slice Grouped Domain Attention. IEEE/ACM Trans Comput Biol Bioinform 2024;21(4):1093-1105.
[16] Mohammed J Z A ,Azmi A M M ,Binti F K , Manshor N.A novel two-stage Lung nodule detection in medical images based on deep learning.Journal of Physics: Conference Series 2023;2467(1).
[17] Shao J, Zhou L, Yeung SYF, Lei T, Zhang W, Yuan X. Pulmonary Nodule Detection and Classification Using All-Optical Deep Diffractive Neural Network. Life. 2023; 13(5):1148.
[18] Gugulothu VK, Balaji S. An early prediction and classification of lung nodule diagnosis on CT images based on hybrid deep learning techniques. Multimed Tools Appl 2023;31:1-21.
[19] Yang L, Cai H, Luo X, et al. A lightweight neural network for lung nodule detection based on improved ghost module. Quant Imaging Med Surg 2023;13(7):4205-4221.
[20] Cao K, Tao H, Wang Z. Three-Dimensional Multifaceted Attention Encoder–Decoder Networks for Pulmonary Nodule Detection. Applied Sciences 2023;13(19):10822.
[21] Jian M, Jin H, Zhang L, Wei B, Yu H. DBPNDNet: dual-branch networks using 3DCNN toward pulmonary nodule detection. Med Biol Eng Comput 2024;62(2):563-573.
[22] Zhang Y, Jiang B, Zhang L, et al. Lung Nodule Detectability of Artificial Intelligence-assisted CT Image Reading in Lung Cancer Screening. Curr Med Imaging 2022;18(3):327-334.
[23] Han L, Li F, Yu H, Xia K, Xin Q, Zou X. BiRPN-YOLOvX: A weighted bidirectional recursive feature pyramid algorithm for lung nodule detection. J Xray Sci Technol 2023;31(2):301-317.
[24] Ren C, Hou S, Hou J, Pang Y. SwiF-YOLO: A Deep Learning Method for Lung Nodule Detection. International Journal of Biology and Life Sciences;2024.
[25] Wu X, Zhang H, Sun J, Wang S,Zhang Y. YOLO-MSRF for lung nodule detection. Biomed Signal Process and Control 2024;94:106318.
[26] Yu Liu & Yongcai Ao.(2024).Deformable attention mechanism-based YOLOv7 structure for lung nodule detection.The Journal of Supercomputing(prepublish),1-20.
[27] Chaosheng Tang,Feifei Zhou,Junding Sun & Yudong Zhang.(2025).Lung-YOLO: Multiscale feature fusion attention and cross-layer aggregation for lung nodule detection.Biomedical Signal Processing and Control106815-106815.
[28] Shital D. Bhatt,Mitesh B. Astik & Himanshu B. Soni.(2024).Automatic Lung Nodules Detection Using a Modified YOLOv5.Journal of Biomimetics, Biomaterials and Biomedical Engineering47-62.
[29] Parveen Rahamathulla M, Sam Emmanuel WR, Bindhu A, Mustaq Ahmed M. YOLOv8's advancements in tuberculosis identification from chest images. Front Big Data 2024;7:1401981.
[30] Chen, H., Wang, Y., Guo, J., & Tao, D. (2023). VanillaNet: the Power of Minimalism in Deep Learning. ArXiv, abs/2305.12972.
[31] Azad, R., Niggemeier, L., Huttemann, M., Kazerouni, A., Aghdam, E.K., Velichko, Y., Bagci, U., & Merhof, D. (2023). Beyond Self-Attention: Deformable Large Kernel Attention for Medical Image Segmentation. 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 1276-1286.
[32] Xiong, Y., Li, Z., Chen, Y., Wang, F., Zhu, X., Luo, J., Wang, W., Lu, T., Li, H., Qiao, Y., Lu, L., Zhou, J., & Dai, J. (2024). Efficient Deformable ConvNets: Rethinking Dynamic and Sparse Operator for Vision Applications. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 5652-5661.
[33] Wang W, Dai J, Chen Z, Huang Z, Li Z, Zhu X, et al. InternImage: Exploring Large-Scale Vision Foundation Models with Deformable Convolutions. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition.CVPR;2023.
[34] X. Ding, X. Zhang, J. Han and G. Ding, "Diverse Branch Block: Building a Convolution as an Inception-like Unit," 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 2021, pp. 10881-10890, doi: 10.1109/CVPR46437.2021.01074.
[35] Ma, S., & Xu, Y. (2023). MPDIoU: A Loss for Efficient and Accurate Bounding Box Regression. ArXiv, abs/2307.07662.
[36] Li X, Wang W, Wu L, et al. (2020). Generalized Focal Loss: Learning Qualified and Distributed Bounding Boxes for Dense Object Detection. ArXiv, abs/2006.04388.

Zichao Liu was born on September 11, 2000, in Weifang, Shandong, China. He is pursuing his master's degree in information science and technology from Qingdao University of Science and Technology, Qingdao, Shandong, China, specializing in deep learning and medical object detection.

He has worked on several research projects in the field of deep learning, focusing on applications in medical imaging and object detection. His current work involves developing advanced algorithms for accurate and efficient medical target detection. His research interests include deep learning, computer vision, and their applications in healthcare technology.

Lili Wei, Ph.D. in Nursing, Chief Nurse, and Fellow of the American Academy of Nursing (FAAN), is the Dean of Qingdao University Affiliated Hospital and a doctoral supervisor.

Her expertise covers chronic disease management, nursing education, maternal and child healthcare, and nursing administration. She holds multiple leadership roles, including Vice Chair of the Chinese Nursing Association's Internal Medicine Nursing Committee and the Shandong Nursing Association, Chair of the Pain Nursing Committee of Shandong Province, and Chair of Qingdao's Nursing Education Committee. She also serves on the editorial boards of Chinese Nursing Journal, China Nursing Management, Journal of Integrative Nursing, and Nursing Open.

Wei has led national and provincial research projects on nursing safety systems, gestational diabetes prediction models, and AI interventions for high-risk pregnancies, earning prestigious awards such as the Second Prize for Scientific and Technological Progress in Shandong (2022) and the First Prize from the Shandong Nursing Association (2020).

Tingqiang Song is an Associate Professor and Vice Dean at the School of Information Science and Technology, Qingdao University of Science and Technology. His research focuses on artificial intelligence, data mining, machine learning, and image processing.

He has led or participated in over 10 major research projects, including National 863 and NSFC programs, and holds five invention patents, three utility model patents, ten software copyrights, and has published over 30 papers. He has co-authored three textbooks, translated eight technical books, and received awards such as the Provincial Third Prize for Computer Applications and Qingdao's Second Prize for Science and Technology Progress.

In education, he has led innovative education projects, developed recognized online courses, and earned top university teaching honors. He has guided students to win over 40 awards in national and provincial competitions, including National Second Prizes in the China College Student Innovation and Entrepreneurship Competition and the National Graduate Electronics Design Competition.