Appendix 1 Supplementary methods

Backbone and feature pyramid network (FPN) modules of the detection system

ResNet101, pre-trained on ImageNet, was used as the backbone for extracting image features, which consists of 5 convolution blocks, with the last 4 blocks having 3, 4, 23, and 3 residual blocks, respectively. The specific structure is shown below.

Layer name	Output size	Structure
Conv1	112×112	7×7,64, stride 2
Conv2.x	56×56	3×3 maxpool, stride 2
		$\begin{bmatrix} 1\times1, & 64 \\ 3\times3, & 64 \\ 1\times1, & 256 \end{bmatrix}\times3$
Conv3.x	28×28	$\begin{bmatrix} 1\times1, & 128 \\ 3\times3, & 128 \\ 1\times1, & 512 \end{bmatrix}\times4$
Conv4.x	14×14	$\begin{bmatrix} 1\times1, & 256 \\ 3\times3, & 256 \\ 1\times1, & 1024 \end{bmatrix}\times23$
Conv5.x	7×7	$\begin{bmatrix} 1\times1, & 512 \\ 3\times3, & 512 \\ 1\times1, & 2048 \end{bmatrix}\times3$

The multi-scale FPN was constructed based on the backbone model. The structure of FPN is divided into 3 parts: bottom-up branch, top-down branch, and transverse connection. The bottom-up branch is the forward propagation process of the backbone network, which computes feature maps (FMs) of ResNet101 at various scales and levels. In the process of forward propagation, the size of FMs will change with the depth of layers. The top-down branch performs bilinear interpolation for the deep FMs with fuzzy position information but rich semantic information to match the FMs of different scales. In the transverse connection, the FMs are fused in the form of element-level addition, in which 1×1 convolution is used to reduce the number of channels. After transverse connection of C2, C3, C4, C5, the fused FM set: M2, M3, M4, M5 is obtained. The 3×3 convolution operation is then performed on each fused FM to reduce the feature discontinuity caused by the superposition and fusion of FMs. Finally, the predicted FM set: P2, P3, P4, P5, P6 is obtained for generating candidate regions. P6 is obtained by FM P5 through down-sampling, and is also used for generating candidate regions for RPN.

RPN module for generating proposal boxes

The RPN module uses a 3×3 convolution with 2 adjacent 1×1 convolutions as sliding windows over all scales of the shared feature set: P2, P3, P4, P5, P6, then combines 9 anchors of different size to generate candidate boxes. Since the size and spatial location of each shared FM are different, it is not necessary to design multi-scale anchors in the FM of a specific scale. Instead, a single scale anchor was assigned for each scale of the shared FM. According to the statistics of lesion sizes, we mapped the receptive field corresponding to the shared FM set: P2, P3, P4, P5, P6 of all scales in FPN, so the anchor sizes designed in this study were 36^2, 72^2, 144^2, 288^2 and 576^2. The anchor at each scale has aspect ratios of 1:1, 1:2 and 2:1, so the RPN contains 15 different sizes of anchors.

Classification and regression networks

The prediction (classification and regression) module used in the FPN systems is named the basic prediction network. In the training phase, the shared FM generated by the backbone and PPN and the proposals (B0) generated by region proposal network (RPN) are fed into the region of interest (ROI) pooling layer, which transforms each input ROI and corresponding FMs into a map of fixed size. After passing through a convolution module identical with the conv5.x module in Resnet101, 2 fully connected layers are used for the classification and regression to generate the final classification (C) and regression results (B).

The prediction module used in the cascade feature pyramid network (CFPN) system is termed the cascade prediction network. Different from the basic prediction network, 3 cascade detectors are used here for the classification and regression. The structure of each detector is basically the same as that in the FPN, and the intersection over union (IoU) thresholds of the 3 detectors are set as 0.5. 0.6, and 0.7, respectively. In addition, more accurate ROI alignment is used to replace the traditional

https://dx.doi.org/10.21037/qims-22-323

ROI pooling to reduce the area migration problem. Shared FM and proposal (B0) generated by RPN are first fed to the ROI Align ("Pool") to generate a map of the same size. After passing through the convolution module and fully connected layers, the classification result (C1) and regression result (B1) of the first detector are obtained. The regression results (B1) of the previous detector and shared FMs are then sent to the next detector for training, and the final classification results (C3) and regression results (B3) are obtained after the iteration process was completed.

Table S1 AP Values of the FPN System with different settings

Settings	AP for benign, mean ± SD	AP for malignant, mean ± SD	mAP
FPN	0.397±0.070	0.869±0.027	0.633±0.032
FPN with augmentation	0.568±0.112	0.875±0.050	0.721±0.076
FPN with focal loss	0.626±0.076	0.917±0.050	0.772±0.045
FPN with both augmentation and focal loss	0.647±0.056	0.922±0.051	0.785±0.037

AP, average precision; FPN, feature pyramid network; mAP, mean average precision; SD, standard deviation.

Table S2 Patient- and slice-level precision, recall, and F1-score in real detection

System	Benign, mean ± SD			Malignant, mean ± SD		
	Precision	Recall	F1-score	Precision	Recall	F1-score
Slice level						
Faster R-CNN	0.811±0.101	0.658±0.089	0.725±0.090	0.922±0.024	0.963±0.022	0.942±0.021
FPN	0.844±0.069	0.561±0.094	0.669±0.069	0.905±0.024	0.976±0.011	0.939±0.014
CFPN	0.829±0.056	0.708±0.101	0.761±0.076	0.933±0.025	0.965±0.013	0.949±0.017
Patient level						
Faster R-CNN	0.812±0.105	0.738±0.102	0.768±0.086	0.944±0.021	0.962±0.024	0.953±0.017
FPN	0.865±0.126	0.647±0.111	0.729±0.070	0.928±0.022	0.974±0.024	0.950±0.012
CFPN	0.860±0.116	0.800±0.146	0.816±0.085	0.958±0.030	0.968±0.029	0.962±0.015

CFPN, cascade feature pyramid network; FPN, feature pyramid network; Faster R-CNN, faster region-based convolutional neural network; SD, standard deviation.

Table S3 Patient- and slice-level precision, recall, and F1-score for large breast lesions in real detection

System	Benign, mean ± SD			Malignant, mean ± SD		
	Precision	Recall	F1-score	Precision	Recall	F1-score
Slice level						
Faster R-CNN	0.814±0.223	0.559±0.266	0.621±0.186	0.936±0.044	0.975±0.026	0.953±0.019
FPN	0.778±0.187	0.503±0.146	0.600±0.141	0.925±0.023	0.977±0.019	0.950±0.013
CFPN	0.782±0.246	0.578±0.238	0.646±0.207	0.938±0.034	0.974±0.024	0.954±0.020
Patient level						
Faster R-CNN	0.792±0.216	0.500±0.373	0.500±0.332	0.945±0.036	0.977±0.023	0.960±0.018
FPN	0.875±0.218	0.625±0.247	0.717±0.218	0.956±0.031	0.989±0.019	0.972±0.024
CFPN	0.833±0.289	0.625±0.247	0.700±0.243	0.955±0.032	0.977±0.039	0.963±0.034

CFPN, cascade feature pyramid network; FPN, feature pyramid network; Faster R-CNN, faster region-based convolutional neural network; SD, standard deviation.

https://dx.doi.org/10.21037/qims-22-323