第 60 卷 第 2 期/2023 年 1 月/激光与光电子学进展 研究论文

激光与光电子学进展 先进成像

基于改进CenterNet的水下目标检测算法

王蓉蓉[1]，蒋中云[2*]

[1]上海海洋大学信息学院，上海 201306；

[2]上海建桥学院信息技术学院，上海 201306

摘要 针对常规目标检测器检测水下目标时存在特征提取困难、目标漏检等问题，提出一种改进CenterNet的水下目标检测算法。首先，使用高分辨率人体姿态估计网络HRNet代替CenterNet模型中的Hourglass-104骨干网络，降低模型参数量，提升网络推理速度；其次，引入瓶颈注意力模块，在空间维度及通道维度进行特征增强，使网络关注重要目标特征信息，提高检测精度；最后，构建特征融合模块，融合网络内部丰富的语义信息和空间位置信息，并利用感受野模块增强融合后的特征，提高网络多尺度目标检测能力。在URPU水下目标检测数据集上进行实验，与CenterNet相比，所提算法的检测精度可达77.4%，提升1.5个百分点，检测速度为7 frame/s，提升35.6%，参数量为30.4 MB，压缩84.1%，同时与其他主流目标检测算法相比具有更高的检测精度，在水下目标检测任务上更具优势。

关键词 机器视觉；水下目标检测；CenterNet；高分辨率网络；注意力机制；特征融合

中图分类号 TP391 **文献标志码** A **doi**：10.3788/LOP212230

Underwater Object Detection Algorithm Based on Improved CenterNet

Wang Rongrong[1], Jiang Zhongyun[2*]

[1]*College of Information, Shanghai Ocean University, Shanghai* 201306, *China*;

[2]*College of Information Technology, Shanghai Jian Qiao University, Shanghai* 201306, *China*

Abstract Aiming at the problems of conventional detectors in detecting underwater objects, such as difficulty in feature extraction and missing detection of objects, an improved CenterNet underwater object detection method is proposed. First, a high resolution human posture estimation network HRNet is used to replace the Hourglass-104 backbone network in CenterNet model to reduce the amount of parameters and improve the speed of network reasoning; then, the bottleneck attention module is introduced to enhance the features in the spatial and channel dimensions, and improve the detection accuracy; finally, a feature fusion module is constructed to integrate the rich semantic information and spatial location information in the network, the fused features are processed by receptive field block to further improve the multi-scale object detection ability of the network. A comparison experiment is carried out on the URPU underwater object detection dataset. Compared with CenterNet network, the detection accuracy of the proposed algorithm can reach 77.4%, increased by 1.5 percentage points, the detection speed is 7 frame/s, increased by 35.6%, the amount of parameters is 30.4 MB, compressed by 84.1%. Compared with the mainstream object detection algorithm, this algorithm also has higher detection accuracy, which has higher advantages in underwater object detection.

Key words machine vision; underwater object detection; CenterNet; high resolution network; attention mechanism; feature fusion

1 引 言

水下目标检测是对水下场景中的目标进行精确定位与识别的技术，该技术因在海洋学、水下导航、渔业养殖等领域的广泛应用，不断受到人们的关注[1]。然而水下环境复杂，并且受水下特殊成像环境的限制，水下光学图像往往存在噪声干扰多、对比度低、纹理特征模糊等问题，水下目标检测依旧面临着巨大的挑战[2]。

收稿日期：2021-08-12；**修回日期**：2021-10-07；**录用日期**：2021-11-08；**网络首发日期**：2021-11-18

基金项目：上海市属高校应用型本科试点专业建设项目(Z32004-17084)、上海市教育委员会一流本科专业建设专项项目(JYLB202002)

通信作者：*jianqiao_jzy@163.com

早期传统水下目标检测算法，首先利用滑动窗口提取水下图像中目标的候选区域，之后人工提取候选区域的目标特征（如HOG、SIFT、Haar），最后利用分类器实现目标的识别。例如：文献[3]利用Haar-like特征和多个级联的分类器实现鱼类目标的检测；文献[4]根据水下图像颜色的均匀性和轮廓的锐度信息检测水下目标。但传统方法主要提取图像底层和中间层次特征，很难捕捉到具有代表性的语义信息，因此算法鲁棒性差，且基于滑动窗口提取候选区域的方式导致算法整体复杂度高、效率低下。随着深度学习的发展，基于卷积神经网络的目标检测算法在各项检测任务上取得先进的结果，主要包括two-stage目标检测算法和one-stage目标检测算法：two-stage目标检测算法将检测任务分两阶段执行，虽然拥有较高的检测精度但速度较慢，代表算法如Faster R-CNN[5]系列；one-stage目标检测算法将目标检测任务看作单一回归问题，精度略低于two-stage算法，但拥有更快的检测速度，代表算法如YOLO[6]系列、SSD[7]系列。目前，利用深度学习方法检测水下目标得到了广泛关注。例如：文献[8]设计了一种深度可分离变形卷积改进SSD模型，提升模型在复杂水下环境下的目标检测能力；文献[9]精简YOLO结构，利用迁移方法训练网络，克服样本集的限制，实现水下鱼类目标的实时、准确检测；文献[10]利用多尺度训练、测试和细粒度特征结合等方法改进YOLOv3网络，实现声呐图像目标的检测；文献[11]通过引入生成式对抗网络生成特定类目标的样本，缓解YOLO网络由于目标数量不平衡引起的检测精度降低的问题，实现水下目标的快速、精确检测。

基于深度学习的水下目标检测器表现出良好的性能，但锚框是检测器常用的组件，锚框的使用使网络极易产生不均衡数量的正、负样本，降低检测精度。因此，本文将无锚框目标检测器CenterNet[12]应用到水下目标检测任务。受特殊成像原理的限制，水下图像存在纹理特征信息模糊、对比度低及颜色失真等情况[13]，且水下目标尺度差异较大，对CenterNet进行以下改进以提升其水下目标的检测性能：首先，引入高分辨率网络HRNet[14]作为骨干网络，使网络保持较强特征提取能力的同时，降低模型参数量，提升检测速度；然后，融入瓶颈注意力模块（BAM）[15]，在空间及通道维度增强目标特征，提升网络对于水下模糊目标的特征提取能力；最后，构建特征融合模块（FFM），融合多分辨率特征图，并通过感受野模块（RFB）[16]进一步增强融合特征的鲁棒性和辨别性，提升模型对于水下多尺度目标的检测能力。

2　基于改进CenterNet的水下目标检测算法

通过改进骨干网络、引入BAM、构建FFM等3个方面改进原始CenterNet，以提升水下目标检测的精度，同时提高模型推理速度。改进后的模型简略结构如图1所示，由FA_HRNet骨干网络和检测模块两部分构成。输入图像首先经过HRNet的4个并行子网络，快速提取4种不同分辨率的特征；之后，4种不同分辨率的特征传入BAM，在空间及通道维度进行特征增强，抑制背景无用信息，增强有效目标信息；然后，经过BAM增强后的4种分辨率特征输入FFM进行特征融合，输出包含丰富语义信息与位置信息的融合特征，提升网络对于多尺度目标的鲁棒性；最后，融合特征传入检测模块通过3个独立的分支生成关键点热图、位置偏移与目标宽高，并通过解码操作得到最终检测结果。

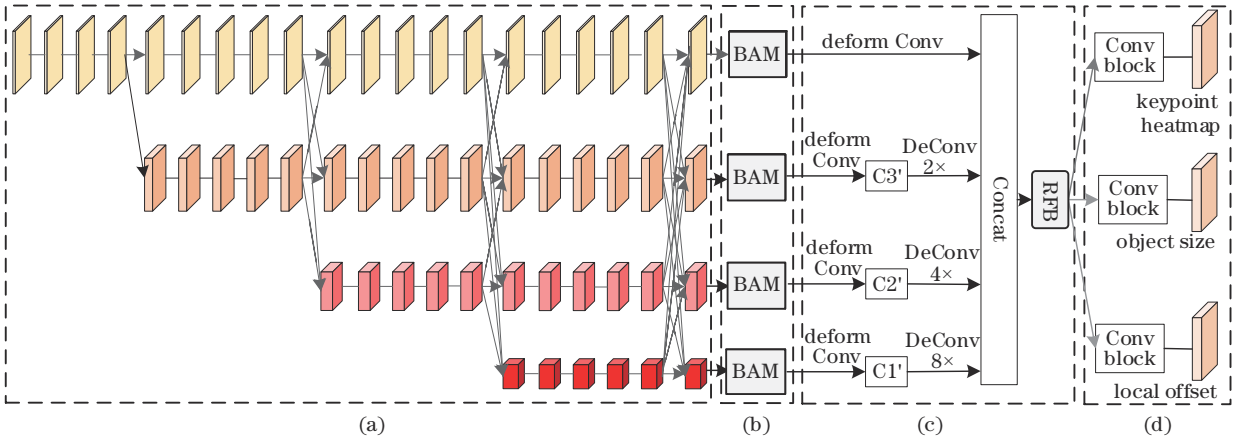

图1　模型结构。(a) HRNet；(b) BAM；(c) FFM；(d)检测模块

Fig. 1　Model structure. (a) HRNet; (b) BAM; (c) FFM; (d) detection module

2.1　骨干网络改进

CenterNet采用Hourglass-104[17]作为骨干网络，虽然可以实现较高的检测精度，但该网络结构复杂，参数量巨大，大大增加了网络的计算复杂度，降低了网络的检测速度。因此为提高算法的检测速度，同时维持较高的检测精度，从轻量化网络结构的角度出发，将骨干网络改为HRNet。

HRNet是一种高分辨率人体姿态估计网络，在整

个网络中保持高分辨率的特征，而不是从低分辨率特征恢复到高分辨率特征，因此可以保留更准确的目标空间信息，并且网络内部通过执行多尺度融合来增强高分辨率特征，在不使用中间监督的情况下，具有较好的计算复杂度和较高的参数效率，网络结构如图 2 所示。

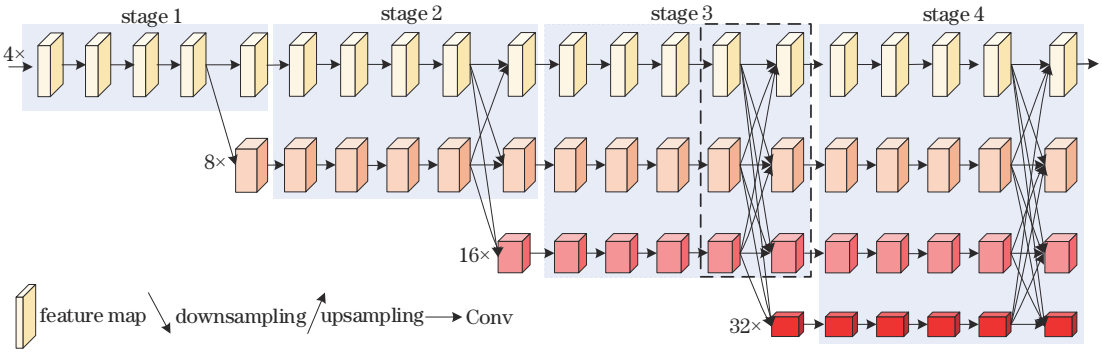

图 2　HRNet结构

Fig. 2　HRNet structure

HRNet包含4个并行的子网络，每个子网络由一系列标准卷积组成，同一子网络特征图分辨率不随网络深度的变化而改变，而并行子网络的特征图分辨率依次降低1/2，同时通道数量增大2倍。HRNet内部通过引入跨并行子网络的交换单元，使每个子网络重复地从其他并行子网络接收信息，聚合不同分辨率的特征得到增强的高分辨率特征，该特征在空间位置上更加准确。交换单元聚合不同分辨率特征的具体实现如图 2 虚线框部分所示，其输入是 s 个响应图 $\{X_1, X_2, \cdots, X_s\}$，输出是输入响应图的聚合，即 $Y_k = \sum_{i=1}^{s} a(X_i, k)$，$s$ 个输入响应图产生 s 个输出 $\{Y_1, Y_2, \cdots, Y_s\}$，每个跨阶段的交换单元还包含1个额外输出 $Y_{s+1} = a(Y_s, s+1)$，其中 $a(X_i, k)$ 函数表示通过下采样或者上采样操作将输入 X_i 的分辨率由 i 更改为 k，下采样使用步长为2的3×3标准卷积实现，上采样使用最邻近差值法，后跟1×1标准卷积匹配通道数，如果 $i=k$，则 $a(X_i, k)$ 为一个恒等映射，即 $a(X_i, k) = X_i$。

所用骨干网络HRNet的详细结构参数如表 1 所示，主体为4个不同分辨率的并行子网络，包含4个阶段：第1阶段包含4个宽度为64的瓶颈残差单元，后跟一个3×3卷积将特征图通道数调整为32，以降低网络内部参数量；第2、3、4阶段分别包含1、4、3个交换块，其中每个交换块包含4个残差单元，在每个分辨率中都包含2个3×3卷积和1个跨分辨率的交换单元（多尺度特征融合）。HRNet的4个子网络分别对图像进行4、8、16、32倍下采样，提取不同分辨率的特征图，原始HRNet仅利用高分辨率子网络的输出，为充分利用多尺度特征信息，本实验组将4个并行子网络产生的特征图作为骨干网络的输出。

表 1　骨干网络结构参数

Table 1　Backbone network structure parameters

Net	Stage 1	Stage 2	Stage 3	Stage 4	Resolution
Subnet_1	$\begin{bmatrix}1\times1,64\\3\times3,64\\1\times1,256\end{bmatrix}\times4\times1$	$\begin{bmatrix}3\times3,32\\3\times3,32\end{bmatrix}\times4\times1$	$\begin{bmatrix}3\times3,32\\3\times3,32\end{bmatrix}\times4\times4$	$\begin{bmatrix}3\times3,32\\3\times3,32\end{bmatrix}\times4\times3$	4×
Subnet_2		$\begin{bmatrix}3\times3,64\\3\times3,64\end{bmatrix}\times4\times1$	$\begin{bmatrix}3\times3,64\\3\times3,64\end{bmatrix}\times4\times4$	$\begin{bmatrix}3\times3,64\\3\times3,64\end{bmatrix}\times4\times3$	8×
Subnet_3			$\begin{bmatrix}3\times3,128\\3\times3,128\end{bmatrix}\times4\times4$	$\begin{bmatrix}3\times3,128\\3\times3,128\end{bmatrix}\times4\times3$	16×
Subnet_4				$\begin{bmatrix}3\times3,256\\3\times3,256\end{bmatrix}\times4\times3$	32×

2.2　引入瓶颈注意力模块

HRNet可产生4种不同分辨率的特征图，这些特征图包含了目标的有效特征，同时也包含了大量无效背景特征，并且这4种特征图存在差异，对最终检测结果的贡献也存在差异。因此为抑制无效特征、增强目标特征，同时使网络自主学习不同分辨率特征图之间的关联性和重要程度，引入BAM，通过两个不同的分支在空间及通道维度对特征进行增强，其结构如图 3 所示。

通道注意力分支通过建模通道间的相关性，使网络关注感兴趣的通道特征。首先输入特征 $\boldsymbol{F} \in \mathbf{R}^{C\times H\times W}$ 经过全局平均池化，编码每个通道的全局

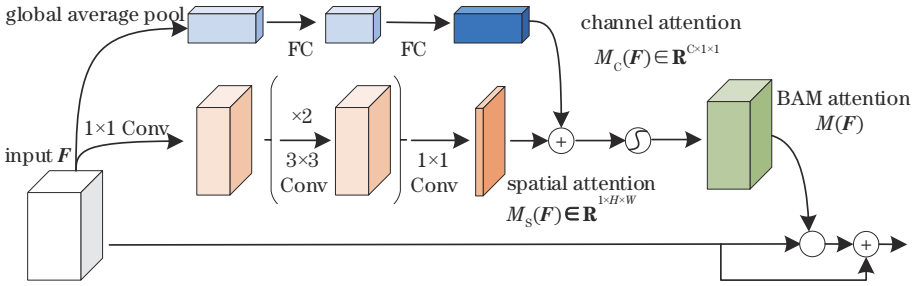

图 3　BAM 结构

Fig. 3　BAM structure

信息，生成一维通道向量；然后利用多层感知机（MLP）处理一维通道向量，估计通道间的注意力；最后利用批量归一化（BN）层调整输出特征尺度，得到通道注意力映射 $M_C(\boldsymbol{F})\in\mathbf{R}^C$。具体可描述为

$$M_C(\boldsymbol{F})=\mathrm{BN}\left\{\mathrm{MLP}\left[\mathrm{AvgPool}(\boldsymbol{F})\right]\right\}=\mathrm{BN}\left\{\boldsymbol{W}_1\left[\boldsymbol{W}_0\mathrm{AvgPool}(\boldsymbol{F})+\boldsymbol{b}_0\right]+\boldsymbol{b}_1\right\},\tag{1}$$

式中：$\boldsymbol{W}_0\in\mathbf{R}^{C/r\times C}$；$\boldsymbol{b}_0\in\mathbf{R}^{C/r}$；$\boldsymbol{W}_1\in\mathbf{R}^{C\times C/r}$；$\boldsymbol{b}_0\in\mathbf{R}^C$。

空间注意力分支可有效捕捉特征空间位置信息，使网络更加关注目标的位置信息。首先将输入 $\boldsymbol{F}\in\mathbf{R}^{C\times H\times W}$ 利用 1×1 卷积压缩通道维度；之后使用 2 个 3×3 空洞卷积聚合具有更大感受野的上下文信息；最后利用 1×1 卷积将特征图维度映射为 $\mathbf{R}^{1\times H\times W}$，并利用批量归一化层进行尺度调整，得到空间注意力映射 $M_S(\boldsymbol{F})\in\mathbf{R}^{H\times W}$。具体可描述为

$$M_S(\boldsymbol{F})=\mathrm{BN}\left\{f_3^{1\times1}\left\{f_2^{3\times3}\left\{f_1^{3\times3}\left[f_0^{1\times1}(\boldsymbol{F})\right]\right\}\right\}\right\},\tag{2}$$

式中：f 为卷积运算，上标为卷积核大小，下标为卷积运算的次序。

BAM 细化输入特征 $\boldsymbol{F}\in\mathbf{R}^{C\times H\times W}$ 的完整计算方式可表示为

$$\boldsymbol{F}'=\boldsymbol{F}+\boldsymbol{F}\otimes\sigma\left[M_C(\boldsymbol{F})+M_S(\boldsymbol{F})\right],\tag{3}$$

式中：$\otimes$ 代表逐元素相乘；σ 代表 Sigmoid 激活函数；$M_C(\boldsymbol{F})$、$M_S(\boldsymbol{F})$ 分别为通道注意力映射与空间注意力映射，两者在相加之前大小被调整为 $\mathbf{R}^{C\times H\times W}$。

一般网络通常串行叠加注意力机制，即在大部分卷积层后添加注意力机制。由于 HRNet 结构的特殊性，仅在并行子网络最后的输出部分并行添加 BAM 注意力机制模块，该模块通过对子网络的输出特征在空间及通道维度进行增强，有效过滤无效背景特征、增强有效目标特征，显著提升各个子网络输出特征的质量，并且不会过多增加网络参数。

2.3　构建特征融合模块

HRNet 的输出经过 BAM 处理后可得到 4 种分辨率的增强特征：高分辨率特征经过较少卷积层处理，语义信息匮乏，但包含丰富的位置及细节信息，适用于小目标的检测；低分辨率特征经过多次卷积下采样，空间位置信息严重流失，但包含丰富的语义信息，适用于大型目标的检测[18]。借鉴 FSSD[19] 的思想，设计了 FFM 聚合多种分辨率特征图，提高网络对于水下多尺度目标的检测能力，其结构如图 4 所示。

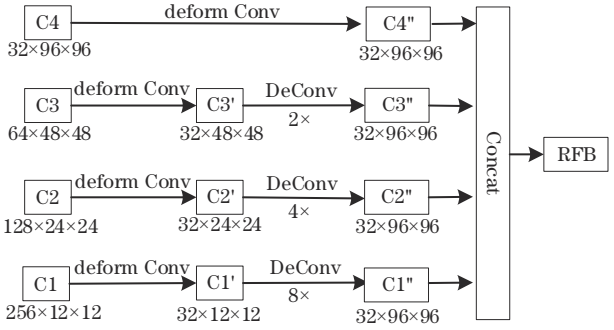

图 4　特征融合模块

Fig. 4　Feature fusion model

FFM 实现特征融合的具体过程如下：首先，利用 3×3 可变形卷积（deformable convolution）[20] 对每一源特征图进行通道降维，使通道数保持一致，同时减少网络内部计算量。选用可变形卷积的原因在于其相对于标准卷积只需额外学习一个偏移量，便可根据实际情况变换自身卷积核的形状，调整感受野大小，仅增加少量参数便可显著提高网络对目标尺度变化的鲁棒性。之后，利用反卷积层对低分辨率特征进行上采样，使其分辨率与高分辨率特征图保持一致。常用的上采样方法有反卷积层和双线性差值法（bilinear interpolation），由于反卷积可以为网络提供可供学习的参数，提升网络的性能，本实验组选择反卷积进行上采样。最后，4 种调整后的特征图经过 Concat 融合操作后输入 RFB 模块，该模块通过模拟卷积核大小和离心率的关系进一步增强目标特征的可辨别性和鲁棒性。RFB 是一个多分支卷积块，如图 5 所示：每个分支首先利用 1×1 卷积进行通道降维和跨通道信息融合；之后紧跟 3×3、1×3 和 3×1 卷积实现多个尺度的感受野，利于检测多种尺度目标，同时保持较低计算量；最后引入空洞卷积作用于每个分支，在保持相同数量参数的同时，在具有更多上下文的更大区域捕获信息，生成更高分辨率特征图，并利用 Concat 实现多分支、多尺度特征融合。

FFM 聚合 4 种分辨率的输入特征图，产生包含丰富语义信息与空间位置信息的增强特征，可显著提升

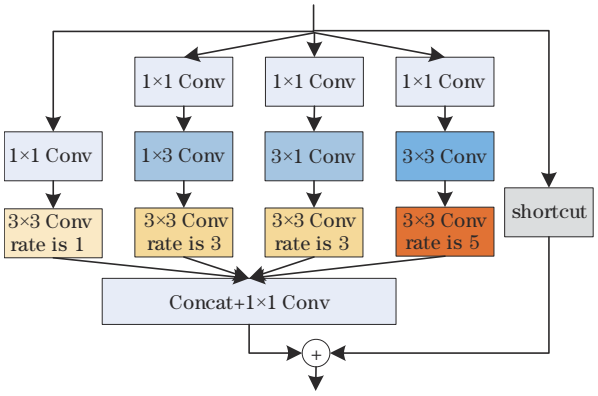

图 5　RFB 模块

Fig. 5　RFB model

网络多尺度目标检测能力，提升模型目标检测的准确度。

2.4　损失函数设计

所提算法的训练损失由中心点预测损失、目标宽高损失及位置偏移损失等 3 部分构成。中心点预测损失采用根据 CenterNet 算法改进的 Focal 损失，用于缓解网络内因正负样本数量分布不均而导致的检测精度下降的问题，其表达式为

$$L_{\mathrm{k}}=\frac{-1}{N}\sum_{xyc}\begin{cases}(1-\hat{Y}_{xyc})^{\alpha}\log(\hat{Y}_{xyc}), & Y_{xyc}=1\\(1-\hat{Y}_{xyc})^{\beta}(\hat{Y}_{xyc})^{\alpha}\log(1-\hat{Y}_{xyc}), & Y_{xyc}\neq 1\end{cases},\tag{4}$$

式中：N 为图像中中心点的数量；$\hat{Y}_{xyc}$ 和 Y_{xyc} 分别为网络预测中心点和 ground truth（GT）中心点；α 和 β 为 Focal 损失的超参数，用于平衡正负样本的重要程度，实验中 α 和 β 的值分别设置为 2 和 4。

位置偏移损失采用 L1 损失，用于修正坐标位置的偏差，其表达式为

$$L_{\mathrm{off}}=\frac{1}{N}\sum_{\tilde{p}}\left|\hat{O}_{\tilde{p}}-\left(\frac{p}{R}-\tilde{p}\right)\right|,\tag{5}$$

式中：p 和 $\tilde{p}$ 分别代表 GT 中心点和预测中心点；R 代表图像下采样率；$\hat{O}_{\tilde{p}}$ 为预测中心点的偏移量；$\frac{p}{R}-\tilde{p}$ 为中心点的实际偏移量。

目标宽高损失，同样采用 L1 损失，其表达式为

$$L_{\mathrm{size}}=\frac{1}{N}\sum_{k=1}^{N}\left|\hat{S}_{pk}-S_k\right|,\tag{6}$$

式中：S_k 和 $\hat{S}_{pk}$ 分别为 GT 宽高和网络预测的目标宽高。

总损失为上述损失的加权和，其表达式为

$$L_{\mathrm{det}}=L_{\mathrm{k}}+\lambda_{\mathrm{size}}L_{\mathrm{size}}+\lambda_{\mathrm{off}}L_{\mathrm{off}},\tag{7}$$

式中：目标宽高损失的权重 $\lambda_{\mathrm{size}}=0.1$；中心点偏移损失的权重 $\lambda_{\mathrm{off}}=1$。

3　实验与分析

3.1　数据集

实验采用 2018 年水下机器人目标抓取大赛（URPU）官方数据集，包含 2901 张训练图像和 800 张测试图像，包含扇贝（scallop）、海参（holothurian）、海星（starfish）、海胆（echinus）等 4 种目标类别，不同类别目标的示例图像如图 6 所示。图像在真实海洋环境中拍摄，存在对比度低、色彩失真、特征信息模糊等问题，并且目标密集、存在遮挡，给目标检测带来极大挑战。

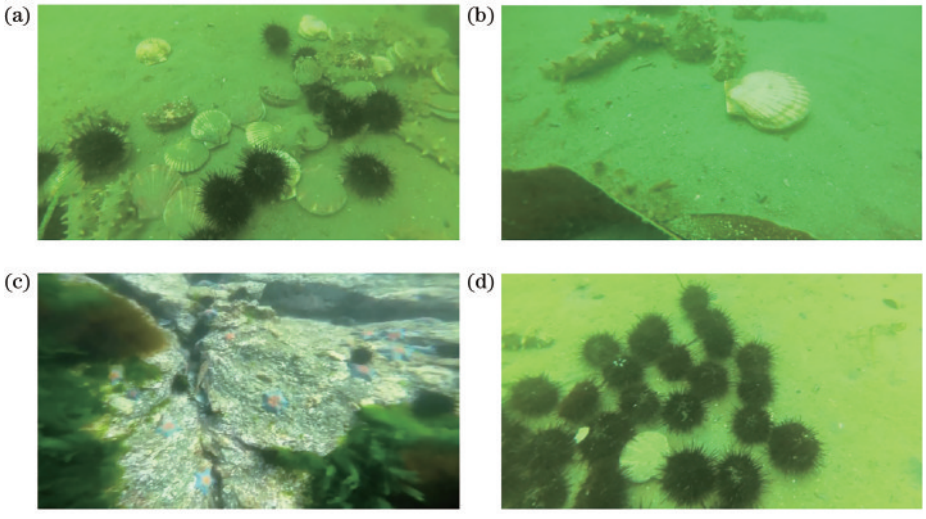

图 6　示例图像。(a) 扇贝；(b) 海参；(c) 海星；(d) 海胆

Fig. 6　Example images. (a) Scallop; (b) holothurian; (c) starfish; (d) echinus

训练集与测试集中各类别目标的样本数量如图 7 所示。在训练集中，扇贝、海参、海星、海胆等 4 类目标分别约占数据集的 6.5%、15.7%、17.9%、60.0%，不同类目标数量分布极其不均衡，给网络训练带来极大的挑战。

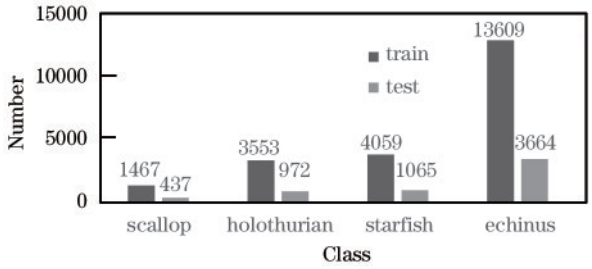

图 7 样本分布

Fig. 7 Sample distribution

3.2 实验环境与设置

训练网络采用的工作站系统为 Ubuntu 18.04，配置 NVIDIA GeForce RTX 2080 SUPER 显卡，使用 PyTorch 深度学习框架，在 Python 3.6、CUDA 11.0、CuDNN 11.0 环境下搭建并训练目标检测模型。

训练时，统一对输入图像进行仿射变换，使其分辨率统一为 384×384；数据集均值设为[0.246859, 0.567928, 0.325002]，标准差设为[0.048203, 0.105985, 0.075544]；初始学习率设置为 1.25×10^{-4}，并在训练 45、60 epoch 时下降为原来的 1/10，采用 Adam 优化器优化整体目标，在 70 epoch 时结束训练。由于训练样本数量较少，易出现过拟合问题，因此训练网络时采用数据增强的策略，包括随机裁剪、随机翻转、缩放和平移、调整对比度、饱和度与亮度等，降低网络对于目标位置的敏感性及对某些特征的依赖性，增强网络的泛化能力。

3.3 评价指标

实验采用每秒处理帧数（FPS）、平均精度（AP）和均值平均精度（mAP）等 3 个指标客观评定网络的性能。利用 FPS 指标评估网络的检测速度，数值越大表明网络检测速度越快。利用 AP 指标评估单个类别的检测精度，其表达式为

$$P_{\mathrm{AP}}=\int_0^1 P(R)\mathrm{d}R, \tag{8}$$

式中：P 代表精确率，指所有检测到的目标为正确检测的概率；R 代表召回率，指所有真实正样本被正确检测的概率。

$$\begin{cases} P=\dfrac{N_{\mathrm{TP}}}{N_{\mathrm{TP}}+N_{\mathrm{FP}}} \\ R=\dfrac{N_{\mathrm{TP}}}{N_{\mathrm{TP}}+N_{\mathrm{FN}}} \end{cases}, \tag{9}$$

式中：N_{TP} 代表检测正确的正样本数量；N_{FP} 代表检测为正样本但实际为负样本的数量；N_{FN} 代表检测为负样本但实际为正样本的数量。

利用 mAP 评价模型的综合检测精度，其表达式为

$$P_{\mathrm{mAP}}=\sum_{i=1}^{N}P_{\mathrm{AP}_i}/N, \tag{10}$$

式中：N 代表所有类别的数量；P_{AP_i} 代表类别 i 的平均精度。

3.4 算法有效性验证

为验证所提算法的合理性，在 PASCAL VOC 公共数据集上将其与主流目标检测算法进行了对比。实验训练集为 VOC07+ VOC12 trainval，共计 16551 张图像，测试集为 VOC07 test-dev，共计 4952 张图像，包含 20 个目标类别。实验结果如表 2 所示。

表 2 PASCAL VOC 数据集测试结果

Table 2 PASCAL VOC dataset test results

Algorithm	Bacbone nework	Size	GPU	mAP /%	FPS
Fast R-CNN	VGG-16	~1000×600	Tian X	70.0	0.5
Reference [21]	ResNet-50		RTX 2080Ti	72.6	
Faster R-CNN	ResNet-101	~1000×600	Tian X	76.4	5.0
SSD300	VGG-16	300×300	Tian X	77.1	45
SSD	VGG-16	320×320	Tian X	77.5	11.2
YOLOv3	Darknet-53	554×554	Tian X	79.3	26.0
RetinaNet	ResNet-101	~1000×600	RTX2080 S	75.3	8.8
Proposed algorithm	FA-HRNet	384×384	RTX2080 S	78.1	7.0
	FA-HRNet	512×512	RTX2080 S	79.5	6.7

从表 2 可以看出：当输入尺寸为 384×384 时，所提算法的平均检测精度达到 78.1%，同 Fast R-CNN、Faster R-CNN 等两阶段目标检测算法相比，检测精度更高、速度更快；输入尺寸增加到 512×512 时，平均检测精度达到 79.5%，相较当前主流的单阶段目标检测算法 SSD、YOLOv3、RetinaNet，检测速度降低，但检测精度更高。在 PASCAL VOC 公开数据集上的实验结果表明，所提算法在通用物体的检测上具备优势，验证了网络结构设计的合理性。

3.5 实验结果与分析

3.5.1 所提算法与 CenterNet 算法检测性能对比

为评估所提改进算法的有效性，从检测精度、检测速度和模型复杂度等 3 个方面与 CenterNet 算法进行了对比分析，实验结果如表 3 和表 4 所示。

从表 3 可以看出：与 CenterNet 算法相比，从检测精度上来看，所提算法对于海星类别的检测精度略微降低，但是对海参、海胆、扇贝类别的检测精度分别提升 1.5 个百分点、1.2 个百分点、2.8 个百分点，最终

表 3　与 CenterNet 算法检测精度、速度对比

Table 3　Comparison of detection accuracy and speed with CenterNet algorithm

Algorithm	Net	Scallop /%	Holothurian /%	Starfish /%	Echinus /%	mAP /%	FPS
CenterNet	Hourglass-104	57.4	71.8	86.0	88.5	75.9	5.2
Proposed algorithm	FA-HRNet	60.2	73.3	85.6	89.7	77.4	7.0

表 4　与 CenterNet 算法模型复杂度对比

Table 4　Comparison of model complexity with CenterNet algorithm

Algorithm	Net	Ipout size	Size /MB	Params /MB	GFLOPs /10^9
CenterNet	Hourglass-104	384×384	765.7	191.2	164.5
Proposed algorithm	FA-HRNet	384×384	123.0	30.4	32.6

mAP 提升 1.5 个百分点，表明所提算法有更强的特征提取能力和多尺度目标检测能力，可以更准确检测各类水下目标；从检测速度上来看，所提算法的检测速度是 CenterNet 的 1.35 倍，处于明显的优势，表明 FA_HRNet 相比 Hourglass-104 是更加轻量级的网络，可显著提升模型的推理速度。

从表 4 可以看出，在相同的实验设置情况下，所提算法模型大小、模型参数量、浮点运算量（GFLOPs）远低于 CenterNet 算法，模型大小降低 642.7 MB，压缩 83.9%，模型参数量降低 160.8 MB，压缩 84.1%，浮点运算量降低 1.319×10^{11}，缩减 80.2%。这表明所提改进算法可显著降低模型的复杂度，在实际应用时更具优势。

图 8 为两种算法在测试集上对 4 类目标的检测效果，所提算法和 CenterNet 算法能检测到大部分目标，但所提算法降低了对模糊、遮挡目标的漏检率，能够检测到更多目标，同时提升了对各类目标的检测置信度，产生位置偏差更小的检测框。

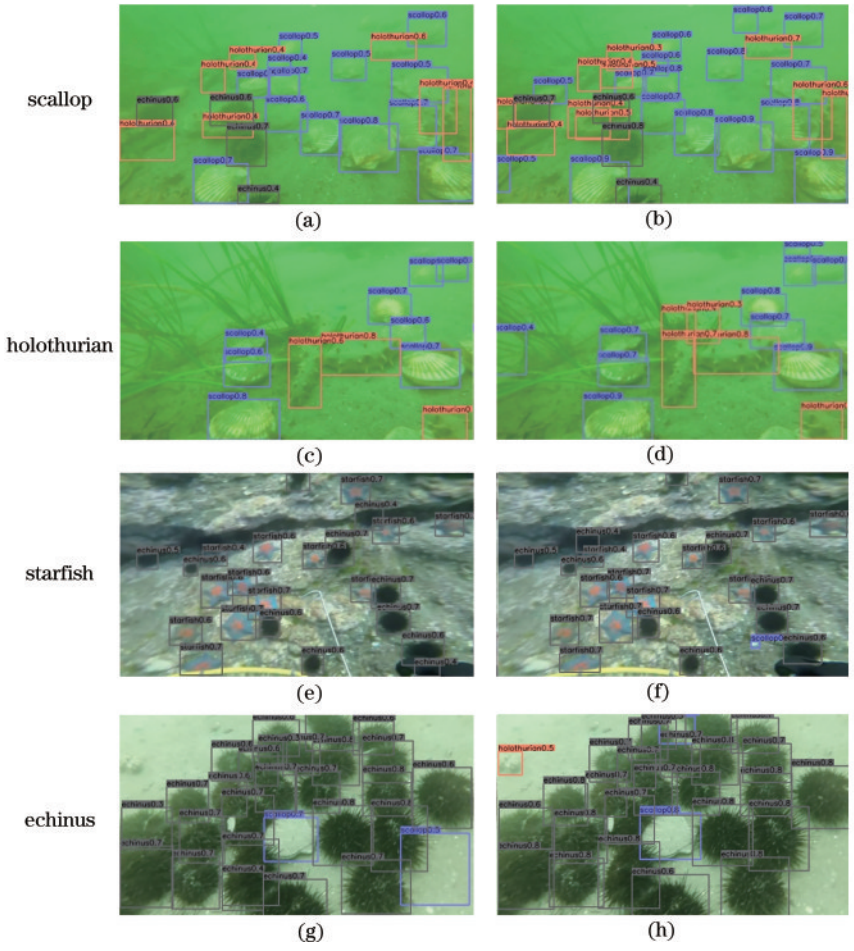

图 8　不同网络检测结果对比。(a) (c) (e) (g) CenterNet 算法；(b) (d) (f) (h) 所提算法

Fig. 8　Comparison of detection results of different networks. (a) (c) (e) (g) CenterNet algorithm; (b) (d) (f) (h) proposed algorithm

综合考虑检测精度、检测速度、模型复杂度等3个方面，所提算法在检测水下目标时比CenterNet更具优势，验证了所提改进策略的有效性。

3.5.2 所提算法与主流目标检测算法的性能对比

为验证所提算法的优越性，将所提算法与Faster R-CNN、文献[22]中的算法、SSD、RetinaNet[23]、YOLOv3、CornerNet[24]、ExtremeNet[25]、CenterNet等8种目标检测算法进行对比，主要从检测精度、检测速度和模型复杂度等方面进行对比分析，结果如表5所示。

表5 与主流目标检测算法性能对比

Table 5 Performance comparison with mainstream object detection algorithms

Algorithm	Bacbone nework	Size /MB	Params /MB	GFLOPs /10^9	mAP /%	FPS
Faster R-CNN	ResNet-101+FPN	552.8	59.5	91.8	73.7	3.5
Reference [22]	ResNet-50				72.6	
SSD	VGG-16	92.61	24.2	30.6	68.6	16.0
YOLOv3	Darknet-53	246.5	61.5	32.8	73.3	15.0
RetinaNet	ResNet-101	228.5	55.2	100.6	72.2	8.8
CornerNet	Hourglass-104	804.6	201.0	453.0	49.0	3.1
ExtremeNet	Hourglass-104	794.1	198.3	229.9	53.5	2.3
CenterNet	Hourglass-104	765.7	191.2	164.5	75.9	5.2
Proposed algorithm	FA-HRNet	123.0	30.4	32.6	77.4	7.0

从表5可以看出，从检测精度上来看，所提算法mAP可达77.4%，远高于其他主流目标检测算法，表明所提改进算法在检测水下目标时更具优势，能够实现较高的检测精度。值得注意的是，无锚框检测器ExtremeNet、CornerNet的各项检测精度显著低于所提算法及其他算法，原因在于上述算法需要定义多个关键角点，并根据关键点之间的距离进行分组，由于水下目标存在严重遮挡、小目标广泛，极易出现误分组显著降低了检测精度。从检测速度上来看，所提算法同Faster R-CNN、ExtremeNet、CornerNet、CenterNet相比检测速度更快，但低于SSD、YOLOv3、RetinaNet算法，原因在于所提算法的骨干网络以HRNet为基础，拥有较低参数量，提升了检测速度，但由于网络内部反复进行多尺度特征融合，一定程度上降低了网络推理速度。从模型复杂度上来看，所提算法的模型大小、参数量、浮点运算量仅为123.0 MB、30.4 MB、3.26×10^{10}，远低于多数主流目标检测模型，表明所提改进网络是一种相对轻量级的网络，可显著减少模型的参数量、降低模型复杂度、提升网络推理速度，相较于其他网络更易部署到移动与嵌入式设备，有利于实际应用。

综合分析，所提算法同其他主流算法相比达到较高的检测精度，同时保持较快的检测速度与较低的模型参数量，在检测水下目标时具有显著优势。

3.5.3 消融实验

为了验证所提改进策略(HRNet骨干网络、BAM和FFM)的有效性，进行了消融实验，从检测精度、检测速度和模型复杂度等方面进行了对比分析，结果如表6所示('√'代表使用此改进点，'—'代表不使用此改进点)，并对不同实验设置下各类别的检测精度进行可视化，结果如图9所示。实验1代表CenterNet，骨干网络为Hourglass-104，实验2在实验1的基础上将骨干网络替换为HRNet，实验3在实验2的基础上融入BAM，实验4在实验3的基础上增加FFM。

表6 不同模块对检测性能的影响

Table 6 Influence of different modules on detection performance

No.	HRNet	BAM	FFM	Size /MB	Params /MB	GFLOPs /10^9	mAP /%	FPS
1	—	—	—	765.7	191.24	164.53	75.9	5.2
2	√	—	—	115.2	28.67	24.16	74.7	8.2
3	√	√	—	115.6	28.74	24.17	76.2	7.9
4	√	√	√	123.0	30.36	32.55	77.4	7.0

实验1与实验2的结果表明，采用HRNet替代Hourglass-104作为检测器骨干网络后，模型大小、参数量、浮点计算量分别降低650.5 MB、162.57 MB、1.4037×10^{11}，模型推理速度提升3，但网络平均检测精度降低1.2个百分点，表明HRNet是一种更加轻量的网络，可显著降低模型复杂度，提升检测速度，但由于网络拥有较少的参数量，特征提取能力略微降低，检测精度略微降低。实验2与实验3的结果表明，引入BAM后，模型大小、参数量、浮点计算量分别增加0.4 MB、0.07 MB、0.01×10^9，模型推理速度略微降低，但平均检测精度提升1.5个百分点，表明BAM增加了模型复杂度，降低了网络推理速度，但该注意力机

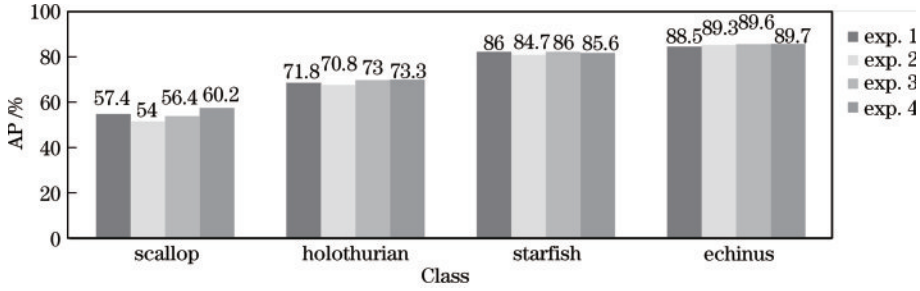

图 9　不同类别检测精度

Fig. 9　Detection accuracy of different categories

制利用通道注意力分支建模通道间的相关性，抑制了非必要特征信息，同时利用空间注意力分支有效捕捉目标空间位置信息，通过两个并行注意力分支使网络更加关注目标特征信息，提高网络提取的特征图质量，显著提升模型的检测精度。实验 4 在实验 3 的基础上，引入 FFM，平均检测精度进一步提升 1.2 个百分点，但模型大小、参数量、浮点计算量分别增加 7.4 MB、1.62 MB、8.83×10^9，模型推理速度降低 0.9，表明 FFM 充分利用了网络内部丰富的语义信息和位置信息，提升了网络对于多尺度目标的检测精度，但由于其内部引入了可变形卷积层、RFB 等结构，增加了模型复杂度。

海胆、海星两类目标数量较多、特征信息较为清晰，因而不同实验设置下的模型均易提取目标特征，实现精确检测。由于海参、扇贝这两类目标仅占总目标数量的 15.6%、6.5%，且这两类目标通常尺寸较小、特征信息模糊，因此检测精度较低，但所提算法（exp. 4）对上述类别的检测精度达到 73.3%、60.2%，显著高于其他网络，表明所提算法可显著提升网络特征提取能力，在一定程度上缓解了目标数据样本匮乏的限制，同时提升了对于小目标及模糊目标的检测能力。

综上所述：利用 HRNet 作为模型骨干网络可使网络整体的复杂度降低，推理速度显著提升，但由于特征提取能力的限制，使得网络整体检测精度略微降低；引入并行 BAM 后，网络推理速度略微降低，但可有效抑制无用信息，显著提升检测精度；FFM 增加了模型复杂度、降低模型推理速度，但同时提升模型对于多尺度目标的检测能力。

5　结　　论

针对水下图像质量低下导致目标特征信息提取困难、目标漏检等问题，提出了一种基于改进的 CenterNet 水下目标检测算法：首先，引入 HRNet 作为模型的骨干网络，降低网络参数量，提高检测速度；然后，引入 BAM 对骨干网络提取的特征在空间及通道维度进行增强，提高网络特征提取能力，提升目标检测精度；最后，构建 FFM 聚合多分辨率特征图，产生具有丰富语义性与空间性的融合特征，并利用 RFB 模块进一步增强融合特征的可辨别性和鲁棒性，提升网络对于多尺度目标的检测能力，进而提升水下目标检测的精度。实验结果表明，与 CenterNet 相比所提算法检测精度、检测速度分别提升 1.5 个百分点、35.6%，模型大小、模型参数量、浮点计算量分别压缩 83.9%、84.1%、80.2%，与其他主流目标检测算法相比在保持较快检测速度的同时，在检测精度、模型参数量上具有明显优势。即所提算法在检测水下目标时更具优势。

参　考　文　献

[1] Chen L, Liu Z H, Tong L, et al. Underwater object detection using Invert Multi-Class Adaboost with deep learning[C]//2020 International Joint Conference on Neural Networks (IJCNN), July 19-24, 2020, Glasgow, UK. New York: IEEE Press, 2020.

[2] 林森, 赵颖. 水下光学图像中目标探测关键技术研究综述[J]. 激光与光电子学进展, 2020, 57(6): 060002.
Lin S, Zhao Y. Review on key technologies of target exploration in underwater optical images[J]. Laser & Optoelectronics Progress, 2020, 57(6): 060002.

[3] Cutter G, Stierhoff K, Zeng J M. Automated detection of rockfish in unconstrained underwater videos using Haar cascades and a new image dataset: labeled fishes in the wild[C]//2015 IEEE Winter Applications and Computer Vision Workshops, January 6-9, 2015, Waikoloa, HI, USA. New York: IEEE Press, 2015: 57-62.

[4] Rizzini D L, Kallasi F, Oleari F, et al. Investigation of vision-based underwater object detection with multiple datasets[J]. International Journal of Advanced Robotic Systems, 2015, 12(6): 77.

[5] Ren S Q, He K M, 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.

[6] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.

[7] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer

science. Cham: Springer, 2016, 9905: 21-37.

[8] 强伟, 贺昱曜, 郭玉锦, 等. 基于改进SSD的水下目标检测算法研究[J]. 西北工业大学学报, 2020, 38(4): 747-754.
Qiang W, He Y Y, Guo Y J, et al. Exploring underwater target detection algorithm based on improved SSD[J]. Journal of Northwestern Polytechnical University, 2020, 38(4): 747-754.

[9] 李庆忠, 李宜兵, 牛炯. 基于改进YOLO和迁移学习的水下鱼类目标实时检测[J]. 模式识别与人工智能, 2019, 32(3): 193-203.
Li Q Z, Li Y B, Niu J. Real-time detection of underwater fish based on improved YOLO and transfer learning[J]. Pattern Recognition and Artificial Intelligence, 2019, 32(3): 193-203.

[10] 王晓, 关志强, 王静, 等. 基于卷积神经网络的彩色图像声呐目标检测[J]. 计算机应用, 2019, 39(S1): 187-191.
Wang X, Guan Z Q, Wang J, et al. Target detection of color image sonar based on convolutional neural network [J]. Journal of Computer Applications, 2019, 39(S1): 187-191.

[11] 刘有用, 张江梅, 王坤朋, 等. 不平衡数据集下的水下目标快速识别方法[J]. 计算机工程与应用, 2020, 56(17): 236-242.
Liu Y Y, Zhang J M, Wang K P, et al. Fast underwater target recognition with unbalanced data set[J]. Computer Engineering and Applications, 2020, 56(17): 236-242.

[12] Zhou X Y, Wang D Q, Krähenbühl P. Objects as points [EB/OL]. (2019-04-16)[2021-04-05]. https://arxiv.org/abs/1904.07850.

[13] 张彩珍, 康斌龙, 李颖, 等. 基于差异通道增益及改进Retinex的水下图像增强[J]. 激光与光电子学进展, 2021, 58(14): 1410004.
Zhang C Z, Kang B L, Li Y, et al. Underwater image enhancement based on differential channel gain and improved Retinex[J]. Laser & Optoelectronics Progress, 2021, 58(14): 1410004.

[14] Huang J J, Zhu Z, Huang G. Multi-stage HRNet: multiple stage high-resolution network for human pose estimation[EB/OL]. (2019-10-14)[2021-04-05]. https://arxiv.org/abs/1910.05901.

[15] Park J, Woo S, Lee J Y, et al. BAM: bottleneck attention module[EB/OL]. (2018-07-17)[2021-04-05]. https://arxiv.org/abs/1807.06514.

[16] Liu S T, Huang D, Wang Y H. Receptive field block net for accurate and fast object detection[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11215: 385-400.

[17] Newell A, Yang K Y, Deng J. Stacked hourglass networks for human pose estimation[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9912: 483-499.

[18] 刘鑫, 陈思溢, 陈小龙, 等. 基于深度学习的深层次多尺度特征融合目标检测算法[J]. 激光与光电子学进展, 2021, 58(12): 1210029.
Liu X, Chen S Y, Chen X L, et al. Deep multi-scale feature fusion target detection algorithm based on deep learning[J]. Laser & Optoelectronics Progress, 2021, 58(12): 1210029.

[19] Li Z X, Zhou F Q. FSSD: feature fusion single shot multibox detector[EB/OL]. (2017-12-04)[2021-04-05]. https://arxiv.org/abs/1712.00960.

[20] Dai J F, Qi H Z, Xiong Y W, et al. Deformable convolutional networks[C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 764-773.

[21] 张涛, 张乐. 一种基于多尺度特征融合的目标检测算法[J]. 激光与光电子学进展, 2021, 58(2): 0215003.
Zhang T, Zhang L. Multiscale feature fusion-based object detection algorithm[J]. Laser & Optoelectronics Progress, 2021, 58(2): 0215003.

[22] Lin W H, Zhong J X, Liu S, et al. ROIMIX: proposal-fusion among multiple images for underwater object detection[C]//ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing, May 4-8, 2020, Barcelona, Spain. New York: IEEE Press, 2020: 2588-2592.

[23] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2999-3007.

[24] Law H, Deng J. CornerNet: detecting objects as paired keypoints[J]. International Journal of Computer Vision, 2020, 128(3): 642-656.

[25] Zhou X Y, Zhuo J C, Krähenbühl P. Bottom-up object detection by grouping extreme and center points[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 850-859.