文章编号:1671-4598(2024)02-0056-09 DOI:10.16526/j.cnki.11-4762/tp.2024.02.009 中图分类号:TP391 文献标识码:A

基于注意力机制和多空间金字塔池化的实时目标检测算法

王国刚，李泽欣，董志豪
（山西大学 物理电子工程学院，太原 030006）

摘要：YOLOv4计算复杂度高，空间金字塔池化模块仅一次增强特征融合网络的深层区域特征图的表征能力、检测头网络的特征图难以突出重要通道特征；针对以上问题，提出一种基于注意力机制和多空间金字塔池化的实时目标检测算法；该算法采用多空间金字塔池化，提取局部特征和全局特征，融合多重感受野，加强特征融合网络的浅、中、深层特征图的表征能力；引入压缩激励通道注意力机制，建模通道间的相关性，自适应调整特征图各个通道的权重，从而使网络更加关注重要特征；特征融合和检测头网络中使用深度可分离卷积，减少了网络参数量；实验结果表明，所提算法的均值平均精度均高于其他7种主流对比算法；与YOLOv4相比，参数量、模型大小分别减少了27.85 M和106.25 MB，所提算法在降低复杂度的同时，提高了检测准确度，且该算法的检测速率达到33.70帧/秒，满足实时性要求。

关键词：YOLOv4；通道注意力；空间金字塔池化；感受野；深度可分离卷积；实时性

Real-Time Object Detection Algorithm Based on Attention Mechanism and Multi-spatial Pyramid Pooling

WANG Guogang, LI Zexin, DONG Zhihao
(College of Physics and Electronic Engineering, Shanxi University, Taiyuan 030006, China)

Abstract: Aimed at the disadvantages of an enhancement to the representation of deep feature map in the enhanced feature fusion network for the spatial pyramid pooling module, higher computational complexity, and difficulty in highlighting important channel features for the feature map of the detection head network in YOLOv4 algorithm, Based on this problems, a real-time object detection algorithm based on attention mechanism and multi-spatial pyramid pooling is proposed. This algorithm adopts multi-spatial pyramid pooling, extracts the local and global features, fuses multiple receptive fields, and strengths the characterization ability of the shallow, middle and deep feature maps for the feature fusion network. The squeeze-and-excitation channel attention mechanism is introduced to model the relativities between channels, the weight of each channel is adaptively recalibrated to make the network pay more attention to important features. Moreover, the deep separable convolution is used to reduce the parameters of the feature fusion and detection head networks. The experimental results show that the mean average precision (mPA) of the proposed algorithm is higher than that of other 7 mainstream comparison algorithms, compared with YOLOv4, the parameters and model size are reduced by 27.85 M and 106.25 MB, respectively. The proposed algorithm not only improves the detection accuracy, but also reduces the computational complexity compared to the baseline algorithm, and the average speed of the algorithm reaches by 33.70 FPS, which meets the real-time requirement.

Keywords: YOLOv4; channel attention; spatial pyramid pooling; receptive field; depthwise separable convolution; real-time

0 引言

作为计算机视觉领域的重要研究课题，目标检测的目的是解决目标分类及定位问题。目前，目标检测已广泛应用于智能视频监控[1-3]、无人驾驶[4-5]、智能化交通[6-7]、人脸识别[8-9]、遥感影像分析[10-12]和医学图像检测[13-15]等领域。

传统目标检测算法主要分3个阶段来完成目标检测任务，首先进行区域的选择，采用不同大小的滑动窗口遍历截取图像，产生多个候选框来尽可能检测所有存在的目标；然后进行特征的提取，由于传统目标检测方法的特征大多是人为手动设计的，因此使用人工设计的特征提取器提取特征；最后提取到特征后，利用支持向量机[16]、可变形部件模型[17]和AdaBoost[18]等分类器来判别目标的类别。该类

收稿日期:2023-03-08； **修回日期**:2023-04-20。
基金项目:国家自然科学基金(11804209)；山西省自然科学基金(201901D111031，201901D211173)；山西省高校科技创新计划(2019L0064，2020L0051)。
作者简介:王国刚(1977-)，男，博士，副教授。
引用格式:王国刚，李泽欣，董志豪.基于注意力机制和多空间金字塔池化的实时目标检测算法[J].计算机测量与控制，2024，32(2)：56-64.

算法通过工程人员的经验提取特征，检测精度低，难以满足实际需求，具有一定的局限性。

2014 年，Girshick 等提出的 R-CNN[19] 算法，成功将深度学习应用到目标检测领域。相较于传统算法，基于深度学习的目标检测算法在性能上更具优势。这类算法可分为两种，一种是以 SSD[20]，Efficientdet[21] 为代表的单阶段目标检测算法，这类算法直接在整张图片上生成若干候选框，再对这些候选框进行类别、框大小以及位置的回归预测。另一种是以 R-CNN，Fast R-CNN[22]，Faster R-CNN[23] 为代表的两阶段目标检测算法，这类算法第一阶段先得到存在物品的候选区域，第二阶段再对候选区域中的候选框进行类别、框大小以及位置的回归预测。

深度学习目标检测算法采用大量特征来提升性能。但是，基于某些特征的目标检测器泛化性能差，采用了加权残差连接（WRC，weighted-residual-connections）、跨阶段部分连接（CSP，cross-stage-partial-connections）等特征的 YOLOv4[24] 目标检测器的泛化性能得到了一定程度的改善。然而，YOLOv4 中的空间金字塔池化（SPP，spatial pyramid pooling）模块未充分提高特征融合网络浅、中、深层特征图的表征能力；检测头网络特征图的每个通道权重都相同，难以凸出重要通道上的信息；且使用了更复杂的网络结构，增大了模型参数量，增加了检测时间。

针对以上问题，提出一种基于注意力机制和多空间金字塔池化的实时目标检测算法（AMMP，real-time object detection algorithm based on attention mechanism and multi-spatial pyramid pooling）。该算法采用多空间金字塔池化，提取多尺度信息，加强特征融合网络的浅、中、深层特征图的表征能力；引入压缩激励通道注意力机制，建模通道间的相关性，自适应调整特征图各个通道的权重，从而使网络更加关注重要特征；特征融合和检测头网络中使用深度可分离卷积，减少网络参数量，提高算法检测速度。实验结果表明，AMMP 算法的均值平均精度（mAP，mean average precision）优于 7 种主流对比算法。与基准算法相比，该算法在降低复杂度的同时，提高了检测准确度。

1　相关网络

YOLOv4 由 Backbone 特征提取网络、Neck 特征融合网络和 Head 检测头网络三部分组成。Backbone、Neck 分别采用 CSPDarknet53 网络和加入了 SPP 结构的路径聚合网络[25]（PANet，path aggregation network）。CSPDarknet53 将 CSPNet 结构应用于 Darknet53 网络中，加速了特征提取过程，该网络由 5 个大残差块组成，这 5 个大残差块包含的小残差单元个数分别为 1，2，8，8，4，即两个 CBM 卷积模块和一个 CSPX 卷积模块共同组成一个大残差块。SPP 结构采用 3 种不同尺寸的池化核对上层特征图做最大池化处理，扩大了网络的感受野。PANet 由上采样和下采样两部分组成，实现了主干网络深层特征与浅层特征的融合，提高了网络的检测精度。Head 检测头网络有 3 个检测头，每个检测头包含 3 组候选框的调整参数，每组候选框的调整参数包含 1 个置信度参数、4 个调整长宽和坐标偏移量的参数和 20 个待检测的类别参数。通过这些参数，YOLOv4 调整候选框的中心点坐标和宽高，生成最终的预测框。

2　基于注意力机制和多空间金字塔池化的实时目标检测算法

2.1　多空间金字塔池化

为提高目标检测准确度，AMMP 在特征融合网络构建多空间金字塔池化模块（MSPP，multi-spatialpyramid pooling module）。MSPP 包含深层空间金字塔池化模块（D-SPP，deep-spatialpyramid pooling module）、中层空间金字塔池化模块（M-SPP，middle-spatialpyramid pooling module）和浅层空间金字塔池化模块（S-SPP，shallow-spatial-pyramid pooling module）3 部分，如图 1 所示。MSPP 借鉴空间金字塔思想，提取局部特征和全局特征，融合多重感受野，从而扩大了主干特征的接收范围，分离出了重要的上下文信息，提高了模型的检测精度。

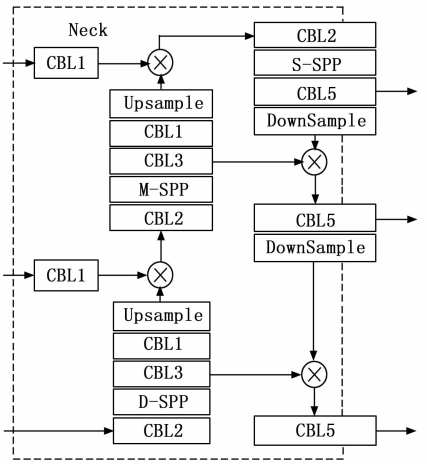

图 1　AMMP 的特征融合网络

D-SPP、M-SPP 和 S-SPP 采用 1×1 卷积减少通道数，再经 3 种不同尺寸的池化核做最大池化处理，提取不同尺度的特征信息，最后，采用 concat 操作，在通道维度上连结各个分支的特征图，使输出特征图通道数变为原来的四倍，从而增强了特征图的表征能力，如图 2 所示。MSPP 三部分采用的最大池化操作的池化核大小为 $K\times K$，步长为 1，填充 $padding=(K-1)/2$，$K\in\{5,9,13\}$。

位于 AMMP 特征融合网络深层特征区域的 D-SPP，提高了深层特征区域特征图的表征能力。深层特征区域特征图经过上采样、特征融合传入中层特征区域。M-SPP 位于中层特征区域，提高了中、深层特征区域特征图的表征能力。中层特征区域特征图经上采样和特征融合传入浅层特征区域。位于浅层特征区域的 S-SPP，提高了浅、中、深层特征区域特征图的表征能力。网络前向传播过程中，特征融合网络深层区域特征图的表征能力得到了 3 次增强，中

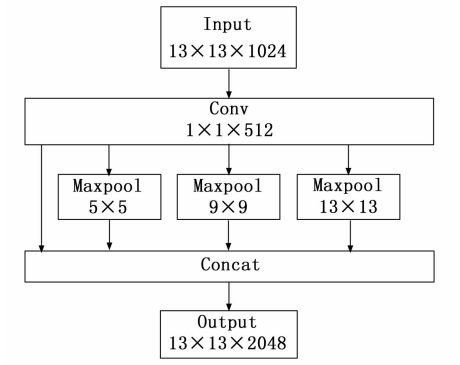

图 2 D-SPP、M-SPP 或 S-SPP 结构

层区域特征图的表征能力得到了两次增强，浅层区域特征图的表征能力得到了一次增强。最后下采样过程中，将不同尺寸的特征图送入 AMMP 的 3 个检测头进行检测，有效提升了模型对大中小 3 种目标的检测能力。

AMMP 特征融合网络通过使用 D-SPP、M-SPP 以及 S-SPP 模块，融合了不同尺度的特征信息，加强了浅、中、深层特征图的表征能力，提高了检测精度。

2.2 压缩激励通道注意力模块

为使最深卷积层关注重要通道信息，AMMP 在检测头网络构建压缩激励通道注意力模块（SECAM，squeeze-and-excitationchannel attention module），如图 3 所示。SECAM 自适应调整最深卷积层中特征图的各通道权重，强化权值高的通道特征，抑制权值低的通道特征，使检测头网络的特征信息表达更充分，从而提升模型检测性能。

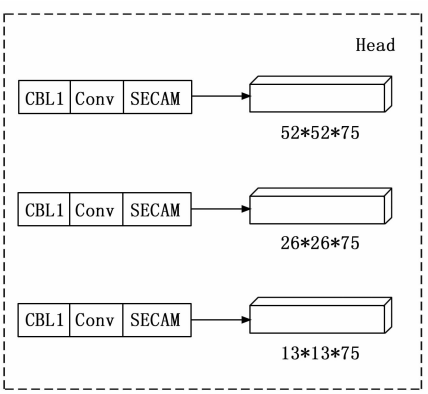

图 3 AMMP 的检测头网络

AMMP 检测头网络中，特征图的不同通道表示不同目标。SECAM 通过建模通道间的相关性，自适应调整特征图各个通道的权重，从而实现目标的自适应选择，使网络更加关注重要通道特征，如图 4 所示。SECAM 通过压缩，激励和重校准 3 个操作得到加权后的特征图。压缩是对特征图 U 的每个通道进行的全局平均池化操作，如公式（1）所示：

$$z = F_{sq}(U) = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W}U(i,j) \tag{1}$$

式中，z 表示压缩通道后得到的特征向量，H 和 W 表示特征图 U 的高度和宽度。

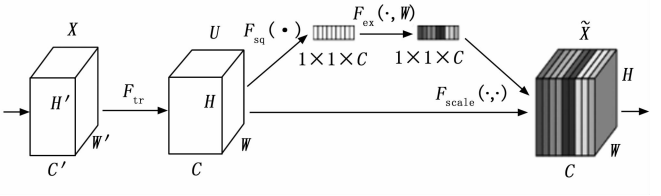

图 4 压缩激励通道注意力模块

激励通过激活函数建模特征通道间的相关性以生成每个通道的权重，如式（2）所示：

$$s = F_{ex}(z,W) = \sigma(W_2\delta(W_1 z)) \tag{2}$$

式中，s 表示生成的通道注意力权重，δ 和 σ 分别表示激活函数 ReLU 和 Sigmoid，W_1 和 W_2 表示用于降维操作的特征矩阵。

重校准将每个通道乘以相应的权重，得到重新标定的特征，如式（3）所示：

$$\tilde{X}_c = F_{scale}(u_c, s_c) = s_c \cdot u_c \tag{3}$$

AMMP 在检测头网络构建压缩激励通道注意力模块。SECAM 调整最深卷积层各通道之间的权重，从而降低干扰信息的影响，强化检测头网络提取特征信息的能力，提升模型的检测性能。

2.3 深度可分离卷积

若标准卷积输入特征图 F、卷积核 K 的尺寸分别为 $D_F \times D_F \times M$ 和 $D_K \times D_K \times M \times N$，得到 $D_F \times D_F \times N$ 的输出特征图 G，则 G 可由公式（4）计算得到：

$$G_{k,l,n} = \sum_{i,j,m} K_{i,j,m,n} \cdot F_{k+i-1,l+j-1,m} \tag{4}$$

由式（4）可知，标准卷积的计算成本为：

$$P_1 = D_K \times D_K \times M \times N \times D_F \times D_F \tag{5}$$

其中：D_K 表示卷积核大小，D_F 表示输入特征图的高度或宽度，M 和 N 分别表示输入、输出通道数，P_1 表示标准卷积的计算量。

由公式（4）和（5）可知，标准卷积是将卷积核作用在输入特征图的所有通道上，这增大了模型参数量，增加了目标检测时间。为克服这一缺陷，AMMP 在图 5 所示的特征融合和检测头网络中使用深度可分离卷积（DSC，depthwise separable convolution），以减少参数量，提高目标检测速度。

深度可分离卷积将标准卷积分解为逐深度卷积和逐点卷积两个操作，从而以较小的精度代价减少了网络计算量。逐深度卷积将不同的卷积核作用在不同的输入通道上，以学习空间特征；逐点卷积将 1×1 的卷积核作用在逐深度卷积操作得到的通道上，以学习通道特征。

若逐深度卷积输入特征图 F、卷积核的尺寸分别为 $D_F \times D_F \times M$ 和 $D_K \times D_K \times M$，得到 $D_F \times D_F \times M$ 的特征图 $\hat{G}$，则 $\hat{G}$ 可由公式（6）计算得到：

$$\hat{G}_{k,l,m} = \sum_{i,j} \hat{K}_{i,j,m} \cdot F_{k+i-1,l+j-1,m} \tag{6}$$

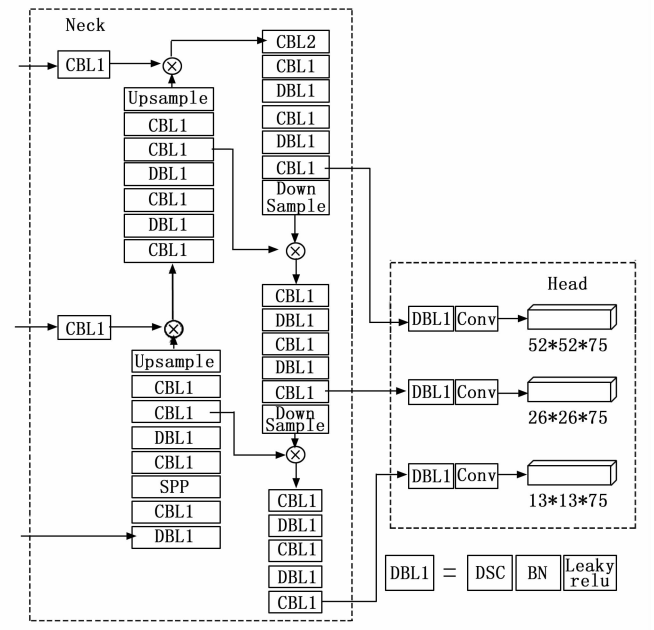

图 5　AMMP 的特征融合和检测头网络

由式（6）可知，逐深度卷积的计算成本为：

$$P_2 = D_K \times D_K \times M \times D_F \times D_F \tag{7}$$

逐深度卷积未对过滤后的信息进行组合。为解决这一问题，深度可分离卷积在逐深度卷积后采用了逐点卷积。若逐点卷积核尺寸为 $1\times1\times M\times N$，则其计算量可由公式（8）得到：

$$P_3 = M \times N \times D_F \times D_F \tag{8}$$

深度可分离卷积由逐深度卷积和逐点卷积构成，故其计算量为：

$$P_4 = P_2 + P_3 = D_K \times D_K \times M \times D_F \times D_F + M \times N \times D_F \times D_F \tag{9}$$

由公式（5）和（9）可得深度可分离卷积与标准卷积计算量之比，如公式（10）所示：

$$\frac{P_4}{P_1} = \frac{D_K \times D_K \times M \times D_F \times D_F + M \times N \times D_F \times D_F}{D_K \times D_K \times M \times N \times D_F \times D_F} = \frac{1}{D_K^2} + \frac{1}{N} \tag{10}$$

由公式（10）可知，卷积核尺寸或输出特征图通道数越大，深度可分离卷积计算量越小。若 N 充分大，则深度可分离卷积比标准卷积减少约（D_k^2-1）倍的计算量。

与基准模型相比，采用了深度可分离卷积的 AMMP 模型，减少了参数量和模型大小，且在保持较高检测精度的前提下，提高了检测效率。

2.4　AMMP 网络模型

AMMP 网络模型如图 6 所示。AMMP 在特征融合网络

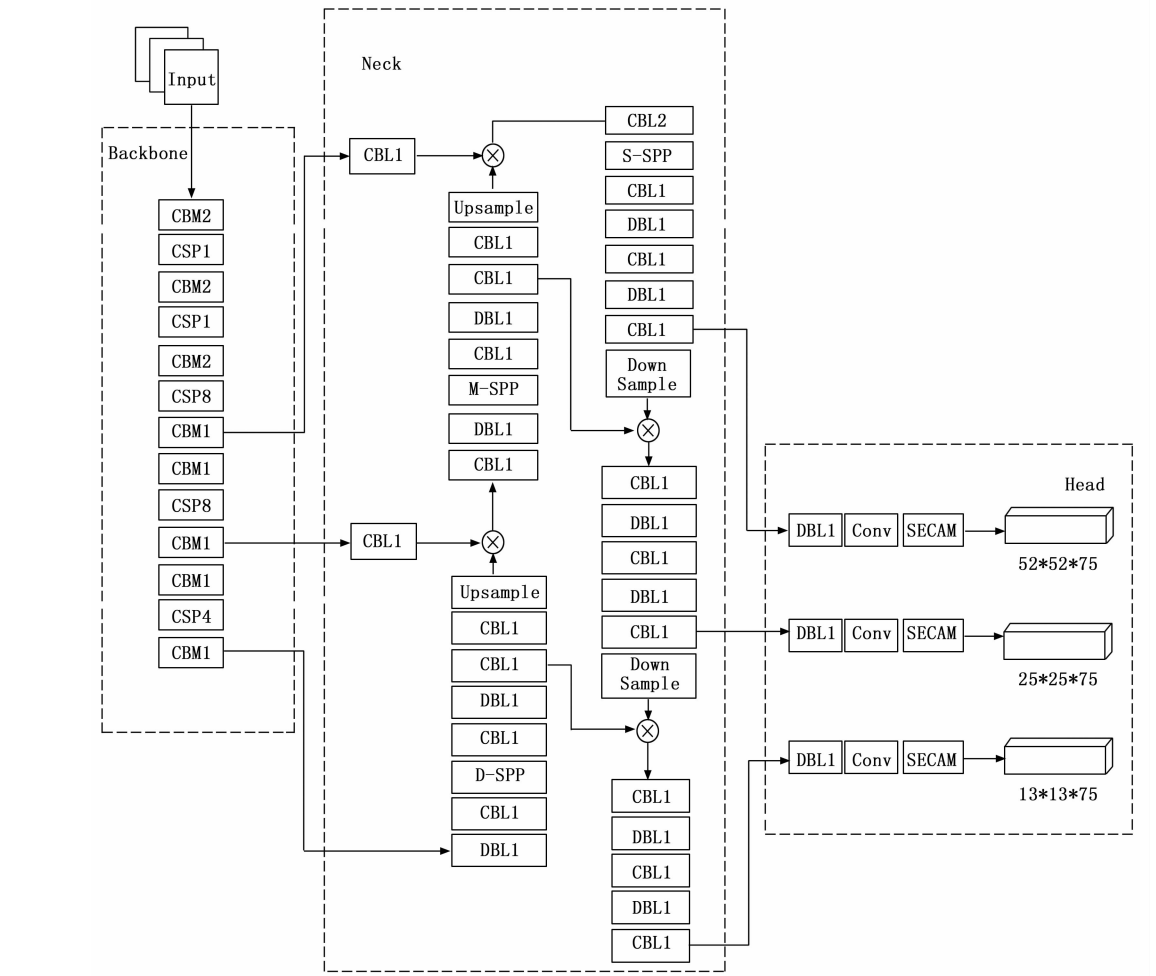

图 6　AMMP 网络模型示意图

构建多空间金字塔池化模块，提取多尺度特征信息，提高 Neck 网络浅、中、深层特征图的表征能力；在检测头网络构建压缩激励通道注意力模块，自适应调整特征图各个通道的权重，强化高权值通道特征的学习，提升模型鲁棒性；特征融合以及检测头网络中使用深度可分离卷积，在保持较高检测精度的前提下，提高模型检测速度。

3 实验结果及分析

3.1 数据集与实验环境配置

实验采用 PASCAL VOC2007 数据集。该数据集包含 20 个类别，5 011 张训练图片，4 952 张测试图片。

训练采用的工作站配置为：Intel（R）Xeon（R）CPU，Tesla T4 GPU，167 G 内存；测试采用的 PC 机配置为：Intel® Core™i5-8265U CPU，NVIDIA GeForce MX250 GPU，8 G 内存。深度学习框架为 pytorch1.2，GPU 加速库版本为 CUDA10.0，Cudnn7.4.1.5。

训练阶段，采用 Adam 优化器更新网络参数，动量、权重衰减系数分别设置为 0.9 和 0.000 5。训练分两个阶段，第一阶段冻结主干网络，调整非主干网络参数；第二阶段释放主干网络，调整整个网络参数。第一阶段学习率设置为 0.001，Batch size 设置为 16，epoch 设置为 20；第二阶段学习率设置为 0.000 1，Batch size 设置为 8，epoch 设置为 100。

3.2 评价指标

实验采用 F_1、查准率（P，precision）、查全率（R，recall）、平均精度（AP，average precision）和均值平均精度 *mAP* 来评估算法性能；且采用参数量、模型大小和每秒传输帧数（FPS，frames per second）来评价算法复杂度。此外，通过绘制查准率－查全率（P-R，precision-recall）曲线和 F_1 曲线，直观地比较不同算法的检测性能。

3.3 定量分析

表 1 给出了 8 种算法在 PASCAL VOC2007 数据集上的 *AP* 和 *mAP* 值。由表 1 可知，AMMP 算法的 5 个类别的 *AP* 值为最优，6 个类别的 *AP* 值为次优，8 个类别的 *AP* 值排名第三。且与 SSD、YOLOv3、YOLOv4-tiny、YOLOv4、YOLOv5m、Centernet、Efficientdet-D0 算法相比，AMMP 算法的 *mAP* 分别提升了 8.99%、3.83%、8.36%、0.16%、1.92%、7.51%、1.97%。

置信度阈值变化，查准率和查全率也随之变化。实验设置不同的置信度阈值，绘制 *P-R* 曲线和 F_1 曲线，以直观比较不同算法的检测性能。通过分析 PASCAL VOC2007 的测试集可知，car 是最难检测的类别之一。通过分析 PASCAL VOC2007 的训练集可知，boat 是最能体现样本不均衡问题的类别之一。故实验选用 car 和 boat 为代表，绘制八种算法在这两种类别上的 *P-R* 曲线和 F_1 曲线，如图 7～10 所示。

P-R 曲线与横轴所围面积越大，算法性能越好。由图 7 可知，AMMP 算法 car 类别的 *P-R* 曲线和横轴间的面积最大，故八种算法中，AMMP 算法对代表类别 car 的检测性能最好。

表 1 8 种算法的 *AP* 以及 *mAP*

算法	*mAP*/%	各类别平均精度 *AP*/%						
SSD	74.72	Aero	Bicycle	Bird	Boat	Bottle	Bus	Car
		79	84	73	68	44	86	87
		Cat	Chair	Cow	D-table	Dog	Horse	M-bike
		87	53	79	75	84	87	83
		Person	P-plant	Sheep	Sofa	Train	Tv	
		78	45	74	72	86	73	
YOLOv3	79.88	Aero	Bicycle	Bird	Boat	Bottle	Bus	Car
		89	87	80	69	70	87	93
		Cat	Chair	Cow	D-table	Dog	Horse	M-bike
		86	62	80	**79**	83	87	87
		Person	P-plant	Sheep	Sofa	Train	Tv	
		89	52	76	77	87	79	
YOLOv4-tiny	75.35	Aero	Bicycle	Bird	Boat	Bottle	Bus	Car
		81	83	70	61	64	81	89
		Cat	Chair	Cow	D-table	Dog	Horse	M-bike
		79	60	81	71	73	85	84
		Person	P-plant	Sheep	Sofa	Train	Tv	
		85	48	77	72	83	79	
YOLOv4	83.55	Aero	Bicycle	Bird	Boat	Bottle	Bus	Car
		85	87	85	64	76	**92**	**94**
		Cat	Chair	Cow	D-table	Dog	Horse	M-bike
		93	73	**93**	77	89	91	91
		Person	P-plant	Sheep	Sofa	Train	Tv	
		91	**59**	**87**	71	90	**85**	
YOLOv5m	81.79	Aero	Bicycle	Bird	Boat	Bottle	Bus	Car
		94	86	**88**	68	**79**	87	93
		Cat	Chair	Cow	D-table	Dog	Horse	M-bike
		89	**76**	91	46	85	**95**	**92**
		Person	P-plant	Sheep	Sofa	Train	Tv	
		93	58	76	67	91	82	
Centernet	76.20	Aero	Bicycle	Bird	Boat	Bottle	Bus	Car
		84	86	76	66	57	80	88
		Cat	Chair	Cow	D-table	Dog	Horse	M-bike
		85	58	82	70	80	87	85
		Person	P-plant	Sheep	Sofa	Train	Tv	
		83	47	76	74	83	77	
Efficientdet-D0	81.74	Aero	Bicycle	Bird	Boat	Bottle	Bus	Car
		87	88	86	**73**	50	**92**	**94**
		Cat	Chair	Cow	D-table	Dog	Horse	M-bike
		90	70	90	77	**90**	91	88
		Person	P-plant	Sheep	Sofa	Train	Tv	
		84	52	83	**82**	91	76	
AMMP	**83.71**	Aero	Bicycle	Bird	Boat	Bottle	Bus	Car
		87	**90**	**85**	**73**	75	91	**94**
		Cat	Chair	Cow	D-table	Dog	Horse	M-bike
		90	70	87	77	**90**	90	89
		Person	P-plant	Sheep	Sofa	Train	Tv	
		90	57	83	78	**93**	84	

注：加粗数字和下划线数字分别表示最优、次优结果。

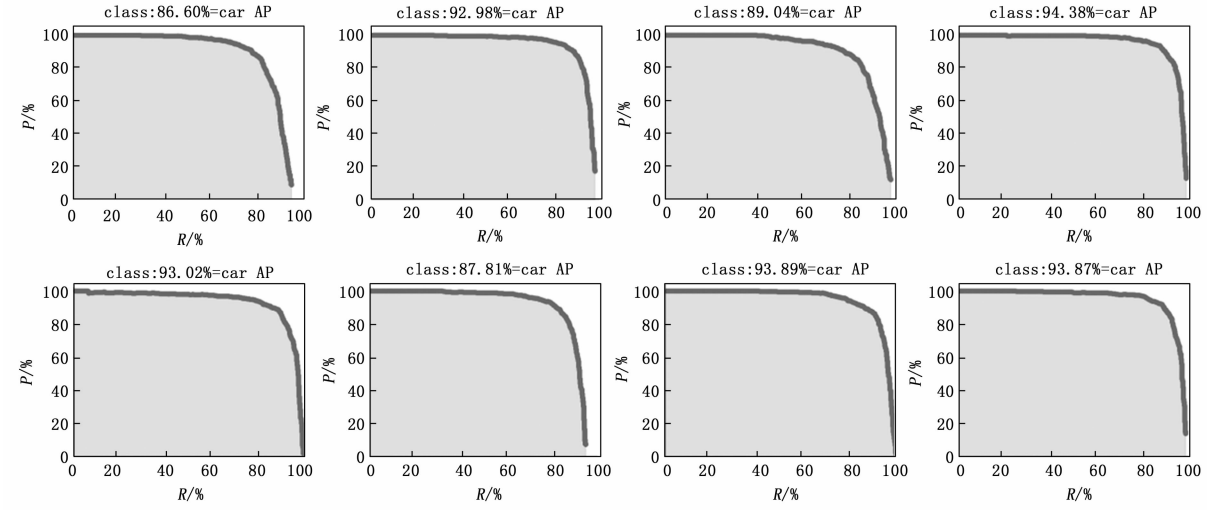

图 7　car 类别的 *P-R* 曲线比较

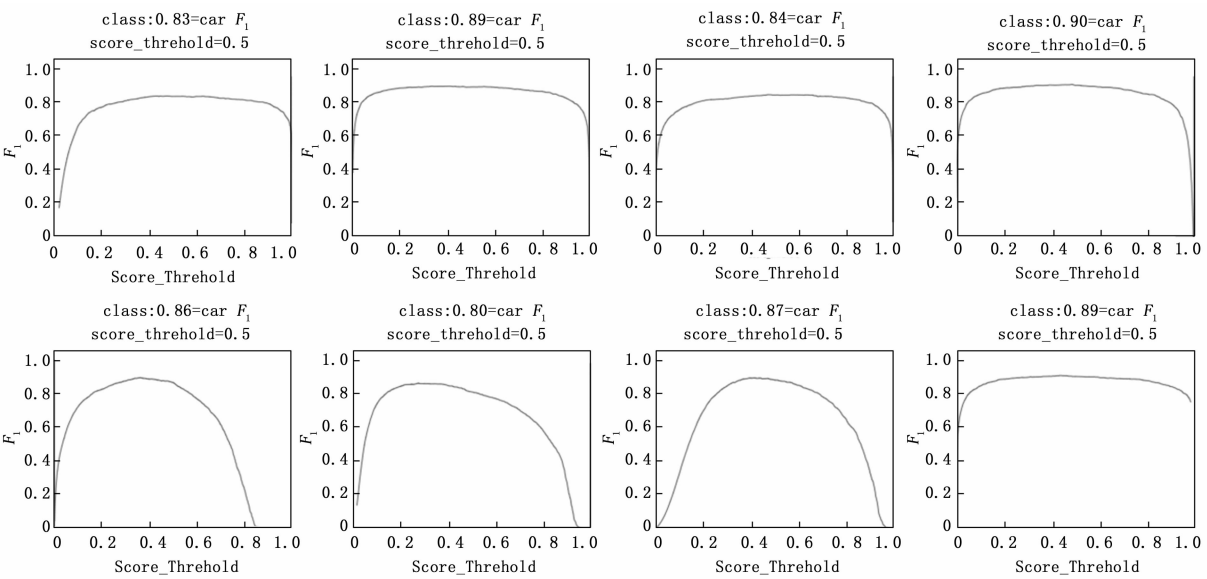

图 8　car 类别的 F_1 曲线比较

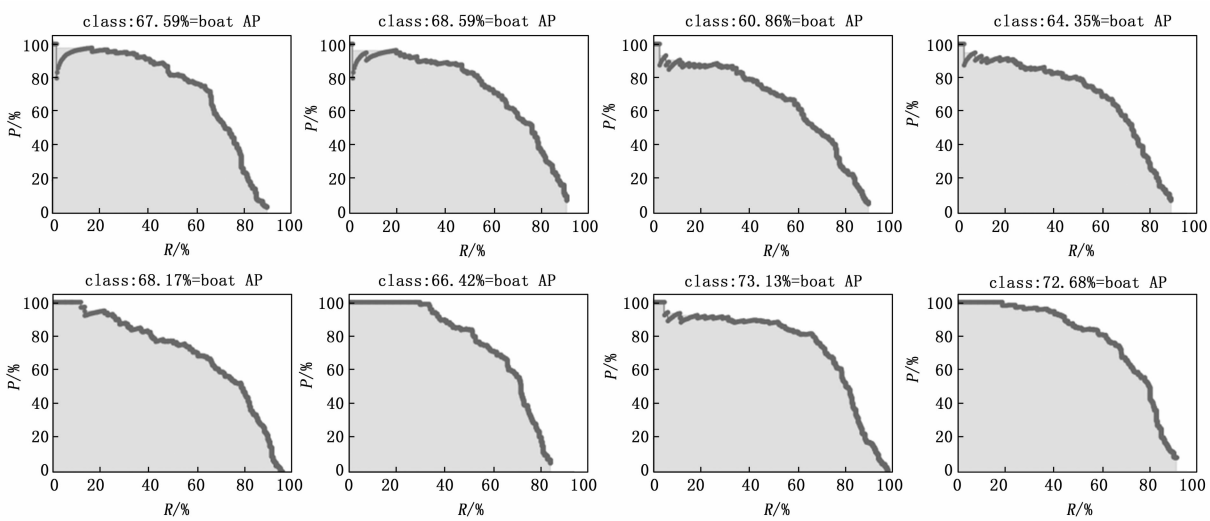

图 9　boat 类别的 *P-R* 曲线比较

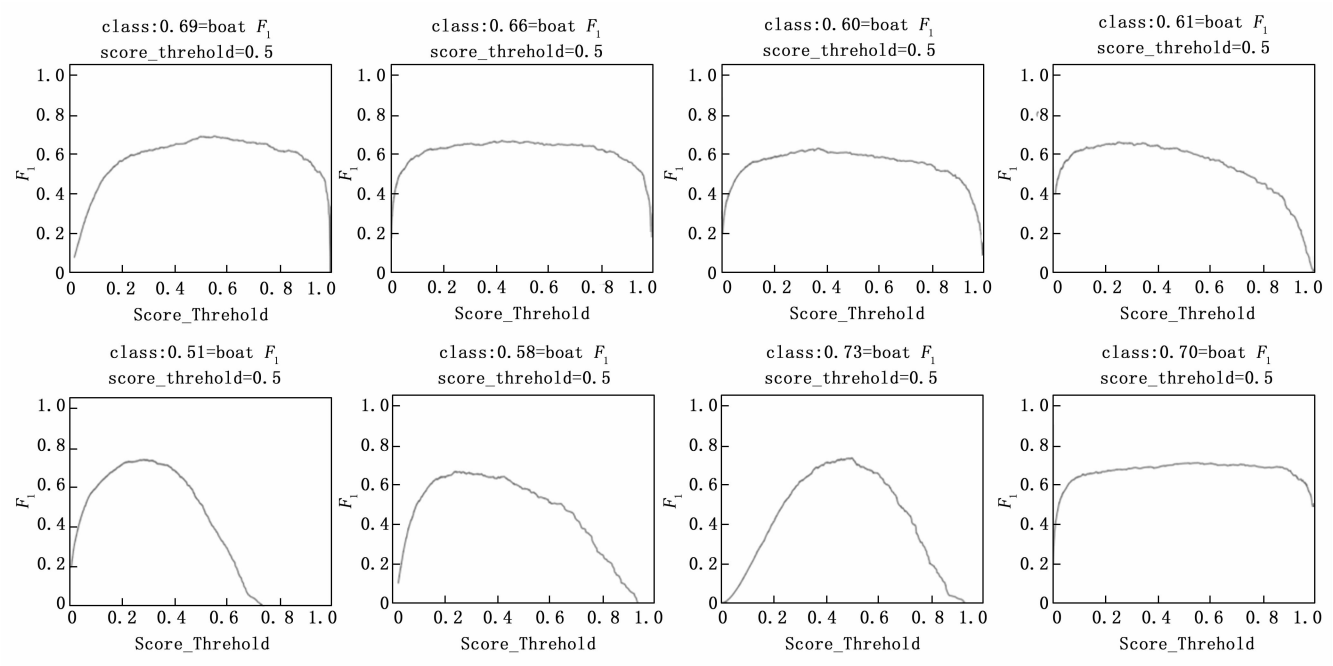

图 10 boat 类别的 F_1 曲线比较

正负样本不均衡时，*P-R* 曲线越平稳且和横轴间的面积越大，算法性能越好。由图 9 可知，AMMP 算法 boat 类别的 *P-R* 曲线最平稳且和横轴间的面积最大，故样本不均衡情况下，AMMP 算法对代表类别 boat 的检测性能优于其他算法。

F_1 曲线越平稳且越接近于 1，算法的检测效果越好。由图 8、图 10 可知，不管是 car 类别还是 boat 类别，AMMP 算法的 F_1 曲线最平稳且最接近于 1，所以 AMMP 算法对这两种代表类别的检测效果最好。

表 2 给出了 8 种算法在 7 项评价指标上的实验结果。由表 2 可知，8 种算法中，虽然 AMMP 算法的查准率比 Efficientdet-D0、YOLOv4、Centernet 算法略低，但 AMMP 算法的均值平均精度，查全率和 *F*1 得分均为最优。由表 2 还可看出，相比基准模型 YOLOv4，AMMP 的参数量、模型大小分别减少 27.85 M 和 106.25 MB；且帧率达到了 33.70 帧/秒，满足实时性要求。

3.4 定性分析

为直观比较 AMMP 与其他 7 种主流目标检测算法的性能，实验选取 PASCAL VOC2007 测试集中 5 张待检测图片以进行定性分析。图 11～15 为 8 种算法的检测结果对比图。

编号 001366 的待检图片 1，包含 2 个 dog 目标和 4 个 person 目标。该图片中的场景，背景复杂，存在遮挡现象。检测结果如图 11 所示，由图可知，SSD、YOLOv4-tiny、YOLOv5m 和 Centernet 算法均漏检了 dog 目标；SSD、Centernet 和 Efficientdet-D0 算法漏检了 person 目标；YOLOv3、YOLOv4 和 AMMP 算法虽无漏检问题，但 AMMP 算法的检测精度更高。

编号 003025 的待检图片 2，包含 4 个 sheep 目标和 1 个 person 目标。该图片场景存在目标重叠，遮挡现象。检测结果如图 12 所示，由图可知，SSD、YOLOv4、YOLOv5m 和 Centernet 算法均漏检了 sheep 目标；YOLOv4-tiny 算法对 sheep 目标存在重复检测问题；YOLOv3 算法 sheep 目标的定位出现严重偏差；Efficientdet-D0 算法对所有目标的定位精度均低于 AMMP 算法；AMMP 算法无漏检、重复检测及定位精度不高的问题，故性能最好。

表 2 8 种算法的客观评价指标对比

算法	*mAP*/%	*P*/%	*R*/%	F_1	参数量/M	模型大小/MB	*FPS*/(Frames·s^{-1})
SSD	74.72	76.54	69.68	0.73	26.29	100.3	52.57
YOLOv3	79.88	83.97	71.70	0.77	61.63	235.08	39.32
YOLOv4-tiny	75.35	76.18	69.58	0.73	5.92	22.58	**138.7**
YOLOv4	83.55	88.73	72.19	0.80	64.04	244.29	33.34
YOLOv5m	81.79	81.28	75.56	0.78	21.13	80.62	59.52
Centernet	76.20	**91.87**	56.40	0.70	32.67	124.61	50.16
Efficientdet-D0	81.74	88.13	67.68	0.77	**3.84**	**15.35**	22.31
AMMP	**83.71**	85.41	**77.17**	**0.81**	36.19	138.04	33.70

注：加粗数字和下划线数字分别表示最优、次优结果。

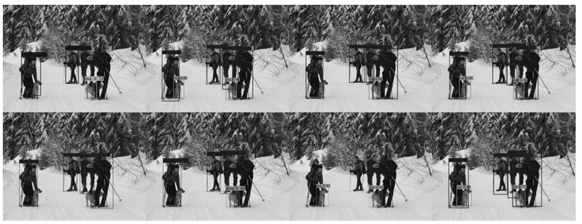

图 11　待检图片 1 的检测结果比较

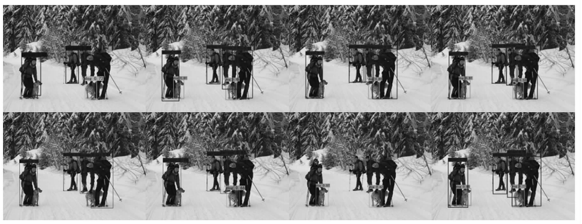

图 12　待检图片 2 的检测结果比较

编号 003858 的待检图片 3，包含 6 个 person 目标和 1 个 train 目标。该图片场景存在遮挡现象，背景阴暗。检测结果如图 13 所示，由图可知，SSD、Centernet 和 Efficientdet-D0 算法均漏检了 person 目标；SSD、YOLOv4-tiny、YOLOv4、YOLOv5m 和 Centernet 算法漏检了 train 目标；YOLOv3 算法将 train 目标错检成 boat 目标；AMMP 算法无漏检、错检问题，检测效果最好。

图 13　待检图片 3 的检测结果比较

编号 006121 的待检图片 4 包含 3 个 car 目标。该图片场景中，目标尺度变化大，背景复杂，有遮挡现象。检测结果如图 14 所示，由图可知，SSD、YOLOv3 和 YOLOv4-tiny 算法均漏检了 car 目标；YOLOv4、YOLOv5m、Centernet 和 Efficientdet-D0 算法虽无漏检问题，但其检测精度均低于 AMMP 算法。

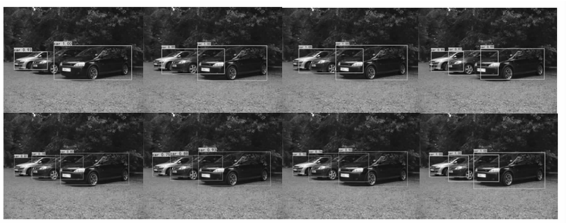

图 14　待检图片 4 的检测结果比较

编号 006771 的待检图片 5，包含 1 个 *tv* 目标，1 个 diningtable 目标以及 6 个 person 目标。该图片场景中，光照不足，背景阴暗。检测结果如图 15 所示，由图可知，YOLOv5m 漏检了 diningtable 目标；Centernet 和 Efficientdet-D0 算法漏检了 person 目标；SSD、YOLOv3、YOLOv5m、Centernet 和 Efficientdet-D0 算法漏检了 *tv* 目标；YOLOv4 算法把背景错检成 *tv* 目标。YOLOv4-tiny 和 AMMP 算法虽无漏检、错检问题，但 AMMP 算法检测精度更高。

图 15　待检图片 5 的检测结果比较

3.5　消融实验

为验证深度可分离卷积（DSC）、多空间金字塔池化（MSPP）以及压缩激励通道注意力（SECAM）方法对 YOLOv4 的优化作用，进行了消融实验，实验结果如表 3 所示。

表 3　消融实验

算法	*mAP*/%	参数量/M	模型大小/MB
YOLOv4	83.55	64.04	244.29
YOLOv4＋DSC	82.99	**35.78**	**136.49**
YOLOv4＋DSC＋MSPP	83.37	36.18	138
AMMP(YOLOv4＋DSC＋MSPP＋SECAM)	**83.71**	36.19	138.04

注：加粗数字表示最优结果。

由表中第 1、2 行可知，引入 DSC 后，*mAP* 下降 0.56%，模型大小、参数量分别下降 107.8 MB 和 28.26 M。这表明 DSC 方法仅以微小的精度代价，换取了模型大小和参数量近一半的减少。由表中第 2、3 行可知，引入 MSPP 后，*mAP* 提升了 0.38%。这是因为 MSPP 方法可提取多尺度特征信息，增强感受野，提高特征图的表征能力。由表中第 3、4 行可知，引入 SECAM 后，*mAP* 提升了 0.34%。这是因为 SECAM 可建模通道间的相关性，自适应调整特征图各通道权重，引导网络更多地关注重要特征。

由表中第 1、4 行可知，融合 DSC、MSPP 以及 SECAM 后，*mAP* 提升 0.16%，模型大小、参数量分别下降 106.25 MB 和 27.85 M。这表明，与基准算法相比，AMMP 算法综合考虑了检测精度和速度，在降低复杂度的同时，提高了检测准确度。

4　结束语

本文提出一种基于注意力机制和多空间金字塔池化的实时目标检测算法。该算法采用多空间金字塔池化模块，提取多尺度信息，融合多重感受野，加强特征融合网络的浅、中、深层特征图的表征能力。引入压缩激励通道注意力机制，建模通道间的相关性，调整特征图各个通道的权重，使网络更加关注重要特征，提升模型鲁棒性。特征融合和检测头网络中使用深度可分离卷积，减少网络参数量，提高模型检测效率。实验结果表明，所提算法的均值平均精度均优于

其他 7 种主流对比算法；与基准算法相比，该算法在降低复杂度的同时，提高了检测准确度。且算法的检测速率达到 33.70 帧/秒，满足实时性要求。之后，本研究将考虑进一步提升算法的准确性，尝试加入不同的特征融合网络结构以及更优的损失函数，在保证算法检测速度不降低的前提下，进一步提高检测精度，并且在工程中实现运用。

参考文献：

[1] XIAOYAN W, PEI C, RUI L, et al. Research on intelligent analysis technology of power monitoring video data based on convolutional neural network [C] //2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE). IEEE, 2020: 2461 - 2464.

[2] AHMED S A, DOGRA D P, KAR S, et al. Query-based video synopsis for intelligent traffic monitoring applications [J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 21 (8): 3457 - 3468.

[3] TIAN Y, SUN T, WANG X, et al. Design and implementation of urban rail transit train passenger flow intelligent monitoring system based on CNN [C] //2021 International Conference on Intelligent Computing, Automation and Systems (ICICAS). IEEE, 2021: 328 - 331.

[4] ADAM R, JANCIAUSKAS P, EBEL T, et al. Synthetic training data generation and domain randomization for object detection in the formula student driverless framework [C] //2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME). IEEE, 2022: 1 - 6.

[5] 王 林，张文卓. 一种融合注意力机制与上下文信息的交通标志检测方法 [J]. 计算机测量与控制，2022，30 (3)：54 - 59.

[6] CHEN C, LIU B, WAN S, et al. An edge traffic flow detection scheme based on deep learning in an intelligent transportation system [J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 22 (3): 1840 - 1852.

[7] LI Y, ZHANG L, LV Z, et al. Detecting anomalies in intelligent vehicle charging and station power supply systems with multi-head attention models [J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 22 (1): 555 - 564.

[8] RANJAN R, PATEL V M, CHELLAPPA R. Hyperface: a deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 41 (1): 121 - 135.

[9] AWAIS M, CHEN C, LONG X, et al. Novel framework: face feature selection algorithm for neonatal facial and related attributes recognition [J]. IEEE Access, 2020, 8: 59100 - 59113.

[10] CHENG G, SI Y, HONG H, et al. Cross-scale feature fusion for object detection in optical remote sensing images [J]. IEEE Geoscience and Remote Sensing Letters, 2020, 18 (3): 431 - 435.

[11] 商俊燕. 基于深度学习的遥感图像微小目标检测方法研究 [J]. 计算机测量与控制，2022，30 (10)：57 - 62.

[12] 张云飞. 基于深度学习的遥感影像目标检测系统设计 [J]. 计算机测量与控制，2021，29 (10)：77 - 82.

[13] WANG Y, WANG N, XU M, et al. Deeply-supervised networks with threshold loss for cancer detection in automated breast ultrasound [J]. IEEE Transactions on Medical Imaging, 2019, 39 (4): 866 - 876.

[14] ZHOU K, LI J, LUO W, et al. Proxy-bridged image reconstruction network for anomaly detection in medical images [J]. IEEE Transactions on Medical Imaging, 2021, 41 (3): 582 - 594.

[15] NITKUNANANTHARAJAH S, ZAHND G, OLIVO M, et al. Skin surface detection in 3D optoacoustic mesoscopy based on dynamic programming [J]. IEEE Transactions on Medical Imaging, 2019, 39 (2): 458 - 467.

[16] CORTES C, VAPNIK V. Support-vector networks [J]. Machine Learning, 1995, 20: 273 - 297.

[17] FELZENSZWALB P F, GIRSHICK R B, MCALLESTER D, et al. Object detection with discriminatively trained part-based models [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 32 (9): 1627 - 1645.

[18] FREUND Y, SCHAPIRE R E. A decision-theoretic generalization of on-line learning and an application to boosting [J]. Journal of Computer and System Sciences, 1997, 55 (1): 119 - 139.

[19] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580 - 587.

[20] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector [C] //Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21 - 37.

[21] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781 - 10790.

[22] GIRSHICK R. Fast R-CNN [C] //Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440 - 1448.

[23] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39 (6): 1137 - 1149.

[24] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection [C] //IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020, 57 (5): 9 - 12.

[25] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 8759 - 8768.