

ICN-DETR: Inter-Slice Continuity and Nodule-Aware Pulmonary Nodule Detection

Man Gong

School of Electronic Engineering, Tianjin University of Technology and Education, Tianjin 300222, China

Email: gongman@tute.edu.cn

How to cite this paper: Gong, M. (2026). ICN-DETR: Inter-Slice Continuity and Nodule-Aware Pulmonary Nodule Detection. Academic Journal of Emerging Technologies, 4(1), 1–10. ISSN Print: 3104-4417, ISSN Online: 3104-4425.

<https://doi.org/10.63313/AJET.9074>

Published: 2026-10-08

Copyright © 2026 by author(s) and Erytis Publishing Limited.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Abstract

Pulmonary nodule detection in computed tomography images remains challenging, especially for small nodules with low contrast and limited image features. To improve detection performance, we propose ICN-DETR, a pulmonary nodule detection model based on RT-DETR-R18. The model uses three consecutive CT slices as input and introduces feature interaction between adjacent slices to strengthen the representation of pulmonary nodules. In addition, shallow high-resolution features are incorporated into multi-scale feature fusion to reduce the loss of boundary and texture information during downsampling. The query selection and localization process is also improved to increase the sensitivity of the detector to small nodules and reduce localization errors. Experiments on the LUNA16 dataset show that ICN-DETR achieves a Precision of 89.5%, Recall of 90.3%, F1-score of 89.9%, and mAP@0.5 of 91.9%. Compared with RT-DETR-R18, the F1-score and mAP@0.5 increase by 3.9 and 4.3 percentage points, respectively. The results demonstrate that ICN-DETR can effectively improve pulmonary nodule detection with a moderate increase in model complexity.

Keywords

Pulmonary Nodule Detection; Computed Tomography; RT-DETR; Inter-Slice Continuity; Cross-Scale Feature Fusion

1. Introduction

Lung cancer remains one of the major causes of cancer-related morbidity and mortality worldwide, and early detection is essential for improving patient prognosis [1]. Low-dose computed tomography (LDCT) has been widely used for lung cancer screening, in which pulmonary nodules are important radiological manifestations associated with early-stage lung cancer [2]. However, a CT examination usually contains a large number of consecutive slices, making manual interpretation time-consuming. Moreover, small pulmonary nodules are easily missed because of their limited size, low contrast, and diverse appearances.

Therefore, accurate automatic lung nodule detection is of considerable value for computer-aided screening.

In recent years, convolutional neural networks (CNNs) have achieved substantial progress in pulmonary nodule detection. Setio et al. [3] employed multi-view CNNs to integrate two-dimensional information from different views, Dou et al. [4] utilized 3D CNNs to exploit volumetric contextual information, and Su et al. [5] introduced Faster R-CNN for automatic nodule detection. Nevertheless, 3D networks generally require considerable computational resources, whereas 2D and multi-slice methods still have limited ability to explicitly model spatial relationships among consecutive CT slices. For small nodules, the visual information in a single axial slice is often insufficient, while adjacent slices provide complementary information about axial continuity and morphological variation. Simply stacking consecutive slices along the channel dimension cannot explicitly characterize these inter-slice relationships.

Another challenge is the progressive loss of fine-grained nodule information during downsampling and cross-scale feature propagation. Although multi-scale architectures such as FPN improve feature aggregation [6], preserving high-resolution details remains difficult for small nodules. The introduction of Transformer-based detectors, particularly DETR [7] and RT-DETR [8], provides stronger global modeling and efficient end-to-end detection. Recent studies have further improved RT-DETR for pulmonary nodule detection through multi-scale fusion and feature re-weighting [9,10]. However, small-nodule features occupy only a small proportion of encoder tokens and may still be overlooked during Object Query selection. In addition, small bounding boxes are more sensitive to coordinate deviations, which can further affect matching and localization accuracy.

To address these limitations, we propose ICN-DETR, an inter-slice continuity and nodule-aware detection framework based on RT-DETR-R18. The main contributions are summarized as follows:

- (1) We propose a Difference-Guided Progressive Inter-Slice Backbone (DPIB), which explicitly models axial relationships between the central slice and its adjacent slices using difference features, consistency features, and dynamic gating.
- (2) We propose a Small-Nodule Detail-Preserving Cross-Scale Fusion Module (SD-CFM) to introduce high-resolution shallow information into cross-scale feature fusion and reduce the loss of fine-grained nodule features.
- (3) We propose a Nodule-Aware Query and Localization Optimization mechanism (NQLO), which incorporates nodule-center information into Query selection and improves localization through scale-adaptive geometric matching and bounding-box regression.

2. Related Work

2.1. CNN-based Lung Nodule Detection

Convolutional neural networks have been widely applied to pulmonary nodule

detection because of their strong ability to extract local image features. Setio et al. [3] proposed a multi-view CNN for false-positive reduction by combining multiple two-dimensional views of nodule candidates. Dou et al. [4] developed a two-stage 3D ConvNet framework, in which a 3D fully convolutional network was used for candidate screening and a residual network was employed for false-positive reduction. Xie et al. [11] proposed a 2D CNN-based detection framework by modifying Faster R-CNN with two region proposal networks and a deconvolutional layer, and fused predictions from different CT slices to incorporate complementary information. Li and Fan [12] subsequently proposed DeepSEED, which combined a 3D encoder-decoder network, a region proposal network, and squeeze-and-excitation blocks for pulmonary nodule detection. These CNN-based approaches have substantially improved detection performance by exploiting local, multi-view, or volumetric information. However, 3D CNNs generally require considerable computational resources, whereas 2D and multi-slice approaches still have limitations in explicitly modeling relationships between adjacent CT slices. In addition, repeated downsampling may weaken fine-grained features of small nodules.

2.2. Transformer-based Lung Nodule Detection

The introduction of Transformer-based object detectors provides an alternative way to model global feature relationships. DETR formulates object detection as a direct set prediction problem using an encoder-decoder architecture and object queries [7]. Deformable DETR further introduces multi-scale deformable attention, improving convergence and detection performance, particularly for small objects [13]. RT-DETR improves real-time end-to-end detection through an efficient hybrid encoder and query selection strategy [8]. Based on RT-DETR, recent studies have explored its application to pulmonary nodule detection. Tang et al. [9] proposed an LN-DETR model incorporating PC-EMA, CTrans, and GS-CCFM to enhance multi-scale feature extraction and cross-scale feature fusion. Zhou et al. [10] developed another RT-DETR-based LN-DETR framework using FasterBlock, Deep and Shallow Detail Fusion, and Joint Context Feature modules to improve feature extraction, feature fusion, and downsampling. Although these studies improve multi-scale representation and computational efficiency, explicit use of inter-slice relationships remains limited, and the query selection and localization process is still largely inherited from general-purpose object detectors. These issues are particularly important for small pulmonary nodules, whose features occupy only a limited spatial region and are sensitive to localization errors.

3. Materials and Methods

3.1. Overall Architecture

The overall architecture of ICN-DETR is shown in Figure 1. ICN-DETR is built on

RT-DETR-R18 and consists of three main components: DPIB, SD-CFM, and NQLO. The input contains three consecutive axial CT slices, including the central slice I_z and its adjacent slices I_{z-1} and I_{z+1} . These slices are processed using shared backbone parameters to extract complementary inter-slice information while controlling model complexity.

DPIB is embedded at the shallow and intermediate backbone stages to progressively model the relationships between consecutive slices and aggregate useful adjacent-slice information into the central branch. The backbone produces multi-level features $\{S_2, S_3, S_4, S_5\}$, which are subsequently processed by SD-CFM to preserve shallow spatial details and generate enhanced multi-scale features $\{P_3, P_4, P_5\}$. After Transformer encoding, NQLO improves candidate Query selection and localization using nodule-center information and scale-related geometric constraints. The selected queries are finally fed into the Transformer decoder for nodule classification and bounding-box prediction.

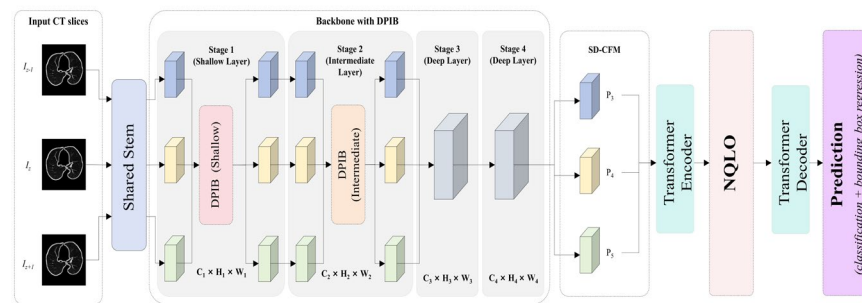


Figure 1. Overall architecture of ICN-DETR.

3.2. Difference-Guided Progressive Inter-Slice Backbone

Small pulmonary nodules contain limited texture and boundary information in a single axial CT slice, while adjacent slices provide complementary information about their axial continuity. Conventional multi-slice approaches usually mix consecutive slices at the network input, making it difficult to explicitly distinguish useful inter-slice relationships from irrelevant anatomical responses. Therefore, we propose DPIB, which maintains independent representations of consecutive slices and performs progressive inter-slice interaction at shallow and intermediate backbone stages.

Let $\mathbf{F}_0^l, \mathbf{F}_-^l$ and $\mathbf{F}_+^l$ denote the features of the central, preceding, and following slices at the l -th stage, respectively. To characterize local variations across consecutive slices, the inter-slice difference features are defined as

$$D_{\pm}^l = |\mathbf{F}_0^l - \mathbf{F}_{\pm}^l| \quad (1)$$

where $D_{\pm}^l$ represents the response difference between the central and adjacent slices.

To further describe spatial responses that remain consistent across consecutive slices, the corresponding consistency features are defined as

$$C_{\pm}^l = F_0^l \odot F_{\pm}^l \quad (2)$$

where $\odot$ denotes element-wise multiplication. Equation (1) describes inter-slice variations, whereas Equation (2) emphasizes regions with consistent responses. Together, they provide complementary information about axial variation and continuity.

The central feature, adjacent feature, and their corresponding difference and consistency features are then concatenated and encoded by a lightweight mapping function. A Sigmoid activation is subsequently applied to generate the dynamic gating weights γ_-^l and γ_+^l for the preceding and following slices, respectively. The enhanced central-slice feature is expressed as

$$\tilde{F}_0^l = F_0^l + \gamma(-) (\gamma_-^l \odot F_-^l + \gamma_+^l \odot F_+^l) \quad (3)$$

where γ_l is a learnable residual scaling parameter. Equation (3) retains the central-slice feature as the main representation while selectively incorporating useful information from adjacent slices. Features with stronger relevance to the central region receive larger gating responses, whereas weakly related adjacent information is suppressed.

The inter-slice interaction is performed at both shallow and intermediate backbone stages. Shallow features retain relatively high spatial resolution and preserve more boundary and local intensity information, whereas intermediate features provide stronger structural representations. After the intermediate-stage interaction, useful adjacent-slice information is aggregated into the central branch, and the subsequent backbone returns to a two-dimensional single-branch structure.

3.3. Small-Nodule Detail-Preserving Cross-Scale Fusion Module

After DPIB, the network produces multi-level features $\{S_2, S_3, S_4, S_5\}$. Shallow features preserve fine spatial information, whereas deeper features contain stronger semantic representations. For small pulmonary nodules, repeated downsampling may weaken boundary, texture, and low-response information before multi-scale fusion.

To reduce this information loss, SD-CFM introduces the high-resolution feature S_2 into the cross-scale fusion process. S_2 is first aligned with S_3 in spatial resolution and channel dimension and is then used for detail compensation:

$$\tilde{S}_3 = S_3 + \mathcal{A}(S_2) \quad (4)$$

where $\mathcal{A}(\cdot)$ denotes spatial and channel alignment. The enhanced feature is subsequently fused with deeper semantic features to generate $\{P_3, P_4, P_5\}$. In this

way, SD-CFM combines shallow spatial details with high-level semantic information, reducing the loss of small-nodule features during cross-scale propagation.

3.4. Nodule-Aware Query and Localization Optimization

After Transformer encoding, a limited number of candidate features are selected to initialize object queries. Because small nodules occupy only a small spatial region, their responses may be suppressed by abundant background features. In addition, small bounding boxes are more sensitive to positional deviations.

NQLO therefore optimizes both Query selection and localization. Nodule-center information is introduced during Query selection to increase the priority of candidate features near potential nodule centers. During target matching, classification information and geometric relationships are jointly considered, with greater emphasis placed on positional information for small targets.

Bounding-box regression is subsequently used to refine the predicted position and scale. By combining nodule-aware Query selection with scale-adaptive localization, NQLO improves the stability of small-nodule detection.

4. Experiments and Results

4.1. Dataset

Experiments were conducted on the publicly available Lung Nodule Analysis 2016 (LUNA16) dataset. LUNA16 is derived from the LIDC-IDRI database, which contains 1,018 thoracic CT examinations annotated by four experienced radiologists through a two-phase annotation procedure [14]. After excluding scans with a slice thickness greater than 2.5 mm, LUNA16 contains 888 CT scans. The reference standard includes 1,186 pulmonary nodules with diameters of at least 3 mm that were accepted as nodules by at least three of the four radiologists. The dataset is officially divided into ten subsets.

4.2. Experimental Setting and Evaluation Metrics

4.2.1. Experimental Setting

All experiments were implemented in Python 3.9 using PyTorch 1.8.0 and conducted on a Lenovo Legion R7000 2021 laptop equipped with an AMD Ryzen 5 5600H processor, an NVIDIA GeForce RTX 3050 Laptop GPU with 4 GB of video memory, and 16 GB of system memory. The operating system was Windows 11, and CUDA 11.1 was used for GPU acceleration.

The input images were resized to 640×640 pixels. The model was trained for 300 epochs using the AdamW optimizer. The initial learning rate and weight decay were both set to 1×10^{-4} . The batch size was set to 2. Automatic mixed precision (AMP) was enabled to reduce GPU memory consumption, and the random seed was fixed at 42.

To ensure a fair comparison, all comparison and ablation experiments were conducted using the same data preprocessing procedure, training strategy, and hyperparameter settings.

Table 1. Experimental Environment Configuration.

Software and Hardware configuration	Version or Model
CPU	AMD Ryzen 5 5600H
GPU	NVIDIA GeForce RTX 3050 Laptop GPU
Python	3.9
PyTorch	1.8.0
Operating System	Windows 11

4.2.2. Evaluation Metrics

The detection performance was evaluated using Precision, Recall, F1-score, and mAP@0.5. Model complexity was additionally assessed using the number of parameters and GFLOPs. Precision and Recall are defined as

$$P = \frac{TP}{TP+FP}, \quad R = \frac{TP}{TP+FN} \quad (5)$$

and the F1-score is calculated as

$$F1 = \frac{2PR}{P+R} \quad (6)$$

where TP, FP, and FN denote true positives, false positives, and false negatives, respectively. mAP@0.5 represents the mean average precision at an IoU threshold of 0.5 and is used to evaluate the overall detection performance of the model.

4.3. Comparison with Existing Methods

To evaluate the overall performance of the proposed method, comparisons were conducted with representative CNN-based and Transformer-based detectors, including Faster R-CNN, YOLOv5-S, YOLOv8-S, Deformable DETR, and RT-DETR-R18. RT-DETR-R18 was selected as the direct baseline because the proposed method was developed based on this architecture. All reproduced models were trained and evaluated using the same LUNA16 data partition, preprocessing procedure, and training settings.

Table 2. Comparison Results on LUNA16.

Model	Precision (%)	Recall (%)	F1-score (%)	mAP @50(%)	Params (M)	GFLOPs
Faster R-CNN	85.2	84.3	84.8	84.8	-	-
YOLOv5-S	86.1	87.5	86.8	87.4	9.2	24.0
YOLOv8-S	88.4	86.9	87.6	88.2	11.3	28.7
Deformable DETR	78.6	74.2	76.3	80.5	42.0	175.0
RT-DETR-R18	85.7	86.3	86.0	87.6	20.0	62.0
ICN-DETR	89.5	90.3	89.9	91.9	20.6	79.0

As shown in Table 2, ICN-DETR achieved a Precision of 89.5%, Recall of 90.3%, F1-score of 89.9%, and mAP@0.5 of 91.9%. Compared with RT-DETR-R18, the F1-score and mAP@0.5 increased by 3.9 and 4.3 percentage points, respectively. The higher Recall indicates a reduction in missed detections. ICN-DETR used 20.6 M parameters and 79.0 GFLOPs, showing that the performance improvement was achieved with a moderate increase in computational cost.

4.4. Ablation Experiment

To investigate the contribution of each proposed component, ablation experiments were performed using RT-DETR-R18 as the baseline. DPIB, SD-CFM, and NQLO were first evaluated individually and then combined in different configurations. All experiments used identical training and evaluation settings.

Table 3. Ablation Results of ICN-DETR.

Model	Precision (%)	Recall (%)	F1-score (%)	mAP@50/%
Baseline	85.7	86.3	86.0	87.6
+DPIB	87.6	89.1	88.3	89.4
+SD-CFM	88.1	87.2	87.6	88.9
+NQLO	87.7	88.1	87.9	89.1
+DPIB +SD-CFM	89.3	89.6	89.4	90.6
+DPIB +NQLO	89.4	88.4	88.9	90.4
+SD-CFM +NQLO	89.1	89.8	89.4	90.2
ICN-DETR	89.5	90.3	89.9	91.9

As shown in Table 3, DPIB provided the largest individual improvement, increasing mAP@0.5 from 87.6% to 89.4% and Recall from 86.3% to 89.1%. SD-CFM and NQLO individually increased mAP@0.5 to 88.9% and 89.1%, respectively. Combining two components produced further gains, while the complete ICN-DETR achieved the highest mAP@0.5 of 91.9%. Compared with the baseline, the full model improved F1-score and mAP@0.5 by 3.9 and 4.3 percentage points, respectively, confirming that the three components provide complementary improvements.

4.5. Visualization Results

Figure 2 presents qualitative detection results of different methods on representative cases from the LUNA16 dataset. Compared with the other detectors, ICN-DETR shows fewer missed detections and more reliable bounding-box predictions, particularly for small and low-contrast nodules. The qualitative observations are consistent with the improvements in Recall and mAP@0.5 reported in Table 2.

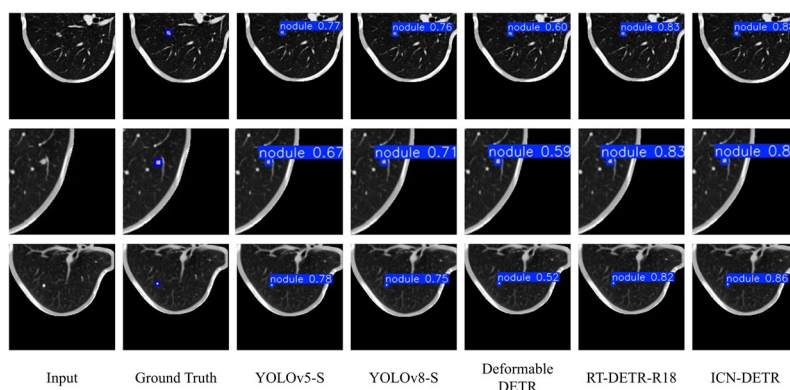


Figure 2. Qualitative detection results on the LUNA16 dataset.

5. Conclusion

This study proposed ICN-DETR, an RT-DETR-R18-based framework for pulmonary nodule detection. DPIB progressively models inter-slice continuity between consecutive CT slices, SD-CFM preserves fine-grained nodule information during cross-scale feature propagation, and NQLO improves query selection and localization.

On the LUNA16 dataset, ICN-DETR achieved a Precision of 89.5%, Recall of 90.3%, F1-score of 89.9%, and mAP@0.5 of 91.9%. Compared with RT-DETR-R18, the F1-score and mAP@0.5 improved by 3.9 and 4.3 percentage points, respectively. Ablation results further confirmed the effectiveness and complementarity of the three proposed components. Future work will focus on reducing computational cost and evaluating the model on additional multi-center CT datasets.

References

- [1] Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 2024, 74(3): 229–263. DOI: 10.3322/caac.21834.
- [2] National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *New England Journal of Medicine*, 2011, 365(5): 395–409. DOI: 10.1056/NEJMoa1102873.
- [3] Setio A A A, Ciompi F, Litjens G, et al. Pulmonary nodule detection in CT images: False positive reduction using multi-view convolutional networks. *IEEE Transactions on Medical Imaging*, 2016, 35(5): 1160–1169. DOI: 10.1109/TMI.2016.2536809.
- [4] Dou Q, Chen H, Jin Y, et al. Automated pulmonary nodule detection via 3D ConvNets with online sample filtering and hybrid-loss residual learning. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2017*, 2017: 630–638. DOI: 10.1007/978-3-319-66179-7_72.
- [5] Su Y, Li D, Chen X. Lung nodule detection based on Faster R-CNN framework. *Computer Methods and Programs in Biomedicine*, 2021, 200: 105866. DOI: 10.1016/j.cmpb.2020.105866.
- [6] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 2117–2125. DOI: 10.1109/CVPR.2017.106.
- [7] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers. In: *European Conference on Computer Vision*, 2020: 213–229. DOI: 10.1007/978-3-030-58452-8_13.
- [8] Zhao Y, Lv W, Xu S, et al. DETRs beat YOLOs on real-time object detection. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024: 16965–16974.
- [9] Tang J, Chen X, Fan L, Zhu Z, Huang C. LN-DETR: An efficient Transformer architecture for lung nodule detection with multi-scale feature fusion. *Neurocomputing*, 2025, 633: 129827. DOI: 10.1016/j.neucom.2025.129827.
- [10] Zhou D, Xu H, Liu W, Liu F. LN-DETR: Cross-scale feature fusion and re-weighting for lung nodule detection. *Scientific Reports*, 2025, 15: 15543. DOI: 10.1038/s41598-025-00309-7.
- [11] Xie H, Yang D, Sun N, Chen Z, Zhang Y. Automated pulmonary nodule detection in CT images using deep convolutional neural networks. *Pattern Recognition*, 2019, 85: 109–119. DOI: 10.1016/j.patcog.2018.07.031.
- [12] Li Y, Fan Y. DeepSEED: 3D squeeze-and-excitation encoder-decoder convolutional

- neural networks for pulmonary nodule detection. In: 2020 IEEE 17th International Symposium on Biomedical Imaging, 2020: 1866–1869. DOI: 10.1109/ISBI45749.2020.9098317.
- [13] Zhu X, Su W, Lu L, Li B, Wang X, Dai J. Deformable DETR: Deformable transformers for end-to-end object detection. In: International Conference on Learning Representations, 2021.
- [14] Armato S G III, McLennan G, Bidaut L, et al. The Lung Image Database Consortium and Image Database Resource Initiative: A completed reference database of lung nodules on CT scans. *Medical Physics*, 2011, 38(2): 915–931. DOI: 10.1118/1.3528204.
- [15] Setio A A A, Traverso A, de Bel T, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge. *Medical Image Analysis*, 2017, 42: 1–13. DOI: 10.1016/j.media.2017.06.015.