



Integrated Channel Equally-Divided and Coordinate Attention Feature Pyramid for subtype detection of lung cancer

Songzhe Ma^{a,c,d}, Huimin Lu^{b,c,d}, Han Xue^{c,d,e}, Chenglin Lin^{b,c,d}, Zexing Zhang^{b,c,d}, Yilong Wang^a

^a School of Mathematics and Statistics, Changchun University of Technology, 2055 Yan 'an Avenue, Changchun 130012, China

^b School of Computer Science and Engineering, Changchun University of Technology, 3000 North Yuanda Avenue, Gaoxin North District, Changchun 130102, China

^c Jilin Province Science and Technology Innovation Center for Multimodal Cognitive Computing and Analysis of Medical Biometrics, 3000 North Yuanda Avenue, Gaoxin North District, Changchun 130102, China

^d Smart Health Joint Innovation Laboratory for the New Generation of AI, 3000 North Yuanda Avenue, Gaoxin North District, Changchun 130102, China

^e Department of Radiology, The First Hospital of Jilin University, Changchun, 130021, China

ARTICLE INFO

Keywords:

Computer vision

Lung cancer

Object detection

Medical imaging

ABSTRACT

Lung cancer, a global leading cause of incidence and mortality, demands accurate detection. Current research primarily targets early nodule detection and distinguishing benign from malignant tumors, with limited focus on lung cancer classification. Identifying various lung cancer types is vital for tailored treatments, while challenges persist in localizing small lesions and early tumors. To tackle these challenges, we propose the framework integrated the Channel Equally-Divided (CED) and the Coordinate Attention Feature Pyramid Network (CAFPN). CAFPN, an innovative feature pyramid structure, integrates the Semantic Information Enhanced (SIE) module and Coordinate Attention. The SIE module filters redundant semantic information, enriches texture features through Coordinate Attention, emphasizes shallow semantic details, and amplifies trustworthy specifics for enhanced deep semantic information. Additionally, the CED module is devised to proficiently extract local contextual information across channels for more precise feature representations. The superior performance of our proposed method was empirically validated through comparative experiments on two mainstream datasets with Competition Performance Metric (CPM) scores and Mean Average Precision (mAP) values reaching 0.942 and 99.18%, respectively, outperforming current state-of-the-art approaches.

1. Introduction

As stated in the 2022 Global Cancer Report, cancer has emerged as a significant health concern profoundly impacting human well-being and is a leading cause of human mortality [1]. In particular, the incidence and fatality rates of lung cancer are steadily escalating, becoming a prevalent contributor to fatalities related to malignant tumors across multiple nations [2]. The insidious nature of lung cancer, often characterized by inconspicuous early-stage symptoms, frequently leads to delayed diagnoses, resulting in a dismal five-year survival rate of merely 5% for advanced-stage cases [3,4]. Consequently, the timely identification, diagnosis, and treatment of lung cancer patients hold paramount significance [5,6]. Nonetheless, the manual review of Computed Tomography (CT) and Positron Emission Tomography (PET) images demands substantial time investment by radiologists [7],

as a single scan could encompass hundreds of slices. Moreover, the number of lesion voxel slices is typically less than 100 [8], and lung lesions exhibit pronounced variations in terms of shape, anatomical context, intensity, and size [9]. Therefore, the screening results of sophisticated Computer-Aided Diagnosis (CAD) systems have the potential to alleviate physicians' workloads engaged in lung cancer screening efforts [10]. Recently, deep learning-based artificial intelligence technologies have demonstrated noteworthy accomplishments in various cancer detection, including lung, breast, cervical cancer and so on [11–13]. Deep learning employs end-to-end training and testing methodologies through supervised learning, resulting in a substantial simplification of the model's training and testing procedures [14,15]. Deep Learning models have eliminated the need for manually extracting lung

* Corresponding author at: School of Computer Science and Engineering, Changchun University of Technology, 3000 North Yuanda Avenue, Gaoxin North District, Changchun 130102, China.

E-mail addresses: 2202103091@stu.ccut.edu.cn (S. Ma), luhumin@ccut.edu.cn (H. Lu), 2202103103@stu.ccut.edu.cn (H. Xue), 2202203031@stu.ccut.edu.cn (C. Lin), 2202303094@stu.ccut.edu.cn (Z. Zhang), 1202112006@stu.ccut.edu.cn (Y. Wang).

<https://doi.org/10.1016/j.bspc.2025.108510>

Received 24 January 2025; Received in revised form 17 June 2025; Accepted 2 August 2025

Available online 13 August 2025

1746-8094/© 2025 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

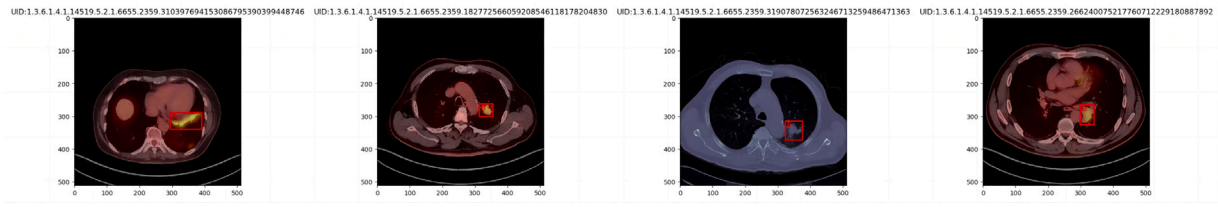


Fig. 1. Various types of lung tumor images, where “A” represents adenocarcinoma, “B” represents small cell carcinoma, “E” represents large cell carcinoma, and “G” represents squamous cell carcinoma.

nodule characteristics, autonomously acquire highly discriminative features that aligns seamlessly with the requirements of lung cancer CAD systems [16]. Furthermore, the scalability of deep learning algorithms with increased data volume has proven advantageous when analyzing extensive medical image datasets [17,18].

Lung cancer [19] is primarily categorized into: small cell lung cancer large cell carcinoma, adenocarcinoma, and squamous cell carcinoma, as illustrated in Fig. 1. Diverse lung cancer types exhibit substantial variations in treatment approaches [20]. Previous studies have primarily concentrated on distinguishing between benign and malignant tumors, overlooking the detection of specific types of lung cancer [21]. While detecting small targets, such as tiny lung cancer lesions or early tumors, is equally crucial. In this context, the emerging role of artificial intelligence has emerged as a potential solution to address these challenges. Artificial intelligence not only has the ability to manage vast medical data but can also learn from it and extract profound information. However, in numerous CT and PET images, certain cancers may exhibit similarities that could potentially confuse radiologists [22]. Therefore, it is crucial to design a lung cancer subtype detection model that assists radiologists in quickly locating the precise location of lung cancer and ascertaining its specific subtype, ensuring a more rapid adoption of appropriate treatment strategies. This endeavor holds considerable implications for lung cancer treatment, opening a new avenue for medical professionals to discern and delineate diverse lesions of varying shapes and sizes from CT and PET images.

Liu [23] proposed the STBi-YOLO method, which utilizes a bidirectional feature pyramid network for multiscale feature fusion and employs the EIoU function to optimize the model for accurate identification of pulmonary nodules in CT images. Jacob et al. [24] detected various types of lung cancer from multimodal images using a method based on the Blind/Referenceless image spatial quality evaluator to extract slices with pulmonary abnormalities. Zhou et al. [25] introduced a Cross-Modal Fusion Transformer Module, facilitating multimodal feature fusion to enhance the capability of extracting key lesion features and performing deep fusion. Sabzalian et al. [26] designed an enhanced form of the Ebola optimization search algorithm to optimize the Bidirectional Recurrent Neural Network (BRNN) in lung cancer detection. Barbouchi et al. [27] employed the DETR model for tumor detection, assisting doctors in staging lung cancer. They established a Transformer-based method for tumor localization and staging. Ji et al. [28] proposed ELCT-YOLO model, redesigning the neck network to reduce model dependencies between feature layers and incorporating complementary features with similar semantic information to enhance detection performance. Mammeri et al. [29] developed an algorithm capable of drawing boundary boxes around pulmonary nodules without losing any important information, thereby assisting radiologists in identifying and tracking nodules in adjacent CT scan slices.

Despite the aforementioned auxiliary methods of accurately locating tumor positions and visualizing the relative relationship between lesions and surrounding tissues, some challenges are still present. Current research on lung tumor detection primarily concentrates on detecting early lung nodules and discerning benign from malignant tumors. However, research on the detection of specific lung cancer types is relatively limited. This is a notable issue since different types of lung cancer

require corresponding approaches. Therefore, conducting in-depth research and developing detection methods for various lung cancer types is of great significance. Furthermore, although deep learning algorithms can learn and capture morphological differences significantly in tumor features, the precise localization and detection of specific small targets, such as tiny lung cancer lesions or early tumors, may still exceed the resolution capability of artificial intelligence, which is a challenge to the accuracy of detection in such cases. To tackle these challenges, we introduce an innovative single-stage detection model tailored to address the issue of variable sizes of lung tumors, which is called iC²Net. Firstly, to maintain tumor detection speed while minimizing hardware costs, we chose YOLOv7 as the underlying framework. Unlike two-stage detection models, YOLOv7 does not need to generate candidate boxes, and can strike a favorable balance between accuracy and tumor detection speed. Secondly, we devised the CED module to enhance spatial information by sensing the target’s local context with a multiscale receptive field. Subsequently, we designed a novel feature pyramid structure named the Coordinate Attention Feature Pyramid Network (CAFPN). Specifically, we design a new module called the Semantic Information Enhanced (SIE) module, which filters out redundant semantic information and enhances texture features by incorporating Coordinate Attention. This highlights shallow semantic information and amplifies credible details to emphasize deep semantic information.

The central contributions of our research are presented below:

- In this study, we propose an innovative CAFPN, designed to precisely localize and detect minuscule lesions using a limited number of pixels.
- Utilizing the CED module facilitates the parallel extraction of tumor image features from diverse channels, enabling the effective capture of local contextual information.
- SIE module generates texture-enhanced features by intelligently filtering out redundant semantic information. This module emphasizes smaller targets in the shallow layers while enhancing significant details to highlight larger targets in the deeper layers.
- The validity and superiority of our approach was validated on the LUNA16 and Lung-PET-CT-Dx datasets. On the LUNA16 dataset, our method achieves a 7.1% increase in CPM compared to the current state-of-the-art. On the Lung-PET-CT-Dx dataset, the proposed approach improves mAP by 1.8% over the best existing method.

2. Related work

With increasing adoption of deep learning within the medical and healthcare domains, numerous researchers of lung tumor detection have focused their attention on deep neural networks (DNN) [30]. Unlike conventional methodologies, DNN have distinctive traits, including a larger parameter space and an enhanced ability to capture semantic features [31]. Gong and colleagues harnessed deep residual networks to distinguish lung adenocarcinoma, achieving superior results compared to, radiologists [32]. Mei et al. conducted lung cancer detection experiments with a slice-aware network (SANet), which performed better than 2D and 3D CNN models, and also decreased the False Positive Rate (FPR) [33]. Xu et al. devised a Slice Group Domain Attention (SGDA) module, facilitating its seamless integration

into existing backbone networks to enhance the generalization ability of detection systems [34]. Su et al. combined Bag of Visual Words (BoVW) and Convolutional Recurrent Neural Networks (CRNN) to initially divide CT images into smaller nano-segments and then subjected them to deep learning-based classification [35]. Mousavi et al. adopted DNN-based detection approach for discerning COVID-19 and other lung infections [36], which extracts features from X-ray images using DNN, sequence modeling using LSTM networks, and classifies using a SoftMax classifier. This approach demonstrated significant effectiveness in the domain of COVID-19 detection, providing valuable assistance to radiologists in rapid diagnosis. Mei et al. adopted depth-wise over-parameterized convolutional layers to formulate residual units, thereby enhancing the network's feature representation capability [37]. They enriched the confidence and focal loss functions further to effectively address the pronounced imbalance sample problem encountered in the training process.

While numerous algorithms exist for lung tumors detection, only a few are tailored to handle the inherent variability in tumor size. Consequently, the varying sizes of lung tumors have become a new challenge in achieving precise tumor detection [38]. The multi-scale nature presents a challenge to the effectiveness of conventional detection methods, prompting researchers to explore and suggest improvement schemes for existing approaches. Causey et al. developed a lung cancer detection algorithm that integrates 3D convolution with Spatial Pyramid Pooling (SPP) [39]. In comparison to isolating the detection of 2D slices, 3D convolutions provide the advantage of capturing comprehensive spatial and volumetric information across adjacent slices, rendering them suitable to sparsely annotated datasets. However, it is worth noting that 3D convolutions demand more computational memory compared to traditional counterparts. Another approach pursued by researchers involved Feature Pyramid Networks (FPNs) [40], where the delineation of small tumors relies on the characteristics of shallow networks, while deeper networks contain richer semantic information, which is crucial for precise tumor differentiation. The primary goal of FPNs is recovering low-resolution details in deep feature maps and enhancing semantic content in shallow feature maps by connecting feature maps across different layers. To achieve effective multi-scale fusion, Guo and Bai developed FPN for multi-scale lung nodule detection [41]. Liu et al. improved the feature fusion structure of YOLOv5 by integrating a Bidirectional FPN (BiFPN), introducing a fusion pathway that connects features at the same level [23]. Several other enhancements to the feature fusion network have demonstrated positive impacts in various tasks. The initial FPN framework and its derivations utilize complex connections across scales to achieve a more robust multi-scale representation capability. However, it is worth noting that this operation necessitates more parameters and increases computational overhead [42].

3. Proposed method

3.1. Overall structure

Fig. 2 depicts the comprehensive structure of iC²Net. The initial input images undergo feature extraction via the CED module within the backbone network. The resulting features are labeled as feature layers, collectively constituting the feature set of the input tumor image. In the backbone segment, three feature layers are extracted to support the subsequent network construction, with each being labeled as an effective feature layer. The CAFPN additionally conducts feature fusion on these three effective feature layers to integrate feature information across different scales. The trio of enhanced effective feature layers (First, Second, and Third Layer) are fed into the SIE module to generate enriched semantic features. This procedure enhances the efficiency of tumor feature extraction, resulting in refined tumor features to fulfill the ultimate goals of classification and regression. Finally, different-sized target boxes are predicted through three YoloHead branches.

3.2. Channel equally-divided module

To allow neurons to adaptively acquire spatial tumor details from the original tumor images, we propose the Channel Equally-Divided (CED) module. Due to the low prevalence and subtle characteristics of minute lung cancer lesions or early tumors in images, the spatial information is often missing. The module uniformly distributes the input feature map into channel segments, executing feature extraction on these discrete channels. Additionally, it harnesses the cascading effect of several convolutional kernels to adeptly extract localized context from the diverse channels. Through the fusion of the extracted information, features related to the spatial environment of tiny lung cancer lesions or early tumors across multiple scales can be obtained. This module can establish context associations to obtain more abundant spatial information. As depicted in Fig. 3, initially, the CED module downsamples the original tumor image to generate a feature map denoted as F . This feature map is then evenly split across channels, resulting in the feature maps F_1 and F_2 . Concurrently, F_1 and F_2 undergo convolution operations. Additionally, by capitalizing on the cascading effect of multiple convolutions, the module effectively extracts local contextual information pertinent to tumors from both F_1 and F_2 . Then, all feature information is fused through skip connections to obtain detailed spatial information favorable for tumor detection. Finally, the fused feature map is subjected to downsampling to generate the final feature map $C1$.

Receptive Fields: The dimensions of receptive fields play a pivotal role in tumor detection, particularly for the identification of minute lung cancer lesions or early-stage tumors. The size of the region mapped onto the original image by the pixels on the feature map output from each convolutional neural network layer is defined as follows:

$$R_{i+1} = R_i + (k + 1) \times \prod_{j=1}^n S_j \quad (1)$$

where R represents the receptive fields, S represents the stride, R_{i+1} represents the $i + 1$ layer of receptive fields, R_i represents the convolution or pooling operation in the i layer, and k represents the size of the convolution kernel in the $i + 1$ layer. As elucidated by Eq. (1), after the operations of F , F_1 , and F_2 , the mapped area size of a pixel on the feature map to the original tumor image is determined as 3, 7, and 11, respectively. Thus, the parallel convolution operations within the CED module can capture local context information at multiple scales. Furthermore, the CED module utilizes a cascade of multi-convolution operations through equal division and fusion to capture rich local context information. These equal-division and fusion operations effectively optimize the time complexity of the CED module.

Equally-Divided: For any given feature map $F \in \mathbb{R}^{C \times H \times W}$, the Equally-Divided process are performed on the feature map. Following these operations, the feature map is evenly divided into two distinct parts as defined below:

$$\mathcal{N}_1 : F \rightarrow F_1 \in \mathbb{R}^{\frac{C}{2} \times H \times W} \quad (2)$$

$$\mathcal{N}_2 : F \rightarrow F_2 \in \mathbb{R}^{\frac{C}{2} \times H \times W} \quad (3)$$

where the operation \mathcal{N}_1 consists of *Conv2D*, Batch Normalization and SiLU activation function. The operation \mathcal{N}_2 is exclusively comprises *Conv2D*.

Fusion: Adaptively learning and fusing the multiscale local context information of tumors establishes spatial positional relationships between tumors and background. This process helps to obtain spatial information more favorable for both common tumors and small lung cancer lesions or early tumors. The steps involved in the computation of the fused feature map F'_3 is presented herein:

$$F'_3 = \alpha(\mathcal{N}(F_3)) + F \quad (4)$$

$$F_3 \in \mathbb{R}^{C \times H \times W} \quad (5)$$

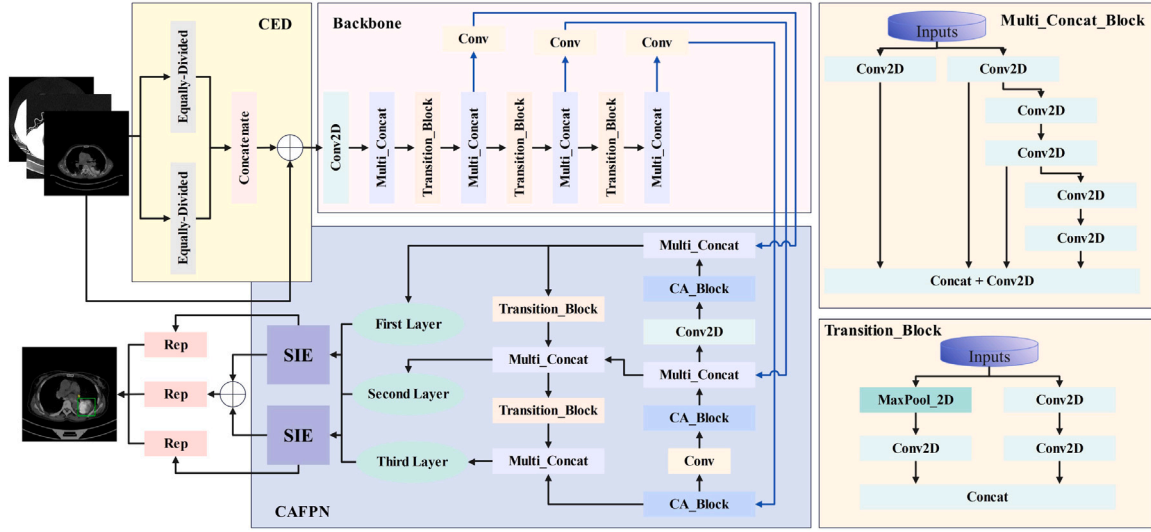


Fig. 2. Overall architecture of ic²Net.

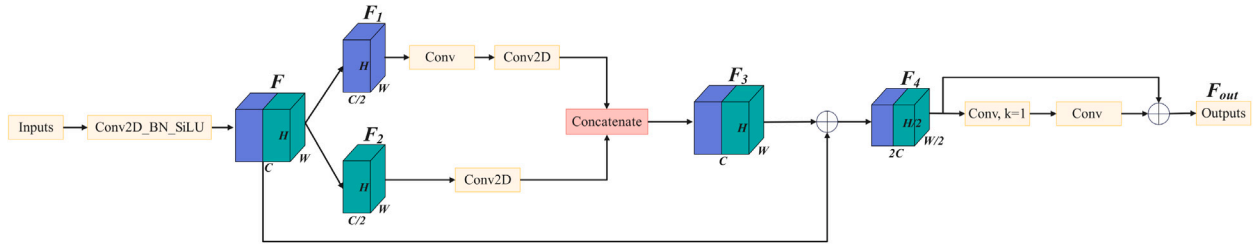


Fig. 3. CED module structure.

where $\alpha(\cdot)$ and $\mathcal{X}(\cdot)$ respectively represent the SiLU activation function and Batch Normalization. F_3 is the feature map obtained following parallel convolution and concatenation operations of F_1 and F_2 . It is defined as:

$$F_3 = \text{Con}[\mathcal{N}_1, \mathcal{N}_2] = \text{Con}[C(y(F_1)), C(F_2)] \quad (6)$$

$$y(\cdot) = \alpha(\mathcal{X}(C(\cdot))) \quad (7)$$

where $\text{Con}[\cdot]$ represents concatenation, and $C(\cdot)$ represents standard convolution operation.

After the equal-division and fusion procedures, multiscale local context information from the same layer are integrated to the CED module, and spatial information is enriched to obtain the fused feature map F_4 . Next, downsample operation of the feature map F_4 is performed to reduce its resolution and undergoes further feature extraction. Subsequently, the output feature map F_{out} is acquired. Its definition is as follows:

$$F_4 = y(F_3') \quad (8)$$

$$F_{out} = F_4 + y(y(F_4)) \quad (9)$$

Finally, the CED module extracts and integrates multiscale local context information through operations such as equal-division and fusion, resulting in a feature map F_{out} enriched with spatial information. This operation aims to enhance the tumor detection capabilities of the proposed model.

As illustrated in Fig. 1, the CED module evenly divides the input feature map under channel dimension and performs feature extraction on these segregated channels, enhancing the model's sensitivity to local contextual information. This ensures that the model can capture multi-scale local contextual information. This effectively addresses the

issue of missing spatial information in the identification of minute lung cancer lesions or early tumors encountered by existing methods, thereby enhancing the model's detection capabilities.

3.3. Coordinate attention feature pyramid network

Although existing attention mechanisms have proven effective in enhancing model performance, they tend to neglect the more critical positional information and spatial structure for visual detection tasks. For instance, Squeeze-and-Excitation Attention only focuses on attention in the channel dimension, overlooking attention in the spatial dimension. Efficient Channel Attention has specific limitations in handling global context dependencies and channel spatial relationships, resulting in decreased accuracy in detecting tiny lesions with fewer pixels. Therefore, we integrate FPN with Coordinate Attention, which incorporates positional information into channel attention, enhancing the model's localization and detection capabilities for small targets such as tiny lung cancer lesions or early tumors.

The Coordinate Attention mechanism decomposes channel attention into two one-dimensional feature representations, traversing two spatial directions, and subsequently aggregates features across the spatial dimension. This strategy adeptly captures remote dependencies in a single spatial direction and simultaneously retains exact location details in the perpendicular spatial dimension. In a subsequent step, the derived feature maps are independently encoded into a pair of attention maps, each being aware of directionality and finely tuned to positional sensitivity. These maps are jointly applied to the input feature maps, collectively enhancing the representation of interested objects. The specific architectural configuration is illustrated in Fig. 4.

Firstly, the input feature map is encoded using pooling kernels of sizes $(H, 1)$ and $(1, W)$ along the horizontal and vertical coordinate directions, respectively, this applies to every channel. Where H, W

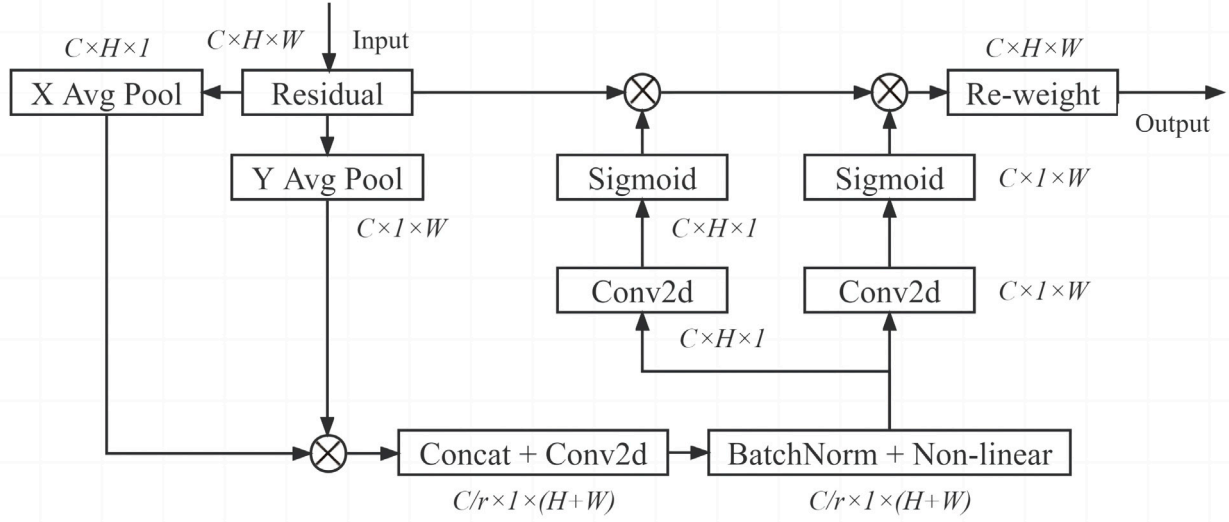


Fig. 4. Structure of coordinate attention.

denote the height and width of the input feature, the output expression for the channel with height is as follows:

$$z_h^c(h) = \frac{1}{W} \sum_{0 \leq i < W} x_c(h, i) \quad (10)$$

where $z_h^c(h)$ is the output for the c channel with height h , and $x_c(h, i)$ represents the feature vector of the row i .

Similarly, the output expression for the c channel with a width of w is given by:

$$z_w^c(w) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w) \quad (11)$$

where z_w^c is the output for the c channel with width w , and $x_c(j, w)$ represents the feature vector of the column j .

Next, the Coordinate Attention is generated. The outputs from the two spatial directions are concatenated, and a 1×1 convolution is applied for the F_1 transformation operation, expressed as:

$$f = \delta(F_1([z^h, z^w])) \quad (12)$$

where f represents the intermediate feature map that encodes spatial information across both the horizontal and vertical dimensions. δ represents non-linear activation function, F_1 is a 1×1 convolutional transformation, and z^h, z^w is the encoding output.

Then, f is split into 2 separate tensors, each of which is transformed using a 1×1 convolution. And the attention weight matrix is obtained by a Sigmoid activation function:

$$g^h = \sigma(F_h(f^h)) \quad (13)$$

$$g^w = \sigma(F_w(f^w)) \quad (14)$$

where f^h and f^w are the tangent tensor of f in the horizontal and vertical directions. F_h and F_w is a 1×1 convolutional transformation, g^h and g^w are the attention weight matrices along the horizontal and vertical directions, respectively, and σ represents Sigmoid function.

The final output $y_c(i, j)$ is given by:

$$y_c(i, j) = x_c(i, j) g_c^h(i) g_c^w(j) \quad (15)$$

where $x_c(i, j)$ represents input feature map, $g_c^h(i)$ and $g_c^w(j)$ represents the attention maps along the horizontal and vertical directions.

Finally, the acquired sensitivity, enhanced with tumor position information, is obtained. Through the top-down structure of the FPN, the improved layers are aggregated, which improves the localization and detection of small targets like small lung cancer lesions or early-stage tumors.

The sensitivity of tumor location information was enhanced by finally acquired $y_c(i, j)$. And $y_c(i, j)$ of each layer are polymerized through the top-down structure of the FPN, which improved the effectiveness of localization and detection of small targets, such as small lung cancer lesions or early-stage tumors.

Therefore, CAFPN is capable of simultaneously capturing long-range dependencies and preserving precise positional information, compensating for the deficiencies of traditional attention mechanisms in the spatial dimension. Through the aggregation of the FPN architecture, it enhances the comprehensive utilization of feature maps at various levels, enabling the model to excel in multi-scale lung tumor detection.

3.4. Semantic information enhanced module

The Feature Pyramid Network (FPN) leverages its inherent feature layer structure to manage feature maps with diverse resolutions. To aggregate multiscale context information, FPN integrates deep semantic insights with shallow details through lateral connections in the top-down pathway. We denote the shallow layer in two adjacent pyramid layers as P_i , and the deep layer as P_{i+1} . In FPN, shallow features P_i are responsible for detecting small objects, while deep features P_{i+1} manage larger counterparts. However, shallow features P_i often encompass unreliable noise in region details, hindering the detection of minute lung cancer lesions or nascent tumors. Furthermore, deep features P_{i+1} harbor certain information pertinent to the detection of small tumors. However, since they are crucial in normal tumor detection, the detector may face challenges in distinguishing the ambiguous semantics between small tumors and their normally sized counterparts.

To overcome challenges in tumor detection, we propose the Semantic Information Enhanced module (SIE). In the SIE module, we extract texture features of tumors and simultaneously capture richer and more abstract information from the tumor region in the image to enhance its semantic information expression. The main objective is to enhance semantic information by minimizing unnecessary redundancies. This technique is crucial for emphasizing small tumors in shallow layers and highlighting regular tumors in deep layers. By infusing reliable details, this approach contributes to improved overall tumor detection efficiency. The SIE module is intricately crafted to seamlessly fuse distinct feature layers, specifically the First and Second Layer, followed by the Second and Third Layer. This design aims to mitigate the presence of regular-sized tumor data in the shallow layers while simultaneously enhancing their significance in the deep layers. This is accomplished by subtracting deep features from their shallow counterparts, effectively

eliminating regular-sized tumor data. Meanwhile, the shallow features are integrated into the deep layers to enhance the comprehensive feature context.

Accurately representing small tumors within shallow features poses a challenge due to significant background noise, which diminishes the discriminative capability of shallow features in detecting these small tumors. While standard atrous convolution can expand the receptive field without adding parameters, there is a limitation that: its output maintains a fixed receptive field size. To overcome this limitation and enhance the representation of shallow features, we propose using atrous convolutions with different rates. This approach enables us to control the receptive field for small tumors located in shallow regions. Furthermore, when considering the differences between deep and shallow features in the context of feature fusion, directly applying addition or concatenation operations could exacerbate gradient computations during training and undermine the effectiveness of feature fusion. To tackle this issue, we utilize the Hadamard product to enhance the feature maps of shallow small tumors, which enables more precise encoding with less noise before performing subtraction and summation operations within the SIE module.

Fig. 5 illustrates the structure of the SIE module, which receives two input features, namely shallow feature maps P_i and high-level feature maps P_{i+1} . These input features are used to generate the output feature maps P'_i and P'_{i+1} of the SIE module, defined as follows:

$$P'_i = P_r \ominus (P_r \odot P_c) \quad (16)$$

where P_r represents the enhanced features from the shallow layers, achieved by utilizing atrous convolutions with dilation rates of 1, 2, and 5 (verified through experimentation) for more accurate representation of small tumors. P_c represents the content extractor, which was used to capture the semantic content of the input features P_i , by suppressing low-resolution features dispersed in informative high-resolution features.

To obtain feature maps P_c , P_i are upsampled to match the P_i size. Subsequently, the upsampled feature maps are then passed through a convolutional layer of 1×1 and a sigmoid function, generating the P_c following:

$$P_c = \varphi(P_{i+1}) \quad (17)$$

where φ consists of upsampling, convolutional layer of 1×1 , and sigmoid function.

The definition of the output feature map P'_{i+1} from the high-level layers is as follows:

$$P'_{i+1} = \mathcal{M}(P_r \otimes P_c) \oplus P_{i+1} \quad (18)$$

where \mathcal{M} is a downsampling operation used to maintain coherent spatial dimensions with P_{i+1} . After conducting feature extraction in the SIE module, the obtained features P_i ($i = 2, 3, 4, 5$) are further processed in the succeeding stage.

The SIE module enhances the accuracy of minute tumor detection using shallow features by mitigating noise and augmenting the feature representation of small tumors. It also enhances the information pertaining to normally sized tumors in deep features and reduces confusion with minute tumors, thereby improving overall detection performance. This addresses the issues of excessive noise in shallow features and ambiguous semantics in deep features present in existing methods. By refining shallow feature representation, optimizing high-level features, and flexibly controlling the receptive field, it significantly boosts the model's detection capabilities for minute lung cancer lesions or nascent tumors.

4. Experiments

4.1. Datasets

4.1.1. LUNA16

The LUNA16 challenge [43] employed CT scans from the public Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) dataset, selecting 888 scans with a slice thickness of 2.5 mm or less, a criterion now deemed standard. Lung nodule annotations in the LIDC-IDRI dataset were provided by four experienced thoracic radiologists through a two-phase process. The challenge adopted a reference standard that included all nodules with diameters of at least 3 mm, annotated by a minimum of three radiologists. Non-nodules, smaller nodules under 3 mm, and those marked by one or two radiologists were excluded from the false positive count, classified instead as irrelevant findings.

4.1.2. Lung-PET-CT-Dx

This study utilized a dataset comprising CT and PET-CT DICOM images from 355 patients with suspected lung cancer who underwent standard-of-care lung biopsy and PET/CT [44]. Images were retrospectively acquired and annotated with tumor location using bounding boxes within XML files. Patients were categorized based on histopathological diagnosis: Adenocarcinoma ("A"), Small Cell Carcinoma ("B"), Large Cell Carcinoma ("E"), and Squamous Cell Carcinoma ("G"). The dataset encompasses 251,135 images across both modalities.

Five experienced thoracic radiologists, two with over 15 years and three with over 5 years of expertise in lung cancer, annotated tumor locations. An initial annotation by one radiologist was subsequently verified by the remaining four, ensuring a robust and reliable dataset for algorithm development in medical diagnosis.

4.1.3. Pre-processing

We conducted preprocessing on two datasets separately. For the LUNA16 dataset, we initially performed thresholding, using -600 as the threshold to filter out water and air. The remaining part underwent dilation to fill small holes inside the lungs, generating a mask. Subsequently, voxel values were truncated to the range of -1200 to +600. Values below -1200 were set to -1200, and values above +600 were set to +600. The data was then normalized to the range of 0-255 with the mask applied subsequently. After cleaning and transformation, the dataset comprised of 3782 images containing nodules, corresponding to 888 CT scans. For the Lung-PET-CT-Dx dataset, an initial step involved organizing the data and discarding incorrect annotation files. As not all XML files have corresponding original images, those XML files without corresponding images were excluded from training. We identified and discarded annotation files that pointed to the same target image in both annotation files. Then, we converted the PETCT's three-channel images into the single-channel image format of conventional CT scans. Finally, we reconstructed the cleaned data into a new dataset with a total of 30,884 images.

4.2. Experimental setup and evaluation metrics

During the training of our method, we divided the two datasets into (training set + validation set): test set in a ratio of 9:1, and the training set: validation set in a ratio of 9:1. The experimental parameters and experimental environment are shown in Table 1.

LUNA16: 3063 training images, 340 validation images, and 379 test images.

Lung-PET-CT-Dx: 25,014 training images, 2780 validation images, and 3090 test images.

The LUNA16 challenge evaluated nodule detection performance using the Free-response Receiver Operating Characteristic (FROC), which considers both sensitivity and the average number of false positive (FP) nodules per scan. Higher average sensitivity across seven FP (1/8, 1/4,

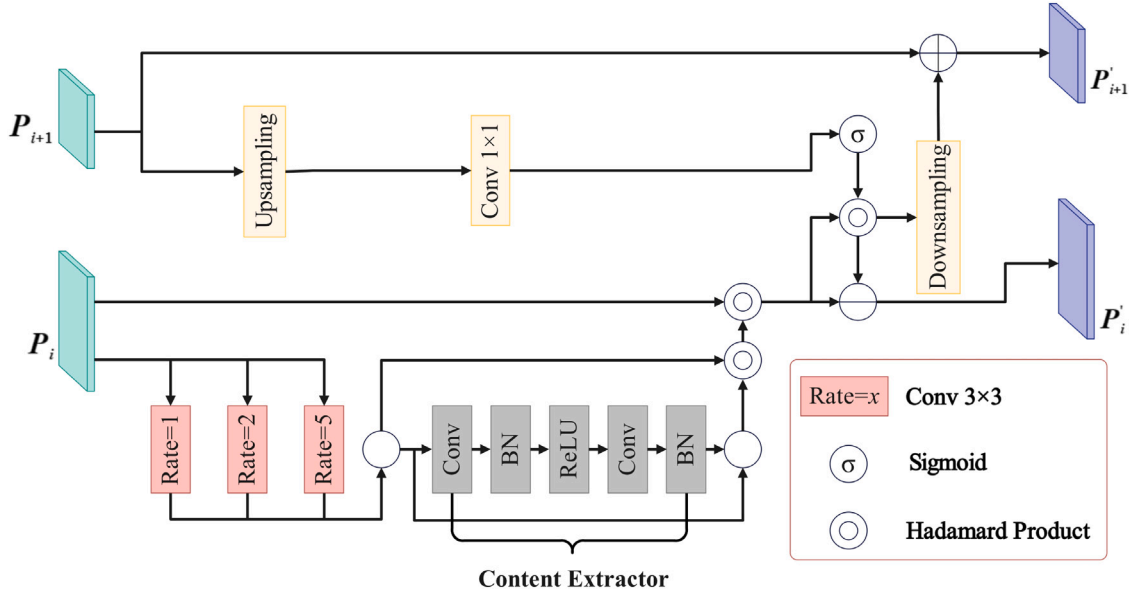


Fig. 5. SIE module structure.

Table 1

Experimental environment and parameter settings.

CPU	Intel Xeon(R) Silver 4110 CPU @ 2.10 GHz
GPU	Two RTX 3090 GPUs
RAM	32 GB
Framework	Pytorch
Optimizer	Adam
Momentum	0.9
Learning rate	0.001
Epochs	300

1/2, 1, 2, 4, and 8) indicates better model performance (Competition Performance Metric, CPM). In addition, we assessed models using Precision, Recall, mean Average Precision (mAP), and F1-score. Their definitions are as follows:

$$Precision = \frac{TP}{TP + FP} \quad (19)$$

$$Recall = \frac{TP}{TP + FN} \quad (20)$$

$$F_1 - score = \frac{2TP}{2TP + FP + FN} \quad (21)$$

$$AP = \int_0^1 P(Recall) dRecall \quad (22)$$

$$mAP = \frac{\sum_{i=1}^N AP_i}{N} \quad (23)$$

4.3. Performance comparison

LUNA16. In recent years, numerous lung nodule detection systems have been proposed and demonstrated commendable performance on the LUNA16 dataset. Khosravan et al. proposed the single-scale model S4ND, founded on dense blocks, and attained superior CPM scores [45], but it exhibited poor sensitivity at 0.125 FP per scan. This deficiency is caused by S4ND's absence of fusion techniques such as FPN. The execution of maximal pooling operations results in the loss of nodule information. Consequently, the model faces challenges in distinguishing between actual nodules and nodule-like tissue. Wang et al. introduced an innovative model for rapid detection of candidates in 3D chest CT scans. The essence of this model lies in its nodule size-adaptive ability, capable of detecting nodules of diverse locations,

types, and sizes for 3D images, achieving a CPM value of 0.903. However, as the majority of FNs are screened during candidate selection and disregarded in FP detection, Wang's method could potentially lead to the partial omission of nodules. In addition to traditional nodule detectors, Table 2 references nodule detection models based on self-encoders that demonstrate comparably elevated detection performance. For instance, Tang et al. [46] developed the NoduleNet model with feature mapping decoupling, while Mei et al. [33] proposed a self-encoder-based model, SANet, specifically tailored for nodule detection. These outcomes demonstrate significant similarities in their utilization of ResBlock for feature map acquisition. However, relying solely on feature maps may not be sufficient for achieving exemplary model performance. Zhu et al. [47] introduced the high-resolution feature extraction network (HRNet) that incorporates a feature pyramid (AC-FPN) to enable the fusion of multi-scale features and facilitate multi-level predictive capabilities. However, the model encounters challenges in confidently recognizing specific nodules with irregular shapes or abnormal sizes. Recognizing the limitations of the aforementioned methods, we propose an integration method that combines channel division and coordinate attention feature pyramid. Comparative experiments with mainstream models were conducted, and the results are presented in Table 2. It can be observed that our proposed method achieved the highest CPM, reaching 0.933. Sensitivity is also the highest at 0.125, 0.25, 0.5, 1, 2, and 4 FPs/scan, with values of 0.862, 0.896, 0.937, 0.948, 0.958, and 0.963, respectively. This demonstrates the comprehensive performance of our proposed model, maintaining relatively high Sensitivity for most false positive rates, reaching the SOTA level. However, at 8 FPs/scan, the results did not surpass HRNet, attributed to HRNet's incorporation of adversarial training, which reduces the FPR of candidates and results in a relatively high Sensitivity.

Lung-PET-CT-Dx. Despite the strong performance of various SOTA algorithms on the Lung-PET-CT-Dx dataset, they tend to overlook the detection of microscopic tumors or early-stage lung cancer lesions, leading to a degradation of model performance. Liu proposed an improved YOLOv5 algorithm for lung nodule detection. He combined the random pooling of multiple convolutional layers, multi-scale fusion, and the EIoU loss function, enhancing the precision of lung nodule localization. However, a single-mode image is influenced by factors such as data distribution, task nature, and specific requirements of the problem, making it unsuitable for all scenarios. Jacob et al. and Zhou et al. conducted studies on the detection of various lung cancer types from

Table 2
Performance comparison on LUNA16 datasets.

Method	Year	False positives per scan(FPs/scan)							
		0.125	0.25	0.5	1	2	4	8	CPM
Nodulenet [46]	2019	0.708	0.783	0.857	0.900	0.943	0.955	0.963	0.873
SANet [33]	2021	0.712	0.802	0.865	0.901	0.937	0.946	0.955	0.874
Wang et al. [48]	2019	0.788	0.847	0.895	0.934	0.952	0.959	0.963	0.903
Vit-Bayesian [49]	2023	–	–	–	–	–	–	–	0.909
HRNet [47]	2023	0.850	0.888	0.927	0.943	0.952	0.961	0.971	0.927
S4ND [45]	2018	0.709	0.836	0.921	0.953	0.953	0.953	0.953	0.897
3D-CenterNet [32]	2020	0.784	0.847	0.906	0.938	0.950	0.955	0.961	0.906
SCPM-Net [50]	2022	0.743	0.829	0.889	0.922	0.939	0.958	0.964	0.892
FFNET [51]	2023	0.711	0.792	0.858	0.899	0.928	0.949	0.961	0.871
MS framework [52]	2022	0.742	0.841	0.896	0.925	0.944	0.946	0.960	0.895
iC²Net	2023	0.862	0.896	0.937	0.948	0.958	0.963	0.969	0.933

Table 3
Performance comparison on Lung-PET-CT-Dx datasets.

Methods	Year	Precision	Recall	F1-Score	FPS	mAP (%)
TL-YOLO [29]	2023	0.827	0.747	0.785	16.41	81.280
ELCT-YOLO [28]	2023	0.923	0.951	0.936	95.00	97.400
STBi-YOLO [23]	2022	–	0.933	–	27.00	95.900
DETR [27]	2023	0.930	0.960	0.940	–	–
Multimodality CNN [24]	2022	0.946	0.946	0.946	61.71	–
Bidirectional Recurrent [26]	2023	0.974	0.953	0.962	28.00	–
CCGL-YOLOv5 [25]	2023	0.977	0.974	0.976	98.59	96.670
iC²Net	2023	0.9908	0.9901	0.9875	98.75	99.183

multimodal images, specifically focusing on adenocarcinoma, squamous cell carcinoma, and small cell carcinoma. Their research aimed to leverage multimodal imaging for enhanced cancer identification. However, it is that fluctuations in threshold values or alterations in scan resolution may significantly impact the performance of multimodal detection systems. Sabzalian et al. refined BRNN and metaheuristic techniques and implemented a series of preprocessing techniques to simplify the diagnosis process in lung cancer. However, their proposed DESO increases its complexity compared with the traditional BRNN model. Therefore, in order to solve the problems in the above model, we carried out comparative experiments. The experimental results are shown in Table 3. Our proposed method is superior to the current best ELCT-YOLO method, and the mAP index is increased by 2.916%. Our method demonstrates a notable proficiency in precisely detecting tumor lesions of varying sizes. This capability holds particular importance in facilitating timely treatment and prognosis assessments. Subsequent investigations and experiments will contribute to our comprehensive understanding of this domain and further improvement of our model for enhanced performance.

4.4. Ablation studies

Effect of Channel Equally-Divided Module. As shown in Tables 4, 5, the Lung-PET-CT-Dx dataset demonstrated a 2.290% increase in mAP value when utilizing the CED module, compared to the baseline. This finding confirms the capability of the CED module to enhance the spatial representation of tumor features to a certain degree and mitigate the degradation of spatial information during feature transmission. Consequently, this augmentation enhances tumor detection capabilities. Upon integrating the SIE module with the baseline and CED module, the mAP values experienced enhancements of 3.687% and 1.397% relative to the baseline and the model incorporating solely the CED module, respectively. Consequently, the fusion of these two modules substantially enhances the model's efficacy by amalgamating multi-scale feature maps that are rich in semantic and spatial details. This integration notably boosts the accuracy in identifying both typical and diminutive tumors. In the context of ablation studies, the CED module demonstrates a marked enhancement in tumor detection accuracy when it is integrated with the SIE module. The results from ablation

Table 4
Ablation experiments of channel without divided and Channel Equally-Divided.

Methods	Precision	Recall	F1-score	mAP (%)
Channel without divided	0.9854	0.9846	0.9752	98.934
Channel Equally-Divided (CED)	0.9908	0.9901	0.9875	99.183

Table 5
Ablation experiments with the Lung-PET-CT-Dx test set.

No.	CAFPN	CED	SIE	Precision	Recall	F1-score	mAP (%)
1	–	–	–	0.9524	0.9516	0.9496	95.182
2	✓	–	–	0.9611	0.9607	0.9567	96.575
3	–	✓	–	0.9708	0.9699	0.9652	97.472
4	–	–	✓	0.9731	0.9725	0.9674	97.630
5	✓	✓	–	0.9886	0.9881	0.9834	98.910
6	✓	–	✓	0.9836	0.9824	0.9729	98.779
7	–	✓	✓	0.9858	0.9849	0.9753	98.869
8	✓	✓	✓	0.9908	0.9901	0.9875	99.183

analyses indicate that the combined operation of these two modules surpasses the performance of either module functioning independently, leading to superior tumor detection capabilities.

Effect of Coordinate Attention Feature Pyramid Network and Semantic Information Enhanced Module. Despite the shallow FPN layer's role in detecting minute tumors and the deep layer's role in detecting typical tumors, these layer remain susceptible to noise and unreliable intricate features. CAFPN and SIE can produce reliable detailed tumor features, thus enhancing the precision of tumor category detection. As shown in Table 5, CAFPN and SIE enhance the mAP of the baseline approach in tumor category detection by 1.393% and 2.448%, respectively. These results indicate that by reducing redundant semantics to attract attention to small tumors in the shallow layers, CAFPN and SIE modules can focus attention on tiny tumors in the superficial layer and normal tumors in the deep layer. Specifically, to suppress surplus regions in normal tumors within the shallow layer and minute tumors within the deep layer, CAFPN and SIE enhance the tumor feature representation via the application of atrous convolutions.

Effect of different dilation rates of atrous convolution. To further assess the effectiveness of distinct dilation rates in the context of tumor species detection, we conducted experiments to evaluate comprehensive performance outcomes. The baseline in Table 5 is the result trained by the original YOLOv7 model. The baseline in Table 6 is the SIE module without atrous convolution processing and directly inputting data into the Content Extractor for content extraction. Accurate tumor representation within the feature map is essential to avoid redundant semantic data. Leveraging atrous convolutions aids in augmenting the feature depiction of tumors, rendering them more discernible for detection purposes. In Table 6, various atrous convolutions employing dilation rates spanning from 1 to 5 are examined. This exploration is conducted to tailor the receptive field of the tumor via the utilization of 3×3 convolution kernels. Experimental results indicate that superior performance is attainable with dilation rates of

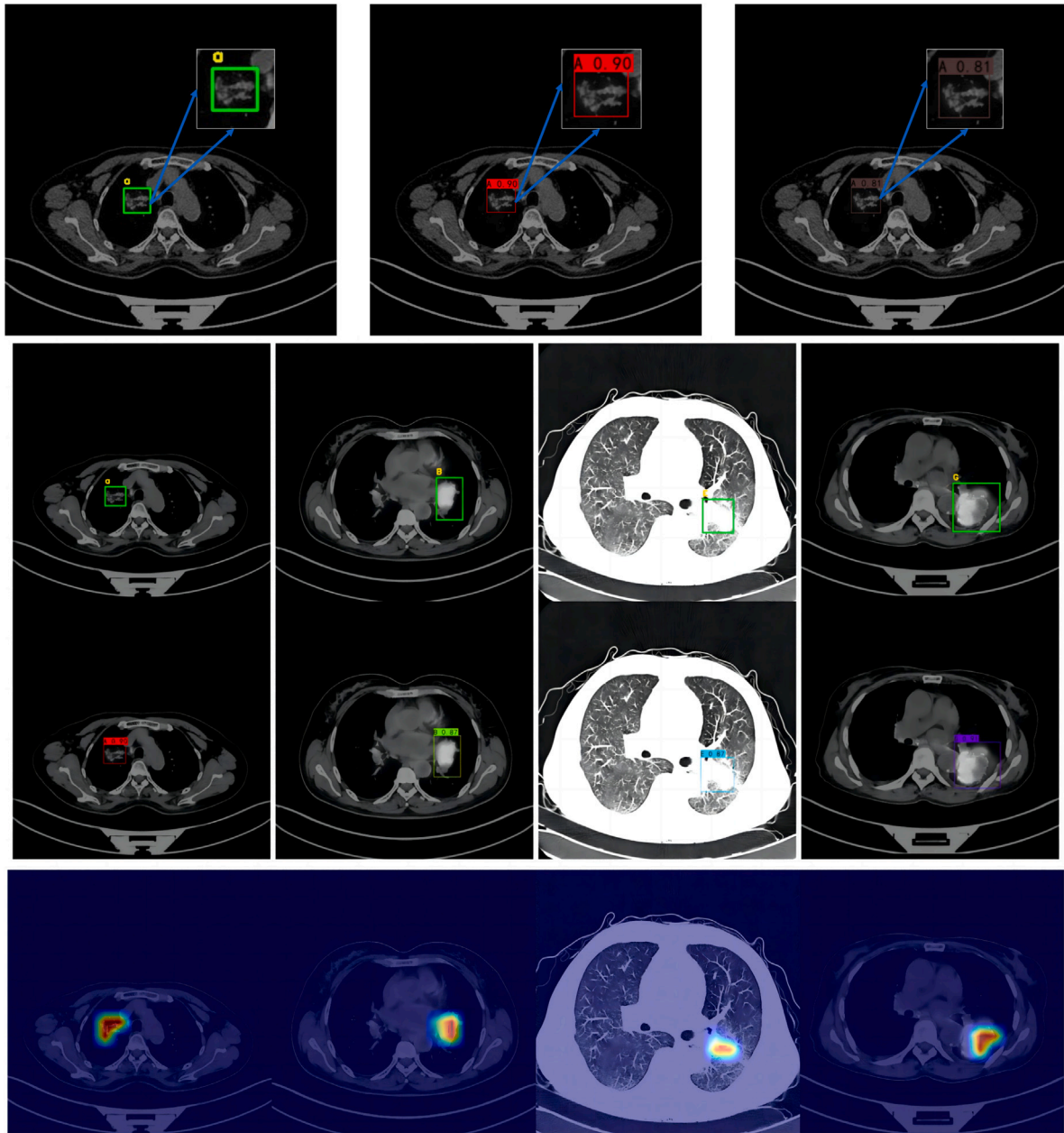


Fig. 6. Ground truth and detection results.

1, 2, and 5. This pattern possibly arises from our intent to prioritize not only the identification of superficial minute tumors but also prominent typical tumors. Consequently, atrous convolution compositions with smaller dilation rates (1, 2, and 3) as well as those with larger dilation rates (4 and 5) are unnecessary to adapt the receptive field. This pattern also signifies that minute tumors exhibit heightened sensitivity to the localized contextual details.

4.5. Qualitative results

In Fig. 6, the ground truth is denoted by the green rectangular box, while the detection results for the Lung-PET-CT-Dx dataset by the proposed method and other comparative methods are represented by rectangles in other colors. According to observations, our method successfully detected all tumors and accurately identified the types of all tumors. Compared to other detection methods, our mAP value is the highest. We enlarged the detection boxes of the images to visualize

Table 6

The effectiveness of different dilation rates on Semantic Information Enhanced (SIE) on Lung-PET-CT-Dx.

Method	Dilation					mAP (%)
	1	2	3	4	5	
Baseline	–	–	–	–	–	97.513
SIE	✓	✓	✓			98.857
	✓	✓		✓		98.532
	✓	✓			✓	99.183
	✓		✓	✓		98.182
	✓		✓		✓	98.267
	✓			✓	✓	98.789

the results, making it clear that our method aligns more closely with nodules and produces results closer to the ground truth. This indicates that our method fits well with the nodules. Finally, we chose to use heatmaps for visualization. The color mapping in the heatmap is based

on data values, where dark colors typically represent higher values, and light colors represent lower values. From the heatmap, it is apparent that the detection results of our method are concentrated in the nodular areas. This comprehensive evidence demonstrates how effectively our method detects the types of tumors.

5. Conclusion

In this research endeavor, we proposed an advanced model, iC²Net, designed for the efficient detection and categorization of both regular-sized lung tumors and early-stage small lung cancer lesions within CT and PET images. Deviating from conventional methods, our model leverages multi-scale tumor data present in the images. To tackle this issue, we implemented the innovative Coordinate Attention Feature Pyramid Network. By integrating the Semantic Information Enhanced (SIE) module, we dynamically refined texture features, removing unnecessary semantic data to emphasize both shallow and deep tumor information. Moreover, to improve the detection of lung cancer features, we proposed the Channel Equally-Divided (CED) module. This module is devised to extract localized contextual information across diverse channels, thus enhancing the feature map representation capabilities. We validated the performance of our proposed model through rigorous evaluation on the LUNA16 and Lung-PET-CT-Dx datasets, revealing outstanding detection accuracy that surpasses existing benchmarks. Looking ahead, we are committed to refining our model, with a focus on minimizing its computational requirements without sacrificing accuracy. Furthermore, we aim to explore emerging technologies further to enhance lung cancer detection. However, our model may exhibit suboptimal performance in certain specific scenarios or datasets, such as when dealing with specific types of lung cancer variants. Additionally, the model's performance in practical applications may be influenced by different scanning devices and parameter settings, requiring broader validation. In future work, we plan to refine the model through lightweight improvements, aiming to reduce the model's parameter count and computational load while maintaining detection accuracy. Furthermore, we will delve into emerging technologies for lung cancer detection, exploring more advanced and efficient methods to adapt to the continually evolving field of medical imaging. This may involve integrating deep learning with other advanced techniques, such as transfer learning or reinforcement learning, to further enhance the model's performance and generalization capabilities.

CRedit authorship contribution statement

Songzhe Ma: Writing – original draft, Software, Methodology, Investigation, Conceptualization. **Huimin Lu:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Formal analysis, Data curation. **Han Xue:** Writing – review & editing, Data curation. **Chenglin Lin:** Visualization, Data curation. **Zexing Zhang:** Visualization, Data curation. **Yilong Wang:** Writing – review & editing, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research is supported by the the Industrial Technology Research and Development Special Project of Jilin Provincial Development and Reform Commission in 2023 (No. 2023C042-6).

Data availability

Data will be made available on request.

References

- [1] Z. Ren, Y. Zhang, S. Wang, A hybrid framework for lung cancer classification, *Electron.* 11 (10) (2022) 1614.
- [2] M.P. Roy, Factors associated with mortality from lung cancer in India, *Curr. Probl. Cancer* 44 (4) (2020) 100512.
- [3] W.H. Organization, et al., Coronavirus disease 2019 (COVID-19) situation report-51, 2020.
- [4] N. Kalaivani, N. Manimaran, S. Sophia, D. Devi, Deep learning based lung cancer detection and classification, in: *IOP Conference Series: Materials Science and Engineering*, Vol. 994, IOP Publishing, 2020, 012026.
- [5] Z. Zhu, L. Liu, R.C. Free, A. Anjum, J. Panneerselvam, OPT-CO: Optimizing pre-trained transformer models for efficient COVID-19 classification with stochastic configuration networks, *Inform. Sci.* 680 (2024) 121141.
- [6] S. Huang, J. Yang, N. Shen, Q. Xu, Q. Zhao, Artificial intelligence in lung cancer diagnosis and prognosis: Current application and future perspective, in: *Seminars in Cancer Biology*, Elsevier, 2023.
- [7] Y. Xiao, J. Wu, Z. Lin, Cancer diagnosis using generative adversarial networks based on deep learning from imbalanced data, *Comput. Biol. Med.* 135 (2021) 104540.
- [8] M. Li, X. Ma, C. Chen, Y. Yuan, S. Zhang, Z. Yan, C. Chen, F. Chen, Y. Bai, P. Zhou, et al., Research on the auxiliary classification and diagnosis of lung cancer subtypes based on histopathological images, *Ieee Access* 9 (2021) 53687–53707.
- [9] M.Z. Islam, R.A. Naqvi, A. Haider, H.S. Kim, Deep learning for automatic tumor lesions delineation and prognostic assessment in multi-modality PET/CT: A prospective survey, *Eng. Appl. Artif. Intell.* 123 (2023) 106276.
- [10] K.S. Pradhan, P. Chawla, R. Tiwari, HRDEL: High ranking deep ensemble learning-based lung cancer diagnosis model, *Expert Syst. Appl.* 213 (2023) 118956.
- [11] M. Usman, A. Rehman, A. Shahid, S. Latif, Y.-G. Shin, MEDS-Net: Multi-encoder based self-distilled network with bidirectional maximum intensity projections fusion for lung nodule detection, *Eng. Appl. Artif. Intell.* 129 (2024) 107597.
- [12] Z. Zhu, S.-H. Wang, Y.-D. Zhang, A survey of convolutional neural network in breast cancer, *Comput. Model. Eng. Sci.: CMES* 136 (3) (2023) 2127.
- [13] S.M. Abd-Alhaleem, H.S. Marie, W. El-Shafai, T. Altameem, R.S. Rathore, T.M. Hassan, Cervical cancer classification based on a bilinear convolutional neural network approach and random projection, *Eng. Appl. Artif. Intell.* 127 (2024) 107261.
- [14] Z. Ren, Q. Lan, Y. Zhang, S. Wang, Exploring simple triplet representation learning, *Comput. Struct. Biotechnol. J.* 23 (2024) 1510–1521.
- [15] Y. Sun, J. Guo, Y. Liu, N. Wang, Y. Xu, F. Wu, J. Xiao, Y. Li, X. Wang, Y. Hu, et al., METnet: A novel deep learning model predicting MET dysregulation in non-small-cell lung cancer on computed tomography images, *Comput. Biol. Med.* (2024) 108136.
- [16] S. Dlamini, Y.-H. Chen, C.-F.J. Kuo, Complete fully automatic detection, segmentation and 3D reconstruction of tumor volume for non-small cell lung cancer using YOLOv4 and region-based active contour model, *Expert Syst. Appl.* 212 (2023) 118661.
- [17] Z. Zhu, S. Wang, Y. Zhang, ROENet: A ResNet-based output ensemble for malaria parasite classification, *Electron.* 11 (13) (2022) 2040.
- [18] N.A. Hroub, A.N. Alsannaa, M. Alowafeer, M. Alfarraj, E. Okafor, Explainable deep learning diagnostic system for prediction of lung disease from medical images, *Comput. Biol. Med.* 170 (2024) 108012.
- [19] R. Raza, F. Zulfikar, M.O. Khan, M. Arif, A. Alvi, M.A. Iftikhar, T. Alam, Lung-EffNet: Lung cancer classification using EfficientNet from CT-scan images, *Eng. Appl. Artif. Intell.* 126 (2023) 106902.
- [20] I. Shafi, S. Din, A. Khan, I.D.L.T. Díez, R.d.J.P. Casanova, K.T. Pifarre, I. Ashraf, An effective method for lung cancer diagnosis from ct scan using deep learning-based support vector network, *Cancers* 14 (21) (2022) 5457.
- [21] S.P. Primakov, A. Ibrahim, J.E. van Timmeren, G. Wu, S.A. Keek, M. Beuque, R.W. Granzier, E. Lavrova, M. Scrivener, S. Sanduleanu, et al., Automated detection and segmentation of non-small cell lung cancer computed tomography images, *Nat. Commun.* 13 (1) (2022) 3423.
- [22] Z. Ren, S. Wang, Y. Zhang, Weakly supervised machine learning, *CAAI Trans. Intell. Technol.* 8 (3) (2023) 549–580.
- [23] K. Liu, Stbi-yolo: A real-time object detection method for lung nodule recognition, *IEEE Access* 10 (2022) 75385–75394.
- [24] C. Jacob, G.C. Menon, Pathological categorization of lung carcinoma from multimodality images using convolutional neural networks, *Int. J. Imaging Syst. Technol.* 32 (5) (2022) 1681–1695.
- [25] T. Zhou, F. Liu, X. Ye, H. Wang, H. Lu, CCGL-YOLOV5: A cross-modal cross-scale global-local attention YOLOV5 lung tumor detection model, *Comput. Biol. Med.* 165 (2023) 107387.

- [26] M.H. Sabzalian, F. Kharajinezhadian, A. Tajally, R. Reihanisarsari, H.A. Alk-hazaleh, D. Bokov, New bidirectional recurrent neural network optimized by improved Ebola search optimization algorithm for lung cancer diagnosis, *Biomed. Signal Process. Control.* 84 (2023) 104965.
- [27] K. Barbouchi, D. El Hamdi, I. Elouedi, T.B. Aïcha, A.K. Echi, I. Slim, A transformer-based deep neural network for detection and classification of lung cancer via PET/CT images, *Int. J. Imaging Syst. Technol.* (2023).
- [28] Z. Ji, J. Zhao, J. Liu, X. Zeng, H. Zhang, X. Zhang, I. Ganchev, ELCT-YOLO: An efficient one-stage model for automatic lung tumor detection based on CT images, *Math.* 11 (10) (2023) 2344.
- [29] S. Mammeri, M. Amroune, M.-Y. Haouam, I. Bendib, A. Corrêa Silva, Early detection and diagnosis of lung cancer using YOLO v7, and transfer learning, *Multimedia Tools Appl.* (2023) 1–16.
- [30] S. Dodia, B. Annappa, P.A. Mahesh, Recent advancements in deep learning based lung cancer detection: A systematic review, *Eng. Appl. Artif. Intell.* 116 (2022) 105490.
- [31] M.F. Serj, B. Lavi, G. Hoff, D.P. Valls, A deep convolutional neural network for lung cancer diagnostic, 2018, arXiv preprint [arXiv:1804.08170](https://arxiv.org/abs/1804.08170).
- [32] J. Gong, J. Liu, W. Hao, S. Nie, B. Zheng, S. Wang, W. Peng, A deep residual learning network for predicting lung adenocarcinoma manifesting as ground-glass nodule on CT images, *Eur. Radiol.* 30 (2020) 1847–1855.
- [33] J. Mei, M.-M. Cheng, G. Xu, L.-R. Wan, H. Zhang, SANet: A slice-aware network for pulmonary nodule detection, *IEEE Trans. Pattern Anal. Mach. Intell.* 44 (8) (2021) 4374–4387.
- [34] R. Xu, Z. Liu, Y. Luo, H. Hu, L. Shen, B. Du, K. Kuang, J. Yang, SGDA: Towards 3D universal pulmonary nodule detection via slice grouped domain attention, *IEEE/ACM Trans. Comput. Biol. Bioinform.* (2023).
- [35] A. SU, F.R. PP, A. Abraham, D. Stephen, Deep learning-based BoVW–CRNN model for lung tumor detection in nano-segmented CT images, *Electron.* 12 (1) (2022) 14.
- [36] Z. Mousavi, N. Shahini, S. Sheykhiand, S. Mojtahedi, A. Arshadi, COVID-19 detection using chest X-ray images based on a developed deep neural network, *SLAS Technol.* 27 (1) (2022) 63–75.
- [37] S. Mei, H. Jiang, L. Ma, YOLO-lung: a practical detector based on improved YOLOv4 for pulmonary nodule detection, in: 2021 14th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, CISP-BMEI, IEEE, 2021, pp. 1–6.
- [38] J. Cai, P.W. Read, K. Sheng, The effect of respiratory motion variability and tumor size on the accuracy of average intensity projection from four-dimensional computed tomography: an investigation based on dynamic MRI, *Med. Phys.* 35 (11) (2008) 4974–4981.
- [39] J.L. Causey, K. Li, X. Chen, W. Dong, K. Walker, J.A. Qualls, J. Stubblefield, J.H. Moore, Y. Guan, X. Huang, Spatial pyramid pooling with 3D convolution improves lung cancer detection, *IEEE/ACM Trans. Comput. Biol. Bioinform.* 19 (2) (2020) 1165–1172.
- [40] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2117–2125.
- [41] N. Guo, Z. Bai, Multi-scale pulmonary nodule detection by fusion of cascade R-CNN and FPN, in: 2021 International Conference on Computer Communication and Artificial Intelligence, CCAI, IEEE, 2021, pp. 15–19.
- [42] Z. Wang, J. Zhu, S. Fu, S. Mao, Y. Ye, RFPNet: Reorganizing feature pyramid networks for medical image segmentation, *Comput. Biol. Med.* (2023) 107108.
- [43] A.A.A. Setio, A. Traverso, T. De Bel, M.S. Berens, C. Van Den Bogaard, P. Cerello, H. Chen, Q. Dou, M.E. Fantacci, B. Geurts, et al., Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge, *Med. Image Anal.* 42 (2017) 1–13.
- [44] P. Li, S. Wang, T. Li, J. Lu, Y. HuangFu, D. Wang, A large-scale CT and PET/CT dataset for lung cancer diagnosis [dataset], *Cancer Imaging Arch.* (2020).
- [45] N. Khosravan, U. Bagci, S4ND: Single-shot single-scale lung nodule detection, in: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16–20, 2018, Proceedings, Part II 11*, Springer, 2018, pp. 794–802.
- [46] H. Tang, C. Zhang, X. Xie, Nodulenet: Decoupled false positive reduction for pulmonary nodule detection and segmentation, in: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22*, Springer, 2019, pp. 266–274.
- [47] L. Zhu, H. Zhu, S. Yang, P. Wang, H. Huang, Pulmonary nodule detection based on hierarchical-split HRNet and feature pyramid network with atrous convolution, *Biomed. Signal Process. Control.* 85 (2023) 105024.
- [48] J. Wang, J. Wang, Y. Wen, H. Lu, T. Niu, J. Pan, D. Qian, Pulmonary nodule detection in volumetric chest CT scans using CNNs-based nodule-size-adaptive detection and classification, *IEEE Access* 7 (2019) 46033–46044.
- [49] H. Mkindu, L. Wu, Y. Zhao, Lung nodule detection in chest CT images based on vision transformer network with Bayesian optimization, *Biomed. Signal Process. Control.* 85 (2023) 104866.
- [50] X. Luo, T. Song, G. Wang, J. Chen, Y. Chen, K. Li, D.N. Metaxas, S. Zhang, SCPM-Net: An anchor-free 3D lung nodule detection network using sphere representation and center points matching, *Med. Image Anal.* 75 (2022) 102287.
- [51] X. Lu, N. Zeng, X. Wang, J. Huang, Y. Hu, J. Fang, J. Liu, FFNET: An end-to-end framework based on feature pyramid network and filter network for pulmonary nodule detection, in: 2023 IEEE 20th International Symposium on Biomedical Imaging, ISBI, IEEE, 2023, pp. 1–6.
- [52] Z. Zhou, F. Gou, Y. Tan, J. Wu, A cascaded multi-stage framework for automatic detection and segmentation of pulmonary nodules in developing countries, *IEEE J. Biomed. Heal. Inform.* 26 (11) (2022) 5619–5630.