



# An attention-based neural network model for automatic partition of abdominal lymph nodes in CT imaging

Jingling Wang<sup>1^</sup>, Hao Huang<sup>2^</sup>, Han Wang<sup>1^</sup>, Mingtian Wei<sup>2^</sup>, Zhang Yi<sup>1^</sup>, Ziqiang Wang<sup>2^</sup>, Haixian Zhang<sup>1^</sup>

<sup>1</sup>The Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu, China; <sup>2</sup>The Gastrointestinal Surgery Center, West China Hospital, Sichuan University, Chengdu, China

*Contributions:* (I) Conception and design: J Wang, H Zhang; (II) Administrative support: Z Wang, Z Yi; (III) Provision of study materials or patients: H Huang, M Wei; (IV) Collection and assembly of data: H Wang, H Huang; (V) Data analysis and interpretation: J Wang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

*Correspondence to:* Ziqiang Wang, MD, PhD. The Gastrointestinal Surgery Center, West China Hospital, Sichuan University, 37, Guoxue Lane, Chengdu 610041, China. Email: wangziqiang@scu.edu.cn; Haixian Zhang, PhD. The Machine Intelligence Laboratory, College of Computer Science, Sichuan University, 24 South Section 1, Yihuan Road, Chengdu 610065, China. Email: zhanghaixian@scu.edu.cn.

**Background:** Abdominal lymph node partition is highly relevant to colorectal cancer (CRC) metastasis, which may further affect patient prognosis and survival quality. In the traditional diagnostic process, medical radiologists must partition all lymph nodes from the computed tomography (CT) images for further diagnostics. The manual interpretation of abdominal nodes is experience-dependent and time-consuming, especially for node partition. Therefore, automated partition methods are desirable to make the diagnostic process more accessible. Automatic abdominal lymph node partition is a challenging task due to the subtle morphological features of the nodes and the complex relative position information of the abdominal structure.

**Methods:** In this paper, a node-oriented dataset containing 6,880 nodes with partition labels was constructed by seasoned professionals through 2-round annotation due to there being no dataset with node-oriented labels to perform the partition task. In addition, specific masking strategies and attention mechanisms were proposed for the primary deep neural networks (DNNs). The specific masking strategy could utilize the positional and morphological information more substantially, which intensively exploits prior knowledge and hones the relative positional information in the lower abdomen. The comprehensive attention mechanism could introduce direction-aware information to enhance the inter-channel relationship of features and capture rich contextual relationships with multi-scale kernels.

**Results:** The experiments were based on the node-oriented dataset. The proposed method achieved superior performance [accuracy (ACC): 89.74%; F1 score (F1): 85.95%; area under the curve (AUC): 88.23%], which is significantly higher than the baseline model with several masking strategies (ACC: 62.05–86.16%; F1: 51.77–80.86%; AUC: 60.44–83.94%). For exploration of attention, the proposed method also outperformed the state-of-the-art convolutional block attention module (CBAM; ACC: 88.90%; F1: 84.09%; AUC: 86.86%) with the same proposed input form.

**Conclusions:** Experimental results indicate that the innovative method performs better in experimental metrics than other prevalent methods. The proposed method is expected to be introduced in future medical scenarios, which will help doctors to optimize the diagnosis workflow and improve partition sensitivity.

<sup>^</sup> ORCID: Jingling Wang, 0000-0003-2348-2009; Hao Huang, 0000-0002-2072-9531; Han Wang, 0000-0001-9282-2105; Mingtian Wei, 0000-0002-9602-3201; Zhang Yi, 0000-0002-5867-9322; Ziqiang Wang, 0000-0002-2874-1535; Haixian Zhang, 0000-0002-9821-508X.

**Keywords:** Abdominal lymph node partition; colorectal cancer (CRC) metastasis; deep neural network (DNN); attention mechanism; masking strategy

Submitted Dec 20, 2022. Accepted for publication Jun 13, 2023. Published online Jul 28, 2023.

doi: 10.21037/qims-22-1412

View this article at: <https://dx.doi.org/10.21037/qims-22-1412>

## Introduction

Colorectal cancer (CRC) is the third most frequent malignant tumor worldwide, with no restriction on gender and race (1). Abdominal lymph node metastasis is the leading avenue of multiple diffusion and metastasis of CRC (2). Metastatic CRC exacerbates when tumor cells spread to other body apparatus or tissues and proliferate (3,4). However, CRC is one of the few cancers that can be entirely prevented by removing the correlative lymph nodes or other precancerous growths (5). Finding lymph nodes and determining the partition are effective means to reduce the risk of CRC (6). Therefore, it is necessary to accurately locate the partition of the abdominal lymph nodes before operating (7). Colorectal computed tomography (CT) is a popular lymph node checker tool that scans the abdominal region directly to help radiologists detect the nodes pre-surgically (8). Although CT scans can provide a suitable method for detecting the abdominal nodes, it is time-consuming and inefficient to determine the partitions of all abdominal lymph nodes artificially. The accuracy of diagnosis is related to the skill level, clinical experience, and even the psychological state of the individual radiologist (9). The development of an automatic abdominal lymph node partition method could bring immense benefits to clinical diagnosis. Therefore, a kind of automated technology is urgently needed for the practical application to assist or even replace manual judgment.

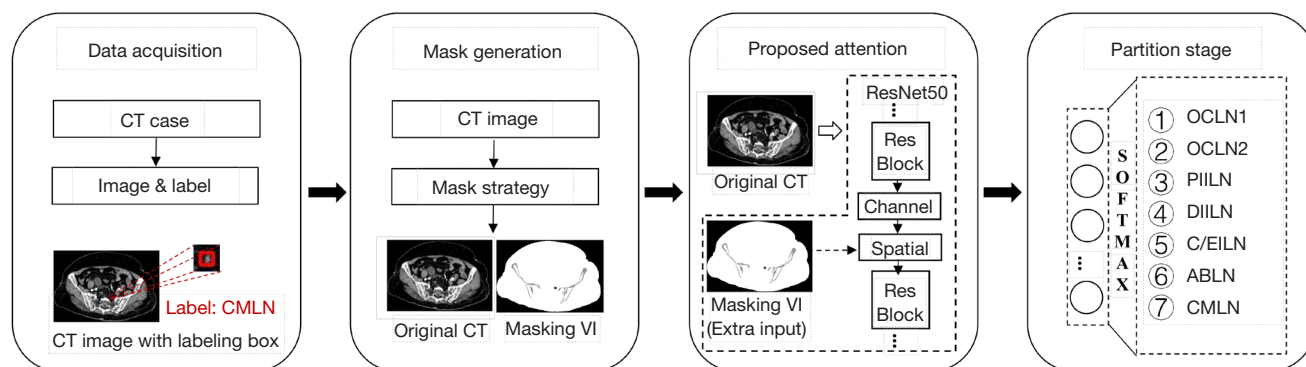
With the development of architecture, deep neural networks (DNNs) have made creative breakthroughs in computer vision (CV) (10). Since the residual network (ResNet) (11) excelled in the image classification tasks (called IISVRC-2015, <http://www.image-net.org/challenges/LSVRC/2015>), DNNs have been widely applied in medical imaging analysis (12-14). Many studies have been conducted in the academic community around colorectal disease diagnosis. Lai *et al.* (15) employed the ResNet to distinguish tissue separation in histological colorectal images. Bagheri *et al.* (16) improved the quality of colonoscopy image segmentation based on DNNs. Sust *et al.* (17) utilized the

patient's physical indicators and the colonoscopy tumor profiles to evaluate the probability of CRC with DNNs. Lewis *et al.* (18) designed Polyp Segmentation Network (PSNet) with a dual encoder and decoder architecture to detect and segment colorectal polyps. Kang *et al.* (19) processed detailed information with a specific attention mechanism, by which we were greatly inspired. In our previous research (20), we achieved accurate abdominal node detection based on CT scanning. It is widely accepted that early detection and partition of the abdominal lymph nodes can prevent CRC. However, only a few studies have addressed the automatic partition of abdominal lymph nodes in CT images. Although Huang *et al.* (21) have conducted similar research, they only explored a simple binary classification problem. Their research method merely used basic mask information and did not incorporate prior knowledge of abdominal organs. Therefore, automated partition methods based on DNNs are still desirable to make the diagnostic process more accessible.

In this study, a node-oriented dataset with partition labels was constructed to address the first problem. It is a large-scale dataset in the field of colorectal diseases, containing nearly 7,000 abdominal lymph nodes. Furthermore, a 2-round annotation calibration approach was adapted to ensure annotating accuracy. A particular masking strategy was devised to solve the second issue, which can enhance the semantic information surrounding the node. The proposed masking strategy integrates the extracted features from 1 examination in a data-driven manner. The comprehensive attention mechanism was designed to address the third challenge, which can further boost the performance of the backbone network. The proposed attention mechanism introduces direction-aware information and captures long-range dependence among spatial positions.

In general, the main contributions of this manuscript are summarized as follows:

- (I) A node-oriented dataset based on abdominal lymph nodes is constructed to implement the node partition task, containing nearly 7,000 nodes. Each node is annotated with accurate partition



**Figure 1** The process flowchart of this study, including 4 stages: data acquisition, mask generation, proposed attention, and partition stage. The experimental dataset was obtained from West China Hospital, Sichuan University, China, which annotates with accurate partition labels. The masking strategy makes full use of prior knowledge contained in the original CT images. The proposed attention mechanism is more targeted, which can capture the inter-channel relationship and build long-range spatial dependencies. The final results are divided into 7 partitions according to the clinical content. CT, computed tomography; OCLN1, obturator cranial lymph node; OCLN2, obturator caudal lymph node; PIILN, proximal internal iliac lymph node; DIILN, distal internal iliac lymph node; C/EILN, common/external iliac lymph node; ABLN, aortic bifurcation lymph node; CMLN, colorectal mesenteries lymph node.

labels through 2-round annotation by seasoned professionals.

- (II) The specific masking strategy for the tiny nodes in the complex abdominal structures is proposed, which exploits prior knowledge intensively and further hones the relative positional information in the lower abdomen.
- (III) The comprehensive attention mechanism takes full advantage of direction-aware information and multi-scale features, which further captures rich contextual relationships and long-range dependence for superior feature representations.
- (IV) Extensive experiments are carried out, including ablation experiments and performance analysis. The proposed method promotes superior performance compared with the other advanced techniques.

## Methods

This section explains the proposed method in detail. Firstly, a detailed description of the dataset used in our experiments is provided. Secondly, the basic architectures are introduced. Thirdly, the masking strategies are explained. Fourthly, the attention mechanisms are introduced. Finally, the pre-training and training strategies are described. This study was evaluated and approved by the Ethics Committee of the West China Hospital, Sichuan University (Nos. 2021-1678, and 2022-1226) and it conformed to the provisions

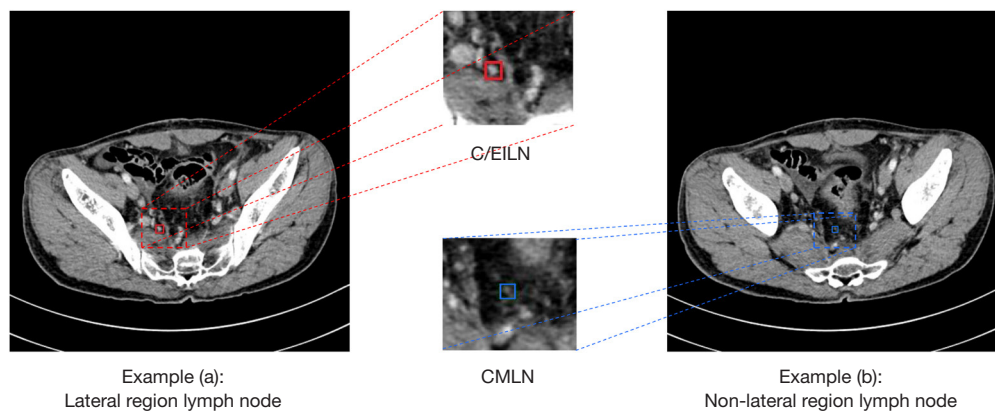
of the Declaration of Helsinki (as revised in 2013). This retrospective study was deemed to carry minimal risk and therefore the requirement for informed patient consent was waived. The general workflow of the whole proposed method is shown in *Figure 1*.

## Data

In this subsection, a detailed description of the dataset used in our experiments is provided. The Japanese Society for Cancer of the Colon and Rectum (JSCCR) (22), as the authority in the field of CRC, declares that all of the abdominal lymph nodes can be divided into 2 regions: lateral region and non-lateral region (shown in *Figure 2*). According to the detailed publication of JSCCR, the abdominal lymph nodes can be further split into 7 partitions: obturator cranial lymph nodes (OCLN1), obturator caudal lymph nodes (OCLN2), proximal internal iliac lymph nodes (PIILN), distal internal iliac lymph nodes (DIILN), common/external iliac lymph nodes (C/EILN), aortic bifurcation lymph nodes (ABLN), and colorectal mesenteries lymph nodes (CMLN), which has more essential clinical consequences compared with 2-region categorizing (*Table 1*). After further data desensitization and processing, the dataset would be conditionally released to interested researchers.

## Collection and division

Only a few open-source datasets are related to abdominal



**Figure 2** Examples of the abdominal lymph node in ultrasound image. The lateral lymph node is identified with red frame, while the non-lateral lymph node is identified with blue frame. C/EILN, common/external iliac lymph node; CMLN, colorectal mesenteries lymph node.

**Table 1** The clinical descriptions with abdominal lymph nodes of different partitions

Region	Partition	Positional description
Lateral region	OCLN1	Crania of the obturator vessels
	OCLN2	Cauda of the obturator vessels
	PIILN	Proximal end of the internal iliac vessels
	DIILN	Distal end of the internal iliac vessels
	C/EILN	Along the common/external iliac vessels
	ABLN	Bifurcation of aorta
Non-lateral region	CMLN	Near the colorectal mesenteries

OCLN1, obturator cranial lymph node; OCLN2, obturator caudal lymph node; PIILN, proximal internal iliac lymph node; DIILN, distal internal iliac lymph node; C/EILN, common/external iliac lymph node; ABLN, aortic bifurcation lymph node; CMLN, colorectal mesenteries lymph node.

lymph nodes, such as the CTLNdataset (23), which still lacks the node-oriented labels and therefore cannot be used to carry out partition tasks. Therefore, a node-oriented dataset was constructed by West China Hospital, Sichuan University, China, which annotates with partition labels. All cases were collected from follow-up patients who agreed to participate. The time span of the cases was from November 2012 to May 2020. A total of 236 CT cases from 236 patients comprising a total of 6,880 lymph nodes were collected and annotated. The distribution of the experimental dataset is shown in *Figure 3*.

During data collection, only 1 case per patient was included in the experimental dataset. As data within the

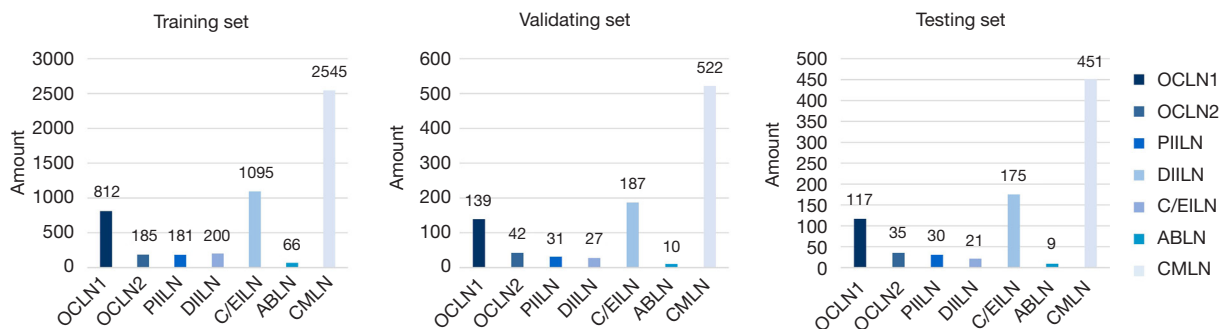
same case are highly correlated, we followed the basic principles of dataset division at the medical case level to divide the dataset into a training set (70%), validation set (15%), and testing set (15%). The division of the experimental dataset is listed in *Table 2*. The training set included 5,084 abdominal lymph nodes from 166 patients of 166 cases. The validating set contained 958 lymph nodes from 35 patients of 35 cases. Similarly, the testing set contained 35 cases from 35 patients with a total of 838 lymph nodes.

### Data annotation

The quality of CT image annotation is highly correlated with the clinical experience and medical knowledge of the annotator. We applied the 2-round annotation method to assure the correctness and quality of abdominal lymph node annotation. Each CT case was annotated by 2 junior annotators independently. The junior annotators marked the locations of lymph nodes and added the corresponding partition label for them. The junior annotators were usually the radiologists or clinical surgeons in gastrointestinal surgery. Then, the middle annotations were merged through non-maximum-suppression (NMS) algorithm. In the case of any conflict, the merged results were reassessed by a senior annotator to obtain the final confirmation. The senior annotators were generally the professional specialist in gastrointestinal surgery. The 2-round annotation process is visualized in *Figure 4*.

### Basic architecture

In this study, some typical diagnostic models were applied.



**Figure 3** Data distribution of abdominal lymph nodes. OCLN1, obturator cranial lymph node; OCLN2, obturator caudal lymph node; PIILN, proximal internal iliac lymph node; DIILN, distal internal iliac lymph node; C/EILN, common/external iliac lymph node; ABLN, aortic bifurcation lymph node; CMLN, colorectal mesenteries lymph node.

**Table 2** Description of the experimental dataset

Dataset	Patient	Case	Node partition						
			OCLN1	OCLN2	PIILN	DIILN	C/EILN	ABLN	CMLN
Training	166	166	812	185	181	200	1,095	66	2,545
Validating	35	35	139	42	31	27	187	10	522
Testing	35	35	117	35	30	21	175	9	451
Total	236	236	1,068	262	242	248	1,457	85	3,518

OCLN1, obturator cranial lymph node; OCLN2, obturator caudal lymph node; PIILN, proximal internal iliac lymph node; DIILN, distal internal iliac lymph node; C/EILN, common/external iliac lymph node; ABLN, aortic bifurcation lymph node; CMLN, colorectal mesenteries lymph node.

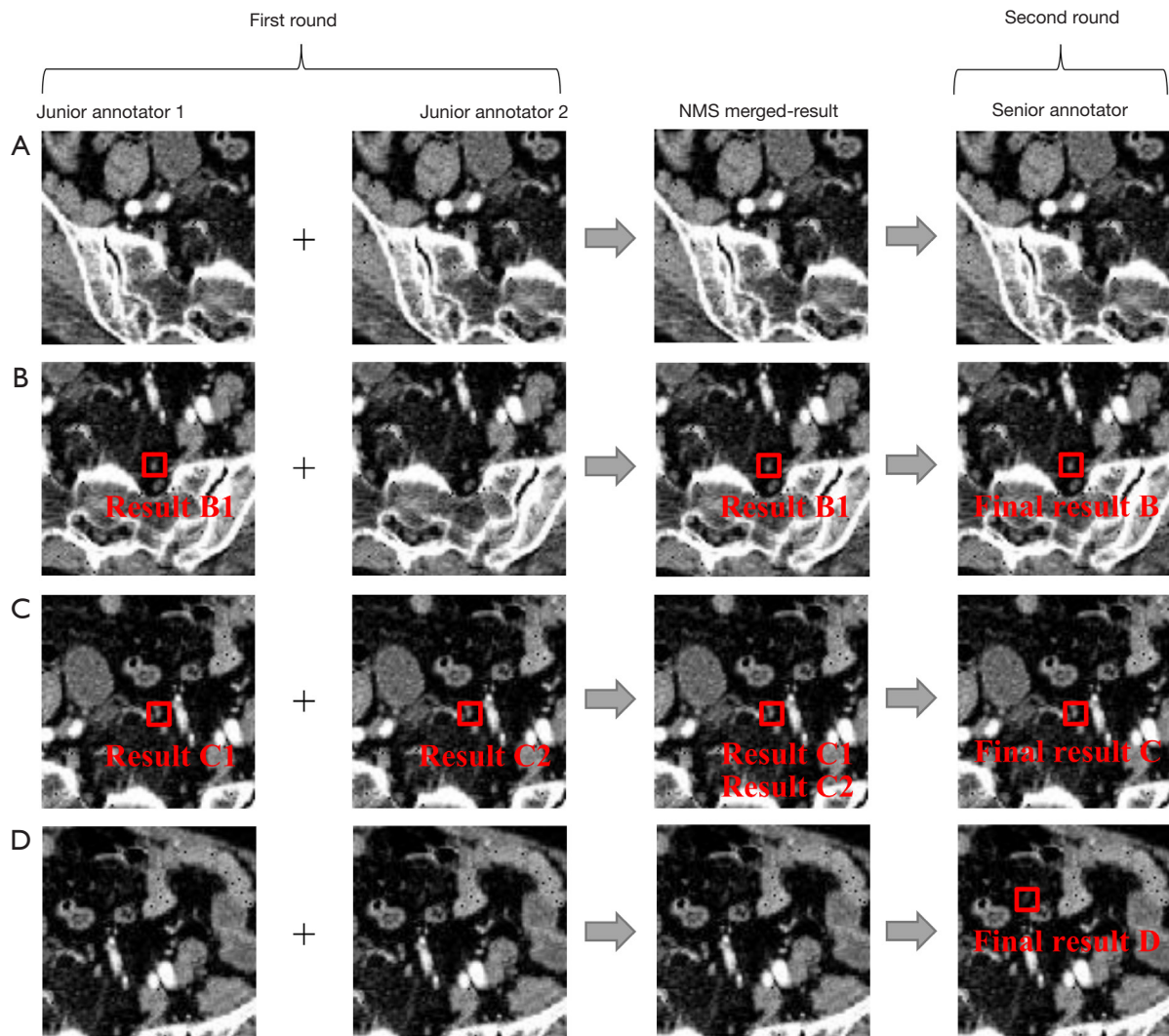
Support vector machine (SVM) is a traditional machine learning technique, which can be brought into medical imaging classification (24). Some high-performance convolutional architectures were also employed, such as AlexNet (25), VGG16 (26), InceptionV4 (27), Faster RCNN (28), YOLOv3 (29), DenseNet (30), ResNet18 (31), ResNet34 (32), and ResNet50 (33). Through experimental results of various network structures (shown in Table 3), the model with best performance is chosen as the basic architecture in subsequent experiments.

**Masking strategy**

Computerized medical imaging has become increasingly powerful in clinical diagnosis and therapy. Mask strategy is an efficient means of medical image analysis, which contributes to understanding contextual information. Zhao *et al.* (34) employed nonlinear masking strategies based on the logarithmic image processing models for medical image enhancement, improving the visual quality of digital medical

images. Zhou *et al.* (35) introduced a patch masking strategy into the pre-training stage and aggregated contextual information to infer content in masked image regions, significantly improving medical image segmentation and classification performance. We had explored some node-oriented masking strategies in our previous research (21). However, further refinements of the node-oriented masking strategy are required to introduce more decision-aided information and achieve more accurate classification results. Some imaging characteristics, such as the abdominal contours and the bone structures, may assist in complex decision making in the clinical setting. Therefore, we designed some novel masking strategies to further enhance the imaging features.

The abdominal lymph nodes of CT images are miniscule and inconspicuous, surrounded by tissues such as blood vessels and muscles. The detailed visualization of masking strategies is illustrated in Figure 5 with the example node marked by a red rectangular box. The labeling box inevitably destroys the semantic information contained

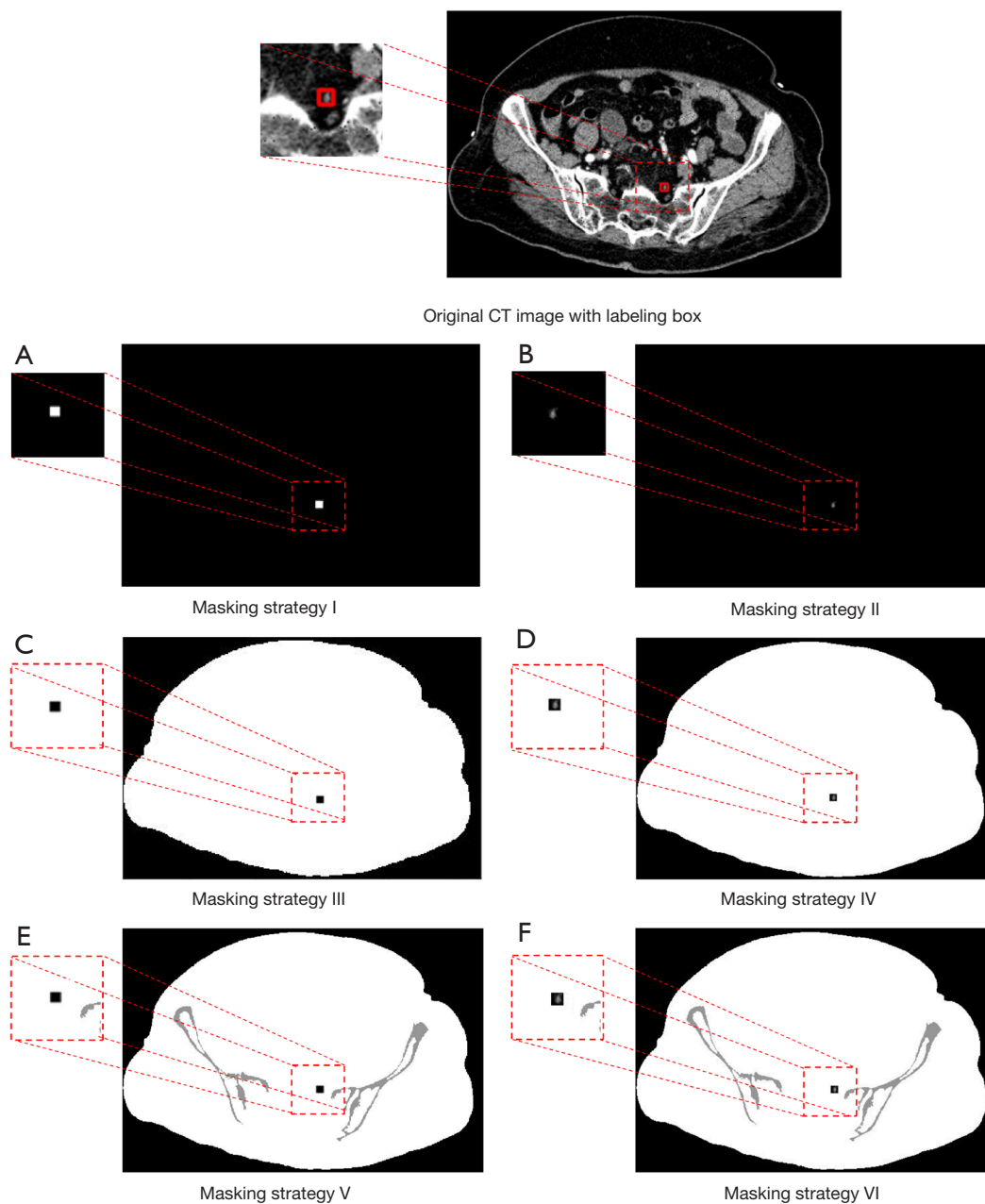


**Figure 4** Visualizations of annotating situations. The red frame denotes that there is an abdominal lymph node contained in this rectangular. The text below the labeling box denotes the annotating results. (A) In this scenario, both junior annotators and the senior annotator annotate that there are no lymph nodes present. (B,C) In these scenarios, the 2 junior annotators hold differing opinions regarding the labeling of lymph nodes, and the final decision relies on the judgment of the senior annotator. (D) In this scenario, both junior annotators overlook the presence of lymph nodes, but the senior annotator identifies and indicates the nodes. NMS, non-maximum-suppression.

**Table 3** Comparison of different basic architectures

Metrics	Basic architecture									
	Traditional	Convolutional								
		SVM	Alex-Net	VGG-16	Inception V4	Faster-RCNN	YOLOv3	DenseNet	ResNet18	ResNet34
ACC (%)	65.27	68.02	69.21	69.45	68.94	68.75	69.03	70.16	70.29	70.41*
F1 (%)	49.71	54.09	54.34	54.94	54.82	55.03	54.11	60.69	60.80	61.45*
AUC (%)	56.54	60.04	60.02	60.60	59.78	60.12	60.43	62.62	63.81	64.94*

ACC, F1, and AUC were calculated on the testing subset. \*, superior indicators than other configurations. SVM, support vector machine; ACC, accuracy; AUC, area under the curve.



**Figure 5** Detailed visualization of the masking strategies. (A-F) represent the diagrammatic drawings processed by Masking strategy I–VI. The example node is marked by red labeling box. The priori information becomes more and more dense from top to bottom. In the right column, the morphological features surrounding the node are introduced compared to the left. CT, computed tomography.

in the original CT scanning images. In order to better utilize the internal feature, an approach to retain the semantic information is desirable. Image masks can retain information in specific areas and block out the interference of irrelevant areas. The mask is a binary matrix mapping from a picture, where pixel value 1 represents the region of

interest (ROI) and pixel value 0 represents the region to be ignored. Image mask has 2 obvious advantages: visualizing node coordinates from the corresponding CT image and reserving useful features based on spatial prior. We proposed 6 masking strategies to address the node partition task, each representing different semantic information.

- ❖ Masking strategy I: Masking strategy I was designed to introduce the approximate positional information of abdominal lymph nodes. The original CT image with a labeling box is processed by binarization treatment. Within the labeling box, the pixel values are set to 1. The pixel values of the other area in this CT image are set to 0.
- ❖ Masking strategy II: On the basis of the introduction of positional information, masking strategy II was designed to bring morphological information into the computation. The pixel values outside the marked area are set to 0. Moreover, the pixel values in the labeling box retain the values of the original CT image based on masking strategy I.
- ❖ Masking strategy III: The abdominal outline information is vital to the partition task. Masking strategy III was designed to better introduce the abdominal contours. Within the abdominal contour region, the pixel values are set to 1, whereas the pixel values in the labeling box and other areas are set to 0.
- ❖ Masking strategy IV: Based on the introduction of abdominal contours, masking strategy IV, aimed to retain the morphological characteristics in the original abdominal CT image. In other words, the pixels in the labeling box retain the values of the original CT image. The rest of the pixel values in the abdominal contour region are set to 1. Moreover, the pixel values outside the abdominal contour region are set to 0.
- ❖ Masking strategy V: On the basis of masking strategy III, masking strategy V further introduces the imaging characteristics of bone. Due to the immutability of bones, the relative positional information is further heightened.
- ❖ Masking strategy VI: Based on masking strategy IV, masking strategy VI is the most complete one, which has the most abundance semantic information. The important characteristics are fully applied, including: the morphological features, the abdominal contours, and the relative positional information between nodes and bones.

In general, the masking strategies I, III, and V only retain the approximate location information and ignore the morphological features of abdominal lymph nodes. The masking strategies II, IV, and VI reserve the morphological features of lymph nodes. The masking strategies I and II only focus on the nodes, whereas the vision is limited. The masking strategies III and IV take the abdominal contours

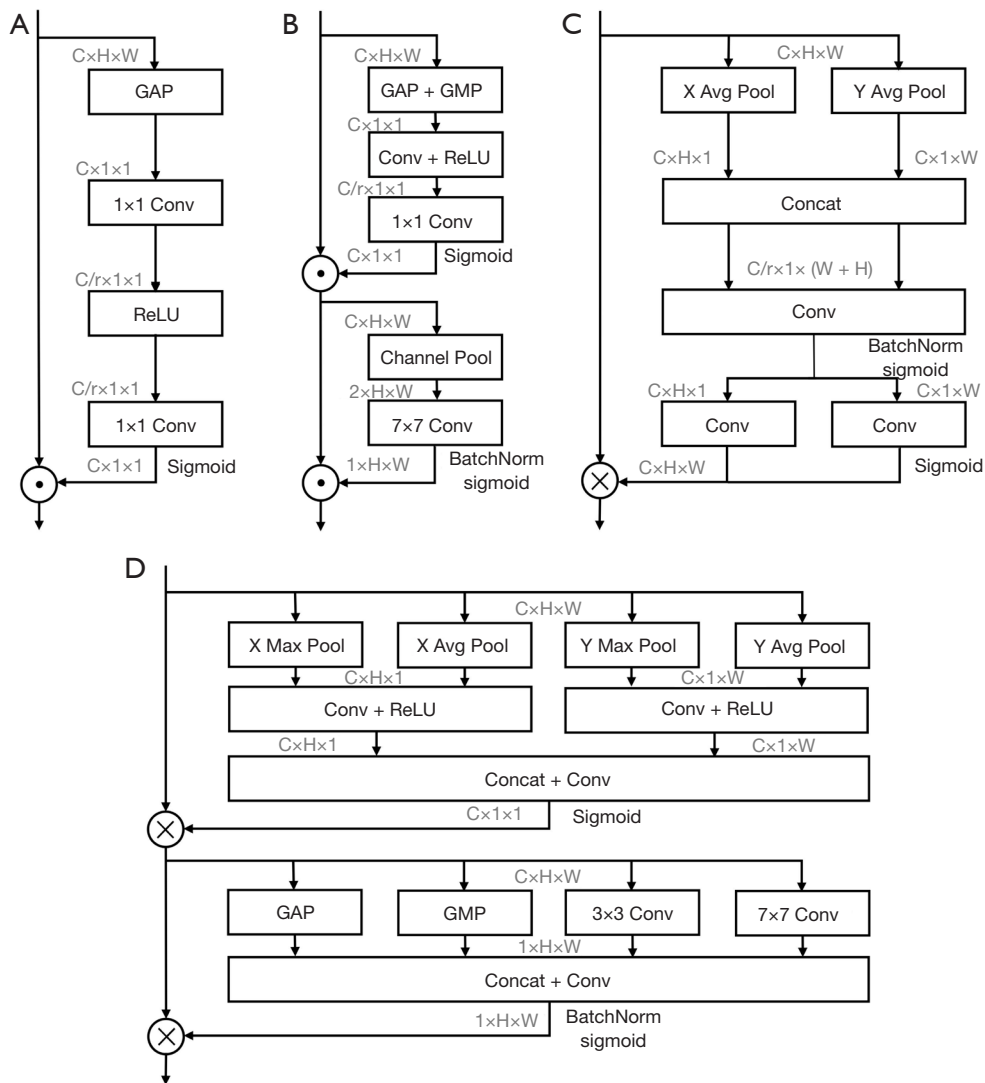
into consideration. The masking strategies V and VI further introduce the imaging features of abdominal bone, improving the relative positional information in the lower abdomen.

### *Comprehensive attention mechanism*

Currently, attention is arguably one of the most productive mechanisms in the deep learning field (36). Specifically, using attention is a useful way to achieve robust performance when there are many features in a network. According to the application of attention mechanisms in different domains, attention mechanisms can be split into 3 categories: spatial attention, channel attention, and mixed attention. There are 2 typical ways for spatial attention: self-attention mechanism (37) and non-local attention mechanism (38). For channel attention mechanism, there are 3 valid modules which can be embedded directly into the network: squeeze excitation (SE) (39), selective kernel (SK) (40), and coordinate attention (CA) (41). For mixed attention mechanism, dual-attention network (DA-net) (42) and convolutional block attention module (CBAM) (43) are 2 typical representations. Some typical attention mechanisms and our proposed attention mechanism are shown in *Figure 6*.

Due to the masking strategy and stacking operation, a masking layer was added based on the original input form. Besides morphological and positional features, channel features between layers should also be considered. Inspired by SE (39) (shown in *Figure 6A*) and CBAM (43) (shown in *Figure 6B*), a novel attention mechanism (shown in *Figure 6D*) directed at lymph node partition was proposed. The proposed comprehensive attention mechanism can be regarded as a targeted means which is conducive to strengthening the expressiveness of the learned features for basic architecture (ResNet50). The proposed attention mechanism captures both inter-channel relationships and long-range spatial dependencies in 2 stages: channel attention stage and spatial attention stage. The arrangement of the whole attention mechanism is shown in *Figure 7*.  $F \in \mathbb{R}^{C \times H \times W}$  represents the intermediate feature map, which is also regarded as input.  $C$  represents the channel number for the image.  $H$  and  $W$  represent the same image's height and weight, respectively. Subsequently,  $C$ ,  $H$ , and  $W$  refer to the components in the  $C$ ,  $H$ , and  $W$  dimensions, respectively.  $F_1$  and  $F_2$  represent the intermediate output after the channel attention stage and spatial attention stage, respectively. The general formulae of the proposed comprehensive attention





**Figure 6** Schematic comparisons with the other advanced attention mechanisms: (A) SE (39) channel attention block; (B) CBAM (43); (C) CA (41) block; (D) comprehensive attention (ours) mechanism. Here, the feature maps refer to feature dimensions,  $C \times H \times W$  represents the dimension of channel (C), height (H) and width (W).  $r$  represents the reduction ratio, which is fixed to 16. Global average pooling denotes as GAP and global max pooling denotes as GMP. X Avg Pool and X Max Pool represent 1D horizontal average pooling and 1D horizontal max pooling. Y Avg Pool and Y Max Pool represent 1D vertical average pooling and 1D vertical max pooling. SE, squeeze excitation; CBAM, convolutional block attention module; CA, coordinate attention; 1D, 1-dimensional.

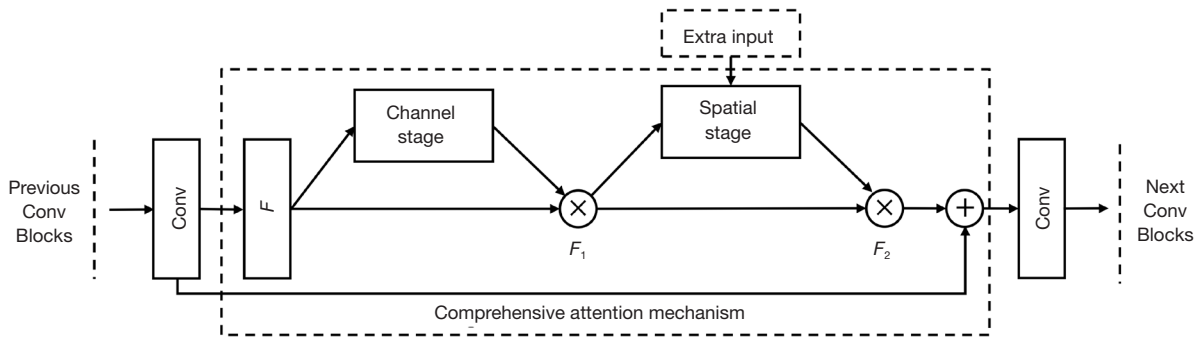
mechanism are as follows:

$$\begin{aligned} F_1 &= M_{channel}(F) \otimes F \\ F_2 &= M_{spatial}(F_1) \otimes F_1 \end{aligned} \quad [1]$$

where  $M_{channel}$  and  $M_{spatial}$  represent the channel and spatial attention operation, respectively.  $\otimes$  denotes element-wise multiplication.

### Channel attention stage

In the channel stages of previous studies (39,40,43), the average and max pooling operations are typical to aggregate deeper information. To further improve the performance of channel attention, some researchers (41) have introduced simplex direction-aware information (shown in Figure 6C). Inspired by these above channel attention mechanisms, we proposed a novel channel attention stage to capture



**Figure 7** The arrangement of our proposed attention method, which is integrated with a residual block in Resnet50. Considering the size of mask, the extra input part is only employed on the first residual block. The extra input part should be deleted in the subsequent residual blocks.

direction-aware information and cross-channel feature for superior feature representations with intra-class compactness. The channel attention map is generated by exploiting the inter-channel relationship of features. The attention mechanism focuses on the most meaningful part of an input image, yet every channel of a feature map can be acknowledged as a feature detector. The input feature map’s spatial dimension is squeezed, aiming to compute the channel attention more efficiently. Both cross-channel feature and direction-aware information are captured simultaneously, which identifies more essential clues of specific object features to infer superior channel-wise attention mechanism.

Firstly, the 2 kinds of directional pooling operations are introduced. Hence, the proposed channel attention can combine features along horizontal and vertical directions, generating a pair of direction-aware channel feature maps. In order to gather important clues, both max pooling operation and average pooling operation along 2 directions are used simultaneously. Secondly, 2 kinds of convolutional operations are applied in their respective directions. The convolution transformations of both directions can fully use the captured channel-sensitive information so that the regions of interest can be accurately stressed. Finally, the intermediate feature maps are concatenated and convolved by a standard convolutional layer. To summarize, the mathematical expressions of the whole channel attention stage can be formulated as below:

$$\begin{aligned}
 M_{channel} &= \sigma_1(F_{1 \times 1}(X(F); Y(F))) \\
 X(F) &= \sigma_2(F_h[XMP(F), XAP(F)]) \\
 XMP_c^h(h) &= \max\{F_c(h, i) | 0 \leq i \leq W\} \\
 XAP_c^h(h) &= \frac{1}{W} \sum_{i=0}^W F_c(h, i) \\
 Y(F) &= \sigma_2(F_w[YMP(F), YAP(F)]) \\
 YMP_c^w(w) &= \max\{F_c(j, w) | 0 \leq j \leq H\} \\
 YAP_c^w(w) &= \frac{1}{H} \sum_{j=0}^H F_c(j, w)
 \end{aligned}
 \tag{2}$$

where  $\sigma_1$  denotes the sigmoid function and  $\sigma_2$  denotes the rectified linear unit (ReLU) function.  $X(F)$  and  $Y(F)$  are intermediate results with the input  $F$  during channel attention stage, which are feature maps along horizontal and vertical directions respectively. Then,  $X(F)$  and  $Y(F)$  are jointed and convolved by a standard convolutional operation.  $F_{1 \times 1}$  represents a weight-shared  $1 \times 1$  convolutional transformation function.  $XMP(F)$  refers to the max-pooled feature map along the horizontal direction and  $XAP(F)$  refers to the average-pooled feature map along the horizontal direction. Conversely,  $YMP(F)$  refers to the max-pooled feature map along the vertical direction and  $YAP(F)$  refers to the average-pooled feature map along the vertical direction.  $h$  and  $w$  are components of  $F$ .  $F_h$  and  $F_w$  are  $1 \times 1$  convolutional operations along different directions.  $XMP_c^h(h)$  denotes the output of the  $c$ -th channel at height  $h$  with 1-dimensional (1D) horizontal global max

pooling operation.  $XAP_c^h(h)$  represents the output of the  $c$ -th channel at height  $h$  with 1D horizontal global average pooling operation.  $YMP_c^w(w)$  denotes the output of the  $c$ -th channel at width  $w$  with 1D vertical global max pooling operation.  $YAP_c^w(w)$  represents the output of the  $c$ -th channel at width  $w$  with 1D vertical global average pooling operation.  $F_c$  represents the linear transformation which can be further learned to capture the significance of each channel.

The proposed channel attention combines features along 2 spatial directions, generating a pair of direction-aware channel feature maps. The experiments indicated that using both average-pooling and max-pooling features can greatly improve networks' representation ability and extraction power rather than only using each independently, which shows the effectiveness of our novel design.

**Spatial attention stage**

A more-comprehensive spatial attention mechanism was proposed to better utilize the spatial features. The spatial attention map can be generated by exploiting the inter-spatial relationship of features. Unlike the channel attention mechanism mentioned above, our spatial attention focuses on the informative spatial part, which can be seen as a supplement to the channel attention mechanism. In order to compute the spatial attention map, the following several kinds of operations were applied to all residual blocks: global average pooling, global max pooling, convolutional operations with 2 kernel sizes, and extra input superposition. Using both global average pooling and global max pooling can help the spatial stage capture more global information. Convolutional operations can help the model further learn the deep expression of features, and convolutional kernels with different sizes focus on different feature types. From this, 2 kernel sizes with different sizes are applied, which can assist the model in capturing more accessible information. In addition, the masking strategy VI is used again as an extra input form in the first block, further enhancing the semantic features during the spatial attention stage. The mathematical expressions of the whole spatial attention stage can be formulated as below:

$$M_{spatial}(F_1) = \sigma_1(F_{1 \times 1}(GAP(F_1); GMP(F_1); F_{3 \times 3}(F_1); F_{7 \times 7}(F_1); EI(F_1))) \quad [3]$$

where  $\sigma_1$  represents the sigmoid function,  $F_{1 \times 1}$  represents a  $1 \times 1$  convolutional operation. All subsequent operations are concatenated and convolved by the standard convolution

$F_{1 \times 1}$ .  $GAP(F_1)$  and  $GMP(F_1)$  denote global average pooling and global max pooling operations with the intermediate input  $F_1$ , separately.  $F_{3 \times 3}(F_1)$  and  $F_{7 \times 7}(F_1)$  represent the convolutional operations with the filter sizes of  $3 \times 3$  and  $7 \times 7$  with the input  $F_1$  after channel processing.  $EI(F_1)$  represents the extra input processed by the means of masking strategy VI, which is only applied to the first residual block. It should be noted that the  $EI(F_1)$  item is not accepted for calculating in the subsequent residual blocks.

**Training stage**

Being different from starting with a series of random initialization parameters, pre-training strategy was applied at the beginning of model training process in this study, which can accelerate the subsequent iterations. The initial parameters were obtained by pre-training on another dataset of the same type, such as the ImageNet (44). It is the most commonly used pre-training dataset and it contributes to train robust and universal image classification models. The obtained parameters were used to initialize the model  $f(x; y; \theta)$ , which can promote the network better approach the global optimal solution.

In addition, the adjustment of learning rate benefits relate directly to the performance of the model during training stage (45). The gradual warmup strategy proposed by Facebook (46) was applied. The formula of learning rate with gradual warmup can be computed as follows:

$$lr_i = \frac{lr_{initial} \times i}{N_{epochs}} \quad [4]$$

where  $i$  represents the current epoch,  $lr_i$  represents the learning rate of  $i$ ,  $lr_{initial}$  represents the initial learning rate by setting, and  $N_{epochs}$  represents the total number of all epochs in warmup strategy.

After the warmup stage, the learning rate reaches the initialized value. To improve global optimization, a cosine annealing algorithm was adopted in the current study:

$$lr = \frac{1}{2} \left( 1 + \cos \left( \frac{t\pi}{T} \right) \right) lr_{initial} \quad [5]$$

The multi-classification cross-entropy loss function is:

$$Loss = - \sum_{i=1}^N y_{(i,m)} \log(\hat{y}_{(i,m)}) \quad [6]$$

where  $y_{(i,m)}$  is the target output and  $\hat{y}_{(i,m)}$  is the actual output of the network.  $i$  denotes the  $i$ -th sample, which belongs

to the class  $m$ . The value of  $m$  varies with the sample  $i$ .  $N$  represents the total number of all samples.

## Results

The section analyzes the experimental results in detail. Firstly, the description of the experimental configuration is stated. Secondly, the detailed analysis of basic architectures is explained. Thirdly, the experimental results of the proposed masking strategies are explained. Finally, the comparison between our proposed attention mechanism and the other advanced attention mechanisms are stated in detail.

### Configuration

The experiments were developed using a universal Python package, PyTorch (47) (<https://pytorch.org/>), which runs on Linux operating system (<https://www.linux.org/>). All the experiments including the models' training process in this study were carried on a high-powered workstation, including: Intel Xeon E5-2620 CPU @2.4 GHz (USA), 3 NVIDIA Tesla K40 GPUs (USA), and 64 GB of random-access memory (RAM). All CT images in this study were resized as the size of 344×224 and the pixel values were normalized to the range of 0 to 1. The parameters pretrained by ImageNet were loaded to initialize the basic convolutional architectures. Stochastic gradient descent (SGD) with momentum was utilized as the optimizer. The batch size was set to 32 and the number of training epochs is 150. The parameters of the model are saved when models achieve the highest accuracy in the 150 epochs.

### Metrics

Accuracy (ACC), Marco F1 score (F1), and area under the curve (AUC) were calculated to evaluate the performances. ACC denoted the number of correct predictions among all instances, which is a measurement of systematic error. F1 reflected the simultaneous effect of between Precision and Recall through harmonic means by imposing more penalties on extreme values. AUC represented the area under the receiver operating characteristic (ROC) curve. The calculation measurements of ACC, F1, and AUC are illustrated as follows:

$$ACC = \frac{\sum_i^N TP_i + \sum_i^N TN_i}{\sum_i^N TP_i + \sum_i^N TN_i + \sum_i^N FP_i + \sum_i^N FN_i} \quad [7]$$

$$Precision_i = \frac{TP_i}{TP_i + FP_i} \quad [8]$$

$$Recall_i = \frac{TP_i}{TP_i + FN_i} \quad [9]$$

$$F1 = \frac{1}{N} \sum_i^N \frac{2 \times Precision_i \times Recall_i}{Precision_i + Recall_i} \quad [10]$$

$$AUC = \frac{1}{N} \sum_i^N \frac{AUC_i}{N} \quad [11]$$

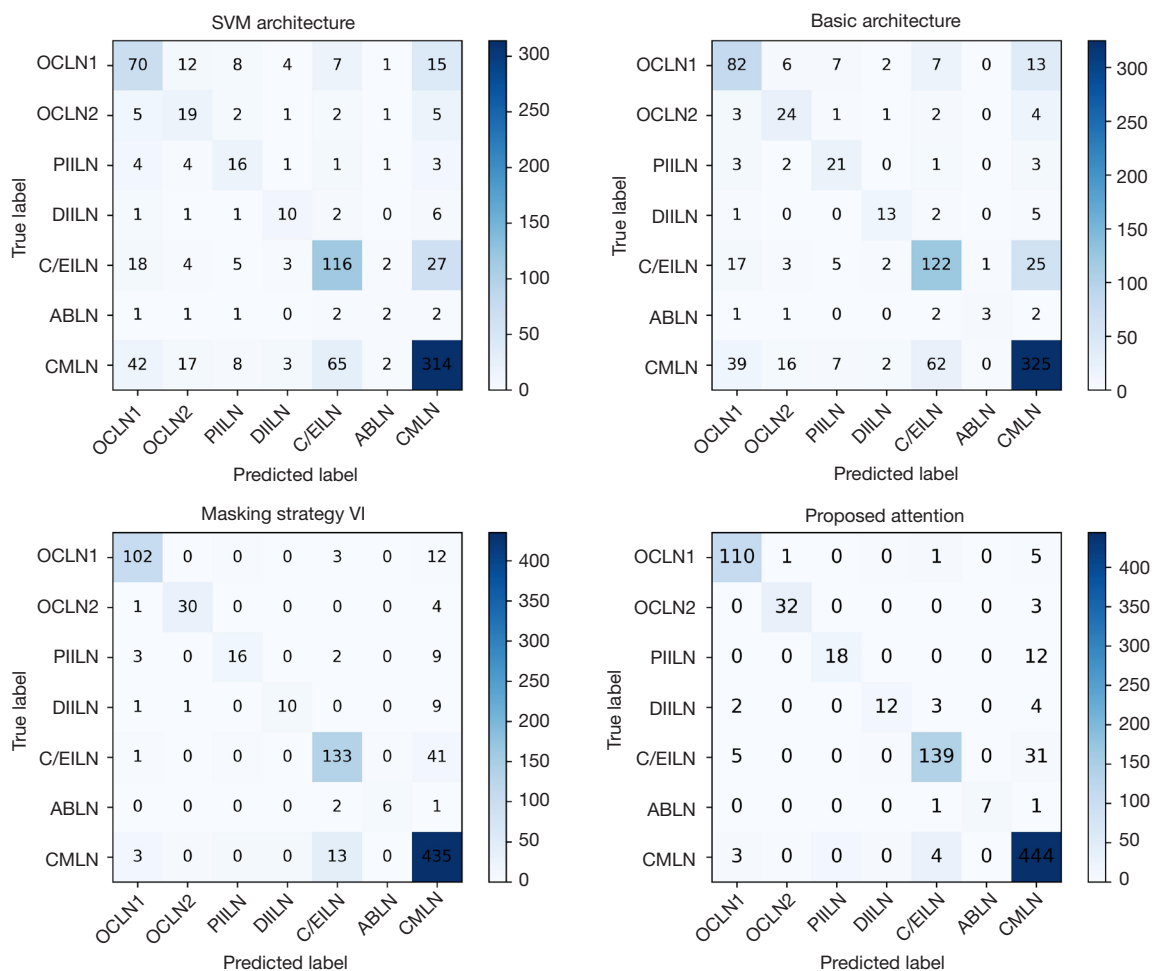
where  $N$  denotes the total number of all categories and  $i$  denotes the current category. In this study,  $N$  equalled 7 and  $i$  ranged from 1 to 7.  $TP_i$  was the number of  $i$ -th true positives.  $TN_i$  was the number of  $i$ -th true negatives.  $FP_i$  was the number of  $i$ -th false positives.  $FN_i$  was the number of  $i$ -th false negatives. The 3 above indicators ( $ACC$ ,  $F1$ ,  $AUC$ ) were computed in subsequent experiments. The higher the 3 indicators, the better the performances.

### Analysis of basic architectures

We set the original CT image with a labeling box to the input form and then applied different architectures to investigate the performance of basic architectures. The experimental results on the testing subset are listed in *Table 3*. The confusion matrix of SVM is shown in *Figure 8A*, and the confusion matrix of ResNet50 is shown in *Figure 8B*. The performance of traditional SVM is quite poor, and it does not cope well with a large amount of data. Some well-performing segmentation structures (including Faster RCNN, YOLOv3, and DenseNet) do not achieve satisfactory results in abdominal lymph node classification tasks, which may increase the likelihood of overfitting on smaller medical image datasets. ResNets are specifically designed for image classification, demonstrating the highest value for the ACC (70.16–70.41%), which is superior to the other convolutional classifiers. ResNets are easier to optimize, and they can gain accuracy from considerably increased depth. Further experimental metrics showed that the deeper the ResNet, the better the architecture performance. As a result, the ResNet50 was chosen as the basic architecture in our subsequent experiments.

### Analysis of masking strategies

From the results shown in *Table 4*, it is evident that using a masking strategy stack with an original CT image



**Figure 8** Confusion matrixes of the methods mentioned in this study. The ordinate represents the true label and the abscissa represents the predicted label. (A) represents the performance of the SVM classifier, with the general input form (the original CT image with a red labeling box). (B) represents the performance of the ResNet50 model, with the general input form. (C) represents the performance of the ResNet50 model, with the specific input form (the original CT image stacking with the mask processed by the proposed masking strategy VI). (D) represents the performance of the ResNet50 model after using the proposed attention mechanism, with the specific input form. CT, computed tomography; OCLN1, obturator cranial lymph node; OCLN2, obturator caudal lymph node; PIILN, proximal internal iliac lymph node; DIILN, distal internal iliac lymph node; C/EILN, common/external iliac lymph node; ABLN, aortic bifurcation lymph node; CMLN, colorectal mesenteries lymph node.

performs better than that only using CT or mask images. The results of the first line without masking strategy can be seen as the baseline. By comparison of the first 7 lines, it can be concluded that the original CT image contains more prodigious amounts of information than the mask. By comparing the baseline and the last six lines, it can be concluded that the labeling box destroys the information contained in the original CT image to some extent. The proposed masking strategy stacking with the original CT image is a reasonable choice. On the one

hand, the information contained in the original CT image is protected and utilized sufficiently. On the other hand, the features of position and shape can be highlighted better. Masking strategy VI stacking with the original CT image performed best, which benefits from introducing the relative positional information and reinforcing the morphological features. The confusion matrix of masking strategy VI is shown in *Figure 8C*. As a result, the proposed masking strategy VI was chosen as the final masking strategy.

**Table 4** Comparison of different input forms and masking strategies

No.	Input form								Metrics		
	CT	Box	Masking strategy						ACC (%)	F1 (%)	AUC (%)
			I	II	III	IV	V	VI			
Baseline	√	√	–	–	–	–	–	–	70.41	61.45	64.94
No. 1	–	–	√	–	–	–	–	–	62.05	51.77	60.44
No. 2	–	–	–	√	–	–	–	–	63.24	51.97	60.54
No. 3	–	–	–	–	√	–	–	–	65.16	52.91	61.50
No. 4	–	–	–	–	–	√	–	–	66.83	53.54	62.55
No. 5	–	–	–	–	–	–	√	–	66.35	53.62	64.55
No. 6	–	–	–	–	–	–	–	√	67.78	53.83	65.09
No. 7	√	–	√	–	–	–	–	–	80.19	70.07	78.03
No. 8	√	–	–	√	–	–	–	–	81.50	71.77	79.43
No. 9	√	–	–	–	√	–	–	–	82.10	72.32	79.94
No. 10	√	–	–	–	–	√	–	–	85.08	78.82	83.04
No. 11	√	–	–	–	–	–	√	–	83.29	74.08	80.62
No. 12	√	–	–	–	–	–	–	√	86.16*	80.86*	83.94*

CT represents the original computed tomography image. Box represents the labeling box. √ in the cell means that the corresponding component is applied, while – represents unused. ACC, F1 and AUC are calculated on testing subset. \*, superior indicators than other configurations. CT, computed tomography; ACC, accuracy; AUC, area under the curve.

### Analysis of attention mechanisms

A series of comparative experiments with the other advanced attention methods mentioned in Subsection *Comprehensive attention mechanism* were further carried out. Through the experimental results shown in *Table 5*, our proposed method achieved the best performance compared with the other state-of-the-art methods in the partition task. Self-attention and non-local belong to early spatial attention research, which performs poorly in the node partition task. As channel attention research continues, several typical architectures achieve some excellent results, and these approaches still have some drawbacks. SE and SK squeeze each 2-dimensional (2D) feature map to build interdependencies among channels, which ignore the importance of positional information. CA introduces the concept of direction but remains channel-level. Hence, mixed attention is entering the mainstream. DA-Net and CBAM can only capture local relations but fail in long-range dependencies modeling. By encoding direction-

aware and position-sensitive feature maps, our proposed comprehensive attention mechanism is designed to capture rich contextual relationships for superior feature representations with intra-class compactness, which contributes to achieving results superior to those of the other advanced methods. The confusion matrix of our proposed attention mechanism is shown in *Figure 8D* and the corresponding visualization is shown in *Figure 9* [the heat maps were obtained by the means of Grad-CAM (48)].

### Discussion

In this study, we proposed a novel masking strategy and developed the comprehensive attention mechanism based on ResNet50 to accurately partition abdominal lymph nodes in CT imaging. The proposed method can improve the efficiency and reliability of abdominal lymph node partition, which has several advantages: (I) the proposed automatic technique with neural network theory can improve the accuracy of abdominal node partition; (II) the

**Table 5** Comparison with the advanced methods

Attention category	Method	Metrics		
		ACC (%)	F1 (%)	AUC (%)
Baseline	ResNet50	86.16	80.86	83.94
Spatial	+self-attention (37)	86.39	80.90	84.42
	+non-local (38)	87.11	80.93	84.79
Channel	+SE (39)	86.51	80.65	84.24
	+SK (40)	86.99	81.00	84.60
	+CA (41)	87.21	81.09	84.88
Mixed	+DA-Net (42)	87.82	82.83	85.35
	+CBAM (43)	88.90	84.09	86.86
	+proposed	89.74*	85.95*	88.23*

Baseline represents without applying any attention components. ACC, F1, and AUC are calculated on testing subset. \*, superior indicators than other configurations. ACC, accuracy; AUC, area under the curve; SE, squeeze excitation; SK, selective kernel; CA, coordinate attention; DA-net, dual-attention network; CBAM, convolutional block attention module.

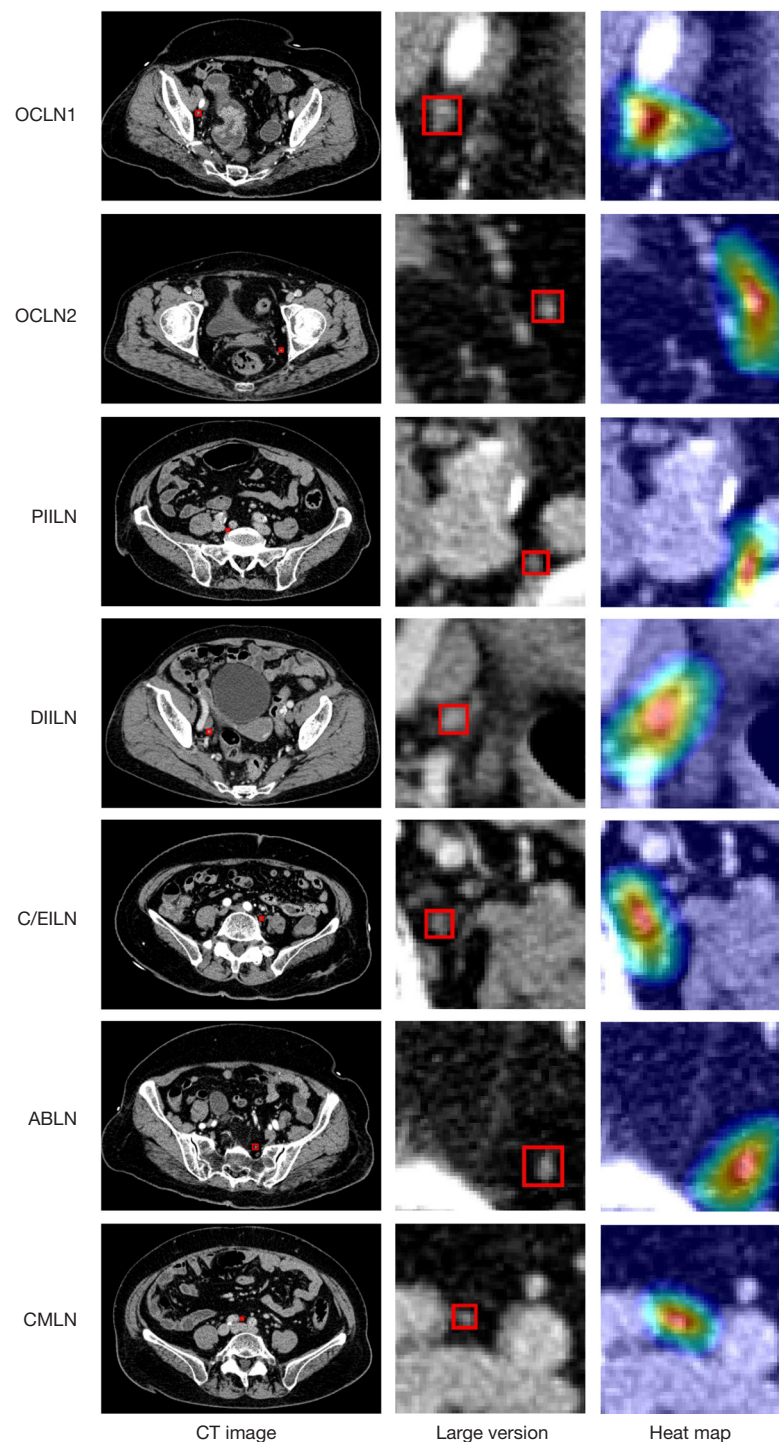
visualizations of experimental results (shown in *Figure 9*) show that the proposed method determines the partition of nodes according to the features with higher correlation, rather than being affected by other non-nodular areas; (III) the accuracy of our proposed attention mechanism is higher than that of the other advanced attention techniques, which is much more suitable for the task of node partitioning.

The actual contents of our ongoing/future research are as follows: (I) as research continues, the dataset size will increase, and the types of labels will be refined; (II) further studies will be conducted on the updated dataset, such as determining whether the node is negative or positive; (III) multi-modality medical information of the same node will be combined and exploited to further improve the experimental metrics.

## Conclusions

In the current study, a large-scale dataset with high-level manual annotation was constructed. A 2-round annotation calibration method was applied to build such a dataset with

node-oriented labels. The proposed method introduces 2 contributions to address the partition task: the innovative masking strategy and the comprehensive attention mechanism. The proposed masking strategy can effectively retain semantic information and enhance relative location information, pushing diagnostic performance further. The comprehensive attention mechanism includes 2 stages: the channel attention stage and the spatial attention stage. The channel attention can subtly capture direction-aware information, which is more beneficial to visual tasks with classification. The spatial attention stage contains the convolutional operations with multiple kernel sizes and the reuse of extra input. The multiple convolutions contribute to capturing multi-scale information. The specific extra input reuses the positional mask, which is conducive to capturing long-range dependence among spatial positions. The attention-based automatic method of abdominal lymph node partition performs better than other state-of-the-art methods, benefitting from the enhancements of the inter-channel relationship and long-range spatial dependencies.



**Figure 9** Visualizations of the abdominal lymph node in 7 partitions. The left column is the original CT images and the tiny nodes are identified with the red labeling boxes. The middle column is the corresponding enlarged parts of the lymph node, which is also remarked by the red rectangles. The right column is the Grad-CAM (48) heat maps stacking with the partial CT image. CT, computed tomography; OCLN1, obturator cranial lymph node; OCLN2, obturator caudal lymph node; PIILN, proximal internal iliac lymph node; DIILN, distal internal iliac lymph node; C/EILN, common/external iliac lymph node; ABLN, aortic bifurcation lymph node; CMLN, colorectal mesenteries lymph node.



## Acknowledgments

*Funding:* This work was supported by the National Key Research and Development Program of China (No. 2018AAA0100201), the Sichuan Province Science and Technology Program (Nos. 21ZDYF1881 & 2019YFS0375), the 1·3·5 Project for Disciplines of Excellence-Clinical Research Incubation Project, West China Hospital, Sichuan University (No. 2019HXFH031), the Wu Jieping Medical Foundation (No. 320.2710.1860), the Bethune Charitable Foundation (No. HZB-20190528-4), and the Major Science and Technology Project from the Science & Technology Department of Sichuan Province (No. 2022ZDZX0023).

## Footnote

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at <https://qims.amegroups.com/article/view/10.21037/qims-22-1412/coif>). The authors have no conflicts of interest to declare.

*Ethical Statement:* The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study was evaluated and approved by the Ethics Committee of the West China Hospital, Sichuan University (Nos. 2021-1678, and 2022-1226) and it conformed to the provisions of the Declaration of Helsinki (as revised in 2013). This retrospective study was deemed to carry minimal risk and therefore the requirement for informed patient consent was waived.

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

## References

1. Vogelstein B, Fearon ER, Kern SE, Hamilton SR, Preisinger AC, Nakamura Y, White R. Allelotyping of colorectal carcinomas. *Science* 1989;244:207-11.
2. Sakuragi M, Togashi K, Konishi F, Koinuma K, Kawamura Y, Okada M, Nagai H. Predictive factors for lymph node metastasis in T1 stage colorectal carcinomas. *Dis Colon Rectum* 2003;46:1626-32.
3. Wu H, Li B, Yang Z, et al. Intravoxel incoherent motion diffusion-weighted imaging for early assessment of combined anti-angiogenic/chemotherapy for colorectal cancer liver metastases. *Quant Imaging Med Surg* 2022;12:4587-600.
4. Qiu L, Hu J, Weng Z, Liu S, Jiang G, Cai X. A prospective study of dual-energy computed tomography for differentiating metastatic and non-metastatic lymph nodes of colorectal cancer. *Quant Imaging Med Surg* 2021;11:3448-59.
5. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin* 2019;69:7-34.
6. Guillem JG, Chessin DB, Cohen AM, et al. Long-term oncologic outcome following preoperative combined modality therapy and total mesorectal excision of locally advanced rectal cancer. *Ann Surg* 2005;241:829-36; discussion 836-8.
7. Flor N, Mezzanica M, Rigamonti P, Rocco EG, Bosari S, Ceretti AP, Soldi S, Peri M, Sardanelli F, Cornalba GP. Contrast-enhanced computed tomography colonography in preoperative distinction between T1-T2 and T3-T4 staging of colon cancer. *Acad Radiol* 2013;20:590-5.
8. Flor N, Ceretti AP, Mezzanica M, Rigamonti P, Peri M, Tresoldi S, Soldi S, Mangiavillano B, Sardanelli F, Cornalba GP. Impact of contrast-enhanced computed tomography colonography on laparoscopic surgical planning of colorectal cancer. *Abdom Imaging* 2013;38:1024-32.
9. Stoitsis J, Valavanis I, Mougiakakou SG, Golemati S, Nikita A, Nikita KS. Computer aided diagnosis based on medical image processing and artificial intelligence methods. *Nucl Instrum Methods Phys Res A* 2006;569:591-5.
10. Miikkulainen R, Liang J, Meyerson E, Rawal A, Fink D, Francon O, Raju B, Shahrzad H, Navruzyan A, Duffy N. Evolving deep neural networks. In: Kozma R, Alippi C, Choe Y, Morabito FC. Editors. *Artificial intelligence in the age of neural networks and brain computing*. Elsevier; 2019:293-312.
11. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA; 2016:770-8.

12. Anthimopoulos M, Christodoulidis S, Ebner L, Christe A, Mouggiakakou S. Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network. *IEEE Trans Med Imaging* 2016;35:1207-16.
13. Wang J, Ju R, Chen Y, Liu G, Yi Z. Automated diagnosis of neonatal encephalopathy on aEEG using deep neural networks. *Neurocomputing* 2020;398:95-107.
14. Wang H, Zhang H, Hu J, Song Y, Bai S, Yi Z. DeepEC: An error correction framework for dose prediction and organ segmentation using deep neural networks. *Int J Intell Syst* 2020;35:1987-2008.
15. Lai KH, Chang MC. Classification of Tissue Types in Histological Colorectal Cancer Images Using Residual Networks. 2019 IEEE 8th Global Conference on Consumer Electronics (GCCE), Osaka, Japan, 2019:299-300.
16. Bagheri M, Mohrekeh M, Tehrani M, Najarian K, Karimi N, Samavi S, Reza Soroushmehr SM. Deep Neural Network based Polyp Segmentation in Colonoscopy Images using a Combination of Color Spaces. *Annu Int Conf IEEE Eng Med Biol Soc* 2019;2019:6742-5.
17. Sust TJ, Leend FC, Kest FX, Wangth SM, Haoth MJ. Estimation of Colorectal Tumor with Colonoscopy Based on Convolutional Neural Network. 021 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS); 2021:1-2.
18. Lewis J, Cha YJ, Kim J. Dual encoder-decoder-based deep polyp segmentation network for colonoscopy images. *Sci Rep* 2023;13:1183.
19. Kang DH, Cha YJ. Efficient attention-based deep encoder and decoder for automatic crack segmentation. *Struct Health Monit* 2022;21:2190-205.
20. Wang H, Huang H, Wang J, Wei M, Yi Z, Wang Z, Zhang H. An intelligent system of pelvic lymph node detection. *Int J Intell Syst* 2021;36:4088-116.
21. Huang Y, Huang S, Wang H, Wei M, Wang J, Zhang H, Wang Z, Yi Z. A prior-based method for colorectal lymph node region classification via deep neural network. 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Houston, TX, USA; 2021:1741-8.
22. Hashiguchi Y, Muro K, Saito Y, Ito Y, Ajioka Y, Hamaguchi T, et al. Japanese Society for Cancer of the Colon and Rectum (JSCCR) guidelines 2019 for the treatment of colorectal cancer. *Int J Clin Oncol* 2020;25:1-42.
23. Roth HR, Lu L, Seff A, Cherry KM, Hoffman J, Wang S, Liu J, Turkbey E, Summers RM. A new 2.5D representation for lymph node detection using random sets of deep convolutional neural network observations. *Med Image Comput Comput Assist Interv* 2014;17:520-7.
24. Camlica Z, Tizhoosh HR, Khalvati F, editors. Medical image classification via SVM using LBP features from saliency-based folded data. 2015 IEEE 14th international conference on machine learning and applications (ICMLA); 2015: IEEE.
25. Krizhevsky A, Sutskever I, Hinton GEJCotA. Imagenet classification with deep convolutional neural networks. *Communications of the ACM* 2017;60:84-90.
26. Simonyan K, Zisserman AJapa. Very deep convolutional networks for large-scale image recognition. 2014.
27. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA, editors. Inception-v4, inception-resnet and the impact of residual connections on learning. Thirty-first AAAI conference on artificial intelligence; 2017.
28. Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans Pattern Anal Mach Intell* 2017;39:1137-49.
29. Redmon J, Farhadi AJapa. Yolov3: An incremental improvement. 2018.
30. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ, editors. Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition; 2017.
31. Yu X, Wang S-HJFI. Abnormality diagnosis in mammograms by transfer learning based on ResNet18. *Fundamenta Informaticae* 2019;168:219-30.
32. Al NMA-MM, Khudeyer RSJI. ResNet-34/DR: A Residual Convolutional Neural Network for the Diagnosis of Diabetic Retinopathy. *Informatica* 2021;45.
33. Ali N, Quansah E, Köhler K, Meyer T, Schmitt M, Popp J, Niendorf A, Bocklitz TJTb. Automatic label-free detection of breast cancer using nonlinear multimodal imaging and the convolutional neural network ResNet50. *Translational Biophotonics* 2019;1:e201900003.
34. Zhao Z, Zhou Y, editors. Comparative study of logarithmic image processing models for medical image enhancement. 2016 IEEE international conference on systems, man, and cybernetics (SMC); 2016: IEEE.
35. Zhou L, Liu H, Bae J, He J, Samaras D, Prasanna PJapa. Self pre-training with masked autoencoders for medical image analysis. 2022.
36. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin IJAinips. Attention is all you need. *Advances in neural information processing systems* 2017;30.

37. Zhao H, Jia J, Koltun V, editors. Exploring self-attention for image recognition. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2020.
38. Wang X, Girshick R, Gupta A, He K, editors. Non-local neural networks. Proceedings of the IEEE conference on computer vision and pattern recognition; 2018.
39. Hu J, Shen L, Sun G, editors. Squeeze-and-excitation networks. Proceedings of the IEEE conference on computer vision and pattern recognition; 2018.
40. Li X, Wang W, Hu X, Yang J, editors. Selective kernel networks. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2019.
41. Hou Q, Zhou D, Feng J, editors. Coordinate attention for efficient mobile network design. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2021.
42. Fu J, Liu J, Tian H, Li Y, Bao Y, Fang Z, Lu H, editors. Dual attention network for scene segmentation. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2019.
43. Woo S, Park J, Lee JY, Kweon IS. Cbam: Convolutional block attention module. Proceedings of the European conference on computer vision (ECCV). 2018:3-19.
44. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA; 2009:248-55.
45. Larochelle H, Bengio Y, Louradour J, Lamblin P. Exploring strategies for training deep neural networks. J Mach Learn Res 2009;1-40.
46. Goyal P, Dollár P, Girshick R, Noordhuis P, Wesolowski L, Kyrola A, Tulloch A, Jia Y, He K. Accurate, Large Minibatch SGD: Training ImageNet in 1 Hour. 2017. doi: 10.48550/arXiv.1706.02677
47. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Gregory Chanan G, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. Advances in Neural Information Processing Systems 32 (NeurIPS 2019). 2019.
48. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy; 2017:618-26.

**Cite this article as:** Wang J, Huang H, Wang H, Wei M, Yi Z, Wang Z, Zhang H. An attention-based neural network model for automatic partition of abdominal lymph nodes in CT imaging. *Quant Imaging Med Surg* 2023;13(9):5483-5501. doi: 10.21037/qims-22-1412