

DOI:10.19651/j.cnki.emt.2211342

高频增强网络与 FPN 融合的水下目标检测*

乔美英 赵岩 史建柯 史有强
(河南理工大学电气与工程学院 焦作 454000)

摘要: 针对水下目标检测中目标对比度低以及水下图像多尺度问题,提出了高频增强网络与特征金字塔(FPN)融合的水下目标检测算法,以提高对水下目标边缘、轮廓信息以及目标底层信息的提取。首先引入八度卷积将卷积层的输出特征按频率分解,将主干网络提取到的特征图进行高、低频信息分离,鉴于水下目标的轮廓信息和噪声信息均包含于高频特征中,在高频信息通道中引入通道注意力具有自适应增强特点的通道注意力机制,形成了一种高频增强卷积,以达到增强有用轮廓特征信息和抑制噪声的目的;其次,将增强的高频特征分量融入 FPN 的浅层网络中,提高原 FPN 对水下多尺度目标的特征表示能力,缓解多尺度目标漏检问题。最后,将所提方法与基线算法 Faster R-CNN 融合,在全国水下机器人竞赛提供的数据集中进行实验。结果表明:改进算法识别准确率达到 78.83%,相比基线提升 2.61%,与其他类型目标检测算法相比,依然具备精度和实时检测优势,证明了从特征图频域角度提升前景和背景对比度的有效性。

关键词: 深度学习;水下目标检测;小目标检测;特征金字塔;八度卷积;通道注意力

中图分类号: TP391 **文献标识码:** A **国家标准学科分类代码:** 510.1050

Underwater target detection based on fusion of high-frequency enhanced network and FPN

Qiao Meiyang Zhao Yan Shi Jianke Shi Youqiang
(School of Electrical and Engineering, Henan Polytechnic University, Jiaozuo 454000, China)

Abstract: Aiming at the problem of low target contrast and multi-scale underwater images in underwater target detection, an underwater target detection algorithm based on the fusion of high-frequency enhanced network and Feature Pyramid Networks (FPN) is proposed. The algorithm improves the extraction of underwater target edge, contour information and target underlying information. Firstly, octave convolution is introduced to decompose the output features of the convolution layer by frequency, and the feature maps extracted by the backbone network are separated from high-frequency and low-frequency information. Since the contour information and noise information of underwater targets are contained in high-frequency features, Squeeze-and-Excitation Network with adaptive enhancement characteristics is introduced into the high-frequency information channel, and a high-frequency enhanced convolution is formed. It can achieve the purpose of enhancing useful contour feature information and suppressing noise. Secondly, the enhanced high-frequency feature components are integrated into the shallow network of FPN. It improves the feature representation ability of the original FPN for underwater multi-scale targets and alleviates the problem of missed detection of multi-scale targets. Finally, the proposed method is fused with the baseline algorithm Faster R-CNN, and the experiment is carried out on the dataset provided by the National Underwater Robot Competition. The results show that the recognition accuracy of the improved algorithm reaches 78.83%, which is 2.61% higher than the baseline. Compared with other types of target detection algorithms, it still has advantages of accuracy and real-time detection. The effectiveness of improving foreground and background contrast from the perspective of feature map frequency domain is demonstrated.

Keywords: deep learning; underwater target detection; small target detection; feature pyramid network; octave convolution; channel attention mechanism

0 引言

主流的目标检测算法主要应用于陆地场景^[1],而在水

下应用的算法研究较少。对于水下目标检测而言,由于受水下环境的影响,导致图像质量降低,目标与背景对比度较低、细节较为模糊等^[2],从而增加了目标检测的难度。

收稿日期:2022-09-09

* 基金项目:国家自然科学基金(41672363)、河南省科技攻关项目(222102220076)资助

传统的目标检测注重特征提取,比如利用纹理、颜色、轮廓等局部特征区分前景和背景。如李鹏等^[2]针对水下声呐图像提出了一种 SURF 算法,结合恒虚警算法来提取特征;Qiao 等^[3]采用自适应轮廓分割的方法对海参目标进行检测;Chen 等^[4]采用加权方式,对光照、颜色、纹理特征融合,用来检测海参目标;Liu 等^[5]在理想条件下,对海参的边缘采用阈值分割的方式捕获其特征;Guo 等^[6]改进了传统的 GrabCut 算法使水下海参目标的检测精度达到 90%。传统方法是针对某些特定场景人工设计完成的,且多用于单一目标的识别任务。当遇到真实水下场景,如多类别目标检测的高度非线性化视觉任务,传统方法泛化能力会变弱而使检测精度降低。

随着深度学习的发展,水下目标检测的研究有了很大进展。相比于人工提取特征,深度学习可以从图像中提取更高阶、更抽象化的层级特征。Fang 等^[7]针对水下目标多尺度的问题,将特征金字塔^[8](feature Pyramid networks, FPN)结构增加了纵向跳跃短接的操作,将底层位置信息和高层语义信息融合,为高层语义信息带来更多底层细节;王红梅等^[9]针对复杂环境下的目标检测,提出 FPN 与多层特征图融合的训练方式;Lin 等^[10]提出将多类别目标区域通过线性叠加的方式来模拟水下遮挡物体;贾振卿等^[11]通过 MSRCR 算法增强水下图像;Pan 等^[12]通过对抗生成网络对图像进行增强,然后提出多尺度 M-ResNet 方法来检测水下目标;Fan^[13]等提出一种具有特征增强和锚框微调功能的水下检测框架,通过引入空洞卷积改变感受野范围以及微调锚框位置增加检测框回归精度来改善水下多类别目标检测效果;张明华等^[14]首先使用 MSRCR 算法对水下图像进行增强,然后再使用 Cascade R-CNN 进行检测。

以上研究在水下目标检测任务中都取得了不错的效果,但针对水下目标边缘和轮廓信息遭到削弱的问题没有进行改进。本文从特征层的频域角度考虑,通过增强目标高频分量的方式来改善其边缘和轮廓被削弱问题,最后针对水下图像中存在目标多尺度问题,在 FPN 的最底层引入高频特征分量,增强了目标底层特征的信息,从而提高了小目标的检测精度。

1 相关工作

1.1 卷积及其变体

Ding 等^[15]提出了非对称卷积,增强模型对旋转失真物体的鲁棒性。Liu 等^[16]提出了自校正卷积,它可以自适应地在每个空间位置周围建立远距离空间和通道的依赖关系,从而生成更具判别力的特征,提取更加丰富的上下文信息。Li 等^[17]提出了内卷卷积,聚合了更广阔的空间上下文信息,从而克服对远距离依赖建模的困难。传统卷积核同时处理高频和低频信息,输出结果不能将输入特征矩阵以不同频率特征分离,如图 1(a)所示,Chen 等^[18]提出了八度卷积,将卷积层的输出特征图按其频率分解。八度卷积可

以通过其特有的设计结构将特征图与卷积核分为两个分量,使卷积核 $W = [W^H, W^L]$ 可以对特征图 $X = \{X^H, X^L\}$ 进行卷积。每个卷积核分量又可细分为频率内部分和频率间部分 $W^H = [W^{H \rightarrow H}, W^{L \rightarrow H}]$ 和 $W^L = [W^{L \rightarrow L}, W^{H \rightarrow L}]$,具体结构如图 1(b)所示。

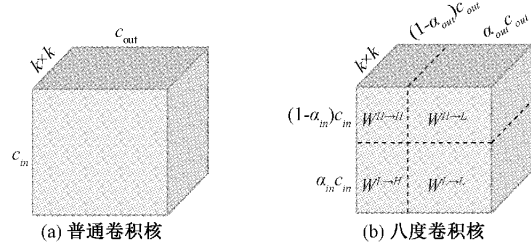


图 1 传统卷积和八度卷积

其中, c_{in} 、 c_{out} 表示输入、输出通道, k 表示卷积核的大小。 $\alpha \in [0, 1]$ 表示分配给低频部分的通道比例。在整个网络中设置 $\alpha_{in} = \alpha_{out} = \alpha = 0.5$, 除了网络的第一个和最后一个八度卷积设置为 $\alpha_{in} = 0, \alpha_{out} = \alpha, \alpha_{in} = \alpha, \alpha_{out} = 0$ 。

1.2 通道注意力机制

在深度神经网络中,不同的通道在特征图中代表着不同的对象^[19],通道注意可以自适应地重新校准每个通道的权重,这个过程可以看作决定关注哪些对象。Jie 等^[20]首次提出了通道注意力的概念,并设计了通道注意力机制(squeeze-and-excitation networks, SeNet)网络结构,他利用通道间的依赖关系进行建模,通过学习的方式自动获取到每个特征通道的重要程度,依照这个重要程度去强调有用的特征并抑制对当前任务用处不大的特征,提高网络的提取能力,其结构如图 2 所示。

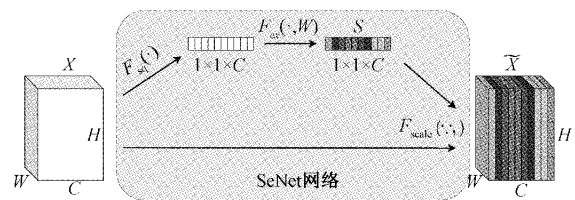


图 2 SeNet 网络结构

SeNet 分为 Squeeze 和 Excitation 两个模块, Squeeze 模块通过全局平均池化收集全局空间信息, Excitation 模块捕获通道方向的向量,并通过全连接层和非线性激活函数输出一个注意力向量,然后将注意力向量中相应的元素相乘,对输入特征图的每个通道进行缩放。

$$s = F_{sc}(X, \theta) = \sigma(W_2 \delta(W_1 \text{GAP}(X))) \quad (1)$$

$$\bar{X} = sX$$

在以后的研究工作中,大都从 Squeeze 和 Excitation 两个模块进行改进,以提高模型捕获全局信息的能力或降低模型的复杂度。如 Qin 等^[21]提出了频域的全局平均池化对 Squeeze 模块进行了改进,增强了获取全局信息的能

力。Wang 等^[22]放弃 Excitation 模块中的全连接层,使用了一维卷积提取通道之间的依赖关系从而降低了模型复杂度。

1.3 FPN

因为卷积神经网络是层层相叠的结构,因此,越深的特征图感受野越大,不同深度的特征图就形成了天然的多尺度表达。SSD 算法^[23]和 MS-CNN 算法^[24]由此被提出,可以利用不同深度的特征图分别检测目标再进行整合,浅层特征图检测小目标,深层特征图检测大目标。但是小目标的检测精度没有得到明显改善,其根本原因是不同深度的特征层表达能力各不相同,浅层特征层虽然保留了更为细粒的空间信息,但语义信息过于薄弱,所以检测效果较差。针对上述缺陷, Lin 等^[8]提出了 FPN,其基本思想在于将浅层特征图的细粒度空间信息与深层特征图的语义信息结合,结构如图 3 所示,算法在区域候选框网络(region proposal network, RPN)网络的基础上增加了由上至下的侧路,从最上层开始经过 1×1 卷积和上采样后与下一层特征图融合得到新的特征图,以此类推,新的特征图经过再与更浅层特征图融合得到新的特征图。实验结果表明, FPN 显著的提高了小目标的检测精度。

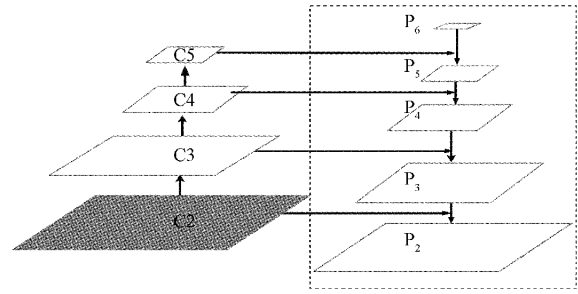


图 3 FPN 结构

2 算法改进

2.1 高频增强卷积的设计

在数字图像领域,高频信息对应着图像的边缘、轮廓和一些噪声信息,在水下目标的高频信息被削弱的环境下,首先利用八度卷积^[18]中的因式分解方式将输出特征层分离为高频特征和低频特征。其次在高频分量上引入 SeNet,通过自适应加权的方式对目标的高频信息进行加权,以达到目标特征增强的目的。该算法称为高频增强卷积(high-frequency enhanced convolution, HFECv)算法。高频增强卷积核的结构如图 4 所示。

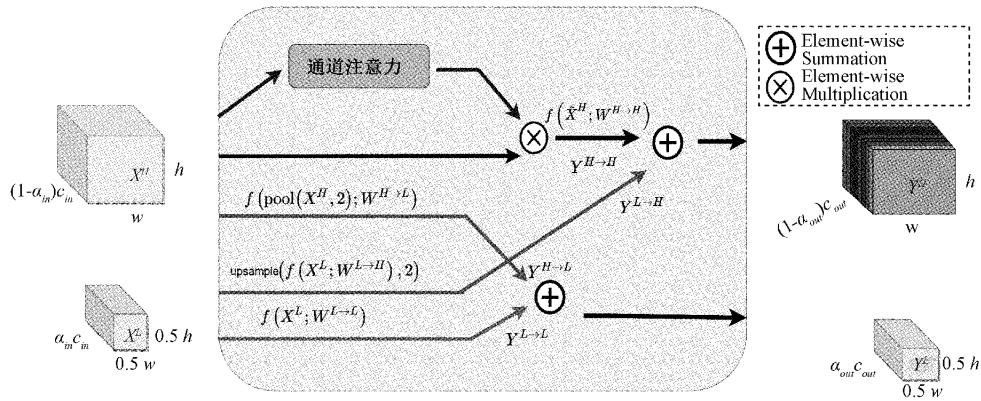


图 4 高频增强卷积细节设计

设 $\mathbf{X} \in \mathbb{R}^{c_{in} \times h \times w}$ 为输入特征,其中 h, w 表示特征图的高和宽, c_{in} 表示通道数。按照通道维度将 \mathbf{X} 因式分解为 $\mathbf{X} = \{\mathbf{X}^H, \mathbf{X}^L\}$, 其中高频特征图 $\mathbf{X}^H \in \mathbb{R}^{(1-\alpha_{in})c_{in} \times h \times w}$ 捕获细节信息,低频特征图 $\mathbf{X}^L \in \mathbb{R}^{\alpha_{in}c_{in} \times \frac{h}{2} \times \frac{w}{2}}$ 得到空间中变换缓慢的信息。

与八度卷积不同的是,高频增强卷积加入了通道注意力结构将高频分量自适应调整,具体实现可由式(2)~(5)表示。

$$\mathbf{X}_c^H \in \mathbb{R}^{H \times W \times C}, c = (1 - \alpha_{in})c_{in} \quad (2)$$

式(2)中 \mathbf{X}_c^H 表示输入特征的高频通道部分, c_{in} 表示输入的总通道数, α_{in} 为输入的低频通道数占总通道数的比值, c 为输入的高频通道数。

$$\mathbf{z}_c^H = F_{sq}(\mathbf{X}_c^H) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W \mathbf{X}_c^H(i, j) \quad (3)$$

式(3)为对高频通道做平均池化操作。 \mathbf{z}_c^H 表示经过全局平均池化后的高频输入特征, F_{sq} 表示压缩操作, H, W 表示特征图的高和宽, (i, j) 表示特征图位置索引。

$$\mathbf{s}_c^H = F_{ex}(\mathbf{z}_c^H, \mathbf{W}) = \sigma(g(\mathbf{z}_c^H, \mathbf{W})) = \sigma(\mathbf{W}_2 \delta(\mathbf{W}_1 \mathbf{z}_c^H)) \quad (4)$$

式(4)表示 SeNet 通道激励部分, \mathbf{s}_c^H 表示输出的高频通道权值, F_{ex} 指通道激励的过程, \mathbf{W} 表示全连接。

\mathbf{z}_c^H 通过权值矩阵 \mathbf{W} 完成激励过程。具体实现为: \mathbf{z}_c^H 与全连接 \mathbf{W}_1 相乘,并将通道压缩至 $\frac{C}{r}$, δ 为 ReLU 激活函数,使得模块具备非线性拟合的能力, \mathbf{W}_2 为另一个全连接层,主要用于将通道调整至初始状态。 $\mathbf{W}_1 \in \mathbb{R}^{\frac{C}{r} \times C}$ 和 $\mathbf{W}_2 \in \mathbb{R}^{C \times \frac{C}{r}}$, r 是一个缩放参数,本文取值 16,目的是减少模型的

计算量。 σ 是 sigmoid 函数,最后通过 σ 进行归一化,得到每个通道对应的权重 s_c 。

$$\tilde{X}^H = F_{scale}(X_c^H, s_c^H) = s_c^H X_c^H \quad (5)$$

式(5)是对高频特征加权的描述, \tilde{X}^H 表示加权后的高频特征。

经过高频增强卷积后输出高频和低频特征图 $Y = \{Y^H, Y^L\}$, 其中 $Y^H = Y^{H \rightarrow H} + Y^{L \rightarrow H}, Y^L = Y^{H \rightarrow L} + Y^{L \rightarrow L}$, $Y^{A \rightarrow B}$ 表示从特征图组 A 到特征图组 B 的卷积更新。其中, $Y^{H \rightarrow H} + Y^{L \rightarrow L}$ 表示同频率之间的信息交互, $Y^{L \rightarrow H} + Y^{H \rightarrow L}$ 表

示不同频率之间的信息交互。

2.2 高频增强残差网络设计

图 5 中是不同卷积核组成的残差模块,图 5(a)是普通卷积核组成的残差模块 Bottleneck,图 5(b)是高频卷积核组成的残差模块 HFEBottleneck,图 5(c)是八度卷积核组成的残差模块 OctBottleneck。Bottleneck 为单输入和单输出,输入数据和输出数据没有将高频和低频特征层分离, HFEBottleneck 和 Bottleneck 将高频和低频特征分离,不同的是 HFEBottleneck 可以将高频分量通过注意力机制进行增强。

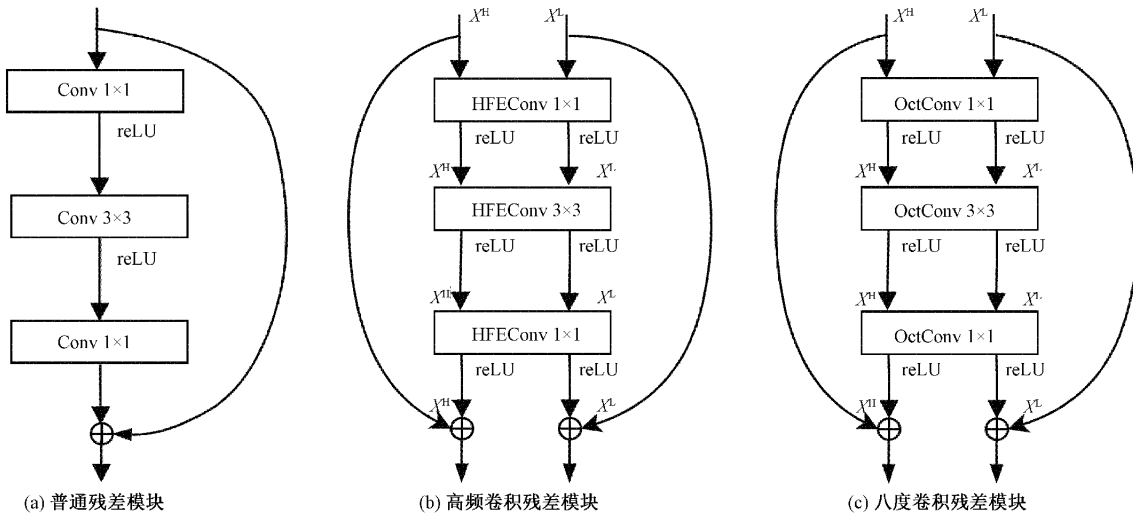


图 5 改进前后残差模块对比

ResNet-50^[25] 由 layer1-4 组成,每一层由不同的残差模块堆叠而成,本文将 ResNet-50 的 layer1 层中的 Bottleneck 替换为 HFEBottleneck,使得主干网络在特征提取过程中可以对目标的高频信息进行增强,即增强目标的轮廓和边缘特征。为了减少主干网络的计算复杂度,将 ResNet-50 其他 layer 层的 Bottleneck 替换为 OctBottleneck,即形成了高频增强残差网络 (high frequency enhanced ResNet, HFEResNet),具体结构如图 6 所示。

2.3 高频增强特征金字塔结构的设计

FPN 结构是多尺度检测的通用手段,但其存在明显的不足,FPN 的底层特征图 P2 在多层特征层间传递融合后,其空间信息遭到削弱,不利于小目标检测。为了减少目标底层特征的信息损失,将 layer1 层输出的高频特征图与 P2 增加残差连接,可以增强目标的轮廓、纹理等信息,提升网络的精度。其中残差链接具体操作为将 \tilde{X}^H 通过 1×1 卷积更改通道数量, P2 特征层进行融合。该 FPN 的改进方案被称为高频增强特征金字塔 (high frequency enhanced FPN, HFE-FPN)。

3 实验结果及分析

本实验的目的是验证本文中改进的算法在水下目标检测中的有效性和优越性。首先针对数据集不足的问题,

采用了迁移学习的方法,实验中的特征提取网络都加载了 ImageNet 数据集训练后权重,然后将本文中的改进网络和基准网络、主流水下目标检测算法进行对比,最后将改进网络部分输出的特征层进行可视化和频率分析。

3.1 实验数据集

实验采用了全国水下机器人大赛数据集,数据集中包含 4 种水下生物模型,将其作为检测目标,去除标注错误的图片共 5 455 张图片,然后按照 7 : 3 的比例划分训练集和测试集。

3.2 实验环境及评价指标

实验采用 GPU 训练方式,并利用 CUDNN 进行加速处理,实验计算机所采用 CPU 为 Intel core i7-10750H 处理器, GPU 为 Nvidia GTX1080TI。

在目标检测任务中交并比 (intersection over union, IoU),表示预测框与真实框交集和并集的比值, IoU 越大对网络的检测精度要求越严格;多个目标类别的平均精度 (mean average precision, mAP) 数值介于 0~1,值越大表明模型识别效果越佳。

3.3 对比实验

考虑到 YOLOv4 和 YOLOv5 对基于 anchor 策略上有一定的过度优化^[26],所以在 YOLO 系列中,本文采用了 YOLOv2^[27]和 YOLOv3^[28]作为代表与本文提出的网络进

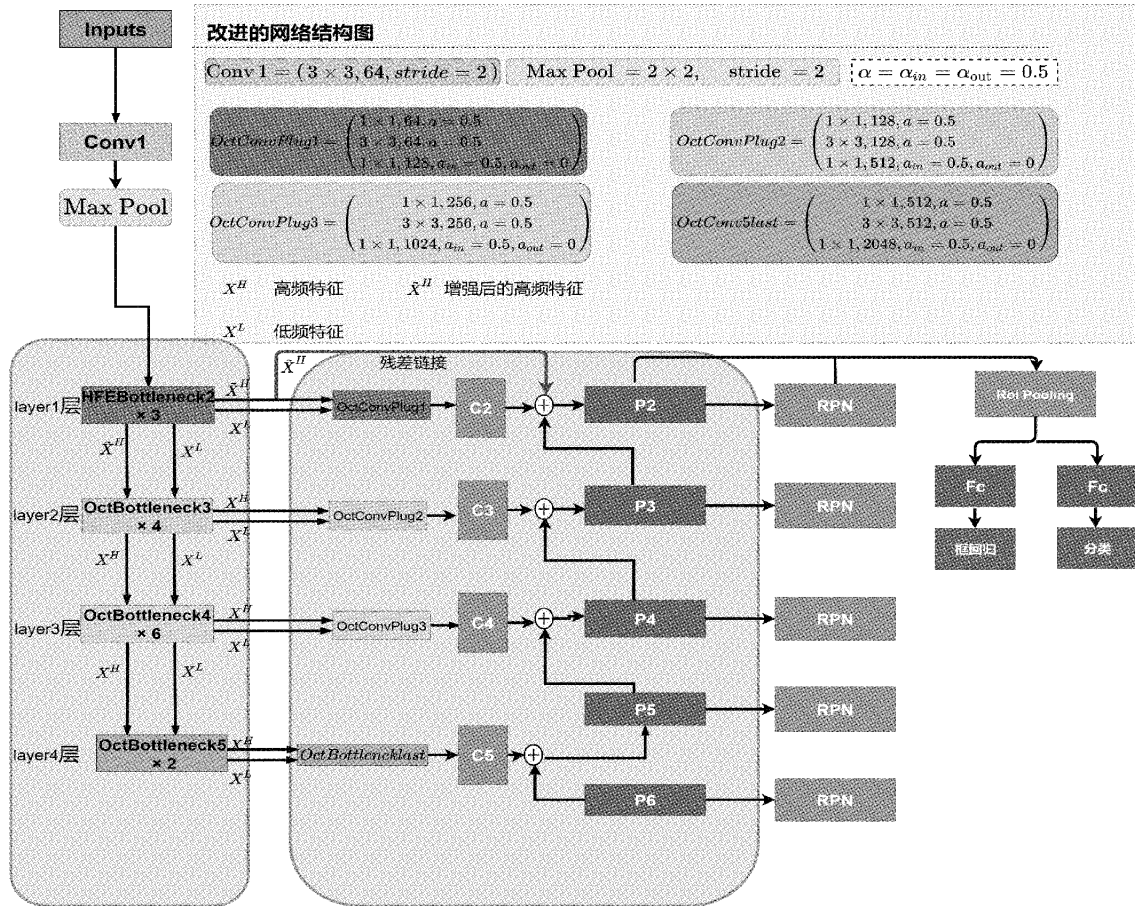


图 6 HFEResNet 结构

行对比,并且本文还选取了一些主流二阶段目标检测网络进行对比试验。

为了测试本算法是否具有优越性,首先在 PASCAL VOC 公共数据集中进行算法测试,其中训练集为 PASCAL VOC2007 和 PASCAL VOC2012 的训练集,测试集为 PASCAL VOC2007 的测试集。总训练轮次为 200,初始学习率为 2×10^{-3} ,训练至 100 轮和 150 轮之后,学习率分别下降至 2×10^{-4} 与 2×10^{-5} ,训练批次值为 32,

实验对比结果如表 1 所示。

由表 1 可知一阶段算法中 YOLOv2 速度最快,但检测精度较低,RefineDet^[31]算法精度达到最高 79.5%,原因是它采用多次微调与特征转移结构。二阶段网络中,R-FCN 算法和 MR-CNN 算法虽然精度达到 76.9%和 78.2%,但检测速度滞后一阶段算法。本算法的主干网络基于八度卷积及其改进,所以低频特征图的高和宽缩小了一半,计算量大幅降低,在不影响精度的情况下检测速度为 12.1 帧/s。

表 1 PASCAL VOC 数据集实验对比

算法	主干网络	输入尺寸/像素	速度/(帧·s ⁻¹)	mAP/%
YOLOv2 ^[27]	Darknet-19	416×416	34.0	76.2
YOLOv3 ^[28]	Darknet-53	416×416	31.0	78.4
RON ^[29]	VGG-16	320×320	12.0	74.2
SSD ^[30]	VGG-16	512×512	15.0	77.9
RefineDet ^[31]	VGG-16	320×320	33.0	79.5
R-FCN ^[32]	ResNet-50	1 000×600	12.0	76.9
MR-CNN ^[33]	VGG-16	1 000×600	0.1	78.2
Faster R-CNN ^[34]	ResNet-50+FPN	600×600	9.0	76.4
Faster R-CNN	HFEResNet-50+HFE-FPN	600×600	12.1	78.7

该实验将本文算法和其它水下目标检测算法^[9,15,29,31,36,37]进行了对比,超参数如下设置:选用 SGD 优化器,学习率设置为 0.05,动量为 0.9,权值衰减为 0.0005,

训练批次 Batch_size 为 2,最大训练次数为 22 884,每 3 814 次进行一次测试,学习率调度为步长调度器,每迭代 3 814 次,衰减 0.33。实验结果如表 2 所示。

表 2 水下数据集实验对比

算法	主干网络	海胆	海参	扇贝	海星	mAP area=all (IoU=0.50)
SSD512 ^[30]	VGG-16	84.50	69.30	66.90	80.50	75.3
YOLOv3 ^[28]	Darknet-53	85.60	71.80	70.60	79.90	76.8
S-FPN ^[7]	ResNet-50	73.09	73.25	82.47	75.16	75.99
RefineDet ^[31]	ResNet-50+FPN	85.40	65.80	71.00	80.90	75.80
FERNet ^[13]	CCB	91.50	71.40	52.20	82.00	74.20
Faster R-CNN ^[35]	VGG16+RoIMiX	86.80	73.27	55.97	83.65	74.92
Faster R-CNN ^[36]	VGG-16+AON	80.20	65.60	63.40	79.10	72.10
Faster R-CNN*	ResNet-50	90.12	70.31	52.41	79.56	73.10
Faster R-CNN	HFEResNet-50	91.60	72.30	55.14	80.6	74.91
Faster R-CNN*	ResNet-50+FPN	86.23	64.07	69.12	80.86	76.22
Faster R-CNN	HFEResNet-50+HFE-FPN	87.3	75.22	65.90	86.79	78.83

本文以 Faster R-CNN 为基准网络,实验中,将基准网络的主干网络为本文提出的 HFEResNet-50 时,可以提升 1.81%,引入 HFE-FPN 后可以提高 2.61%,从而证明了本文中改进模型的有效性。不同于陆地场景数据,在面对噪声更多的水下数据时,其他目标检测算法的检测精度都有明显的下降,但本文算法在面对复杂场景时表现出了良好的抗噪性,精度与 PASCAL VOC 数据上的精度相比,精度提高了 0.13%。

为了可以直观的展示本算法的有效性,本实验从多目标、单目标图像中选取了几组不同位姿和模糊度的图像,如图 7 和 8 所示,基线算法存在明显的漏检现象,本文提出的算法增强了目标底层特征的信息,所以在远距离和近距离图像中减少了漏检现象,并且在目标预测的准确率上有很大提升。

多类别目标检测对比图如图 7 所示。左列以 ResNet-50+FPN 为基线算法的检测结果,右列是本文算法的检测结果。图 7(a)是近距离多目标清晰图像,图 7(b)是近距离多目标模糊图像,图 7(c)是多目标远距离清晰图像,图 7(d)是多目标远距离模糊图像。

单类别目标检测效果对比图如图 8 所示。左列为基线算法的检测结果,右列是本算法的检测结果。图 8(a)是单类别近距离清晰图像,图 8(b)是单类别的情况下近距离模糊图像,图 8(c)是单类别远距离清晰图像,图 8(d)是单类别远距离模糊图像。

为了验证本文提出的主干网络对输入图像高频特征的增强能力,将输入图像主干网络的 layer1 层输出的特征层可视化并对输出的特征层进行频域分析,如图 9 所示。图 9(a)为原始输入图像,图 9(b)和(c)是对输入图像的特

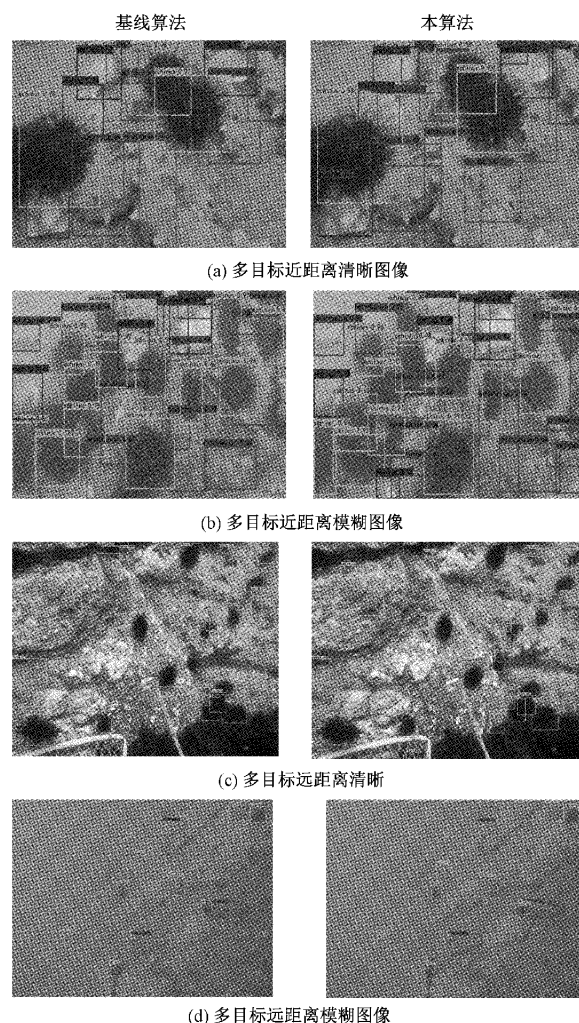


图 7 多类别目标检测结果

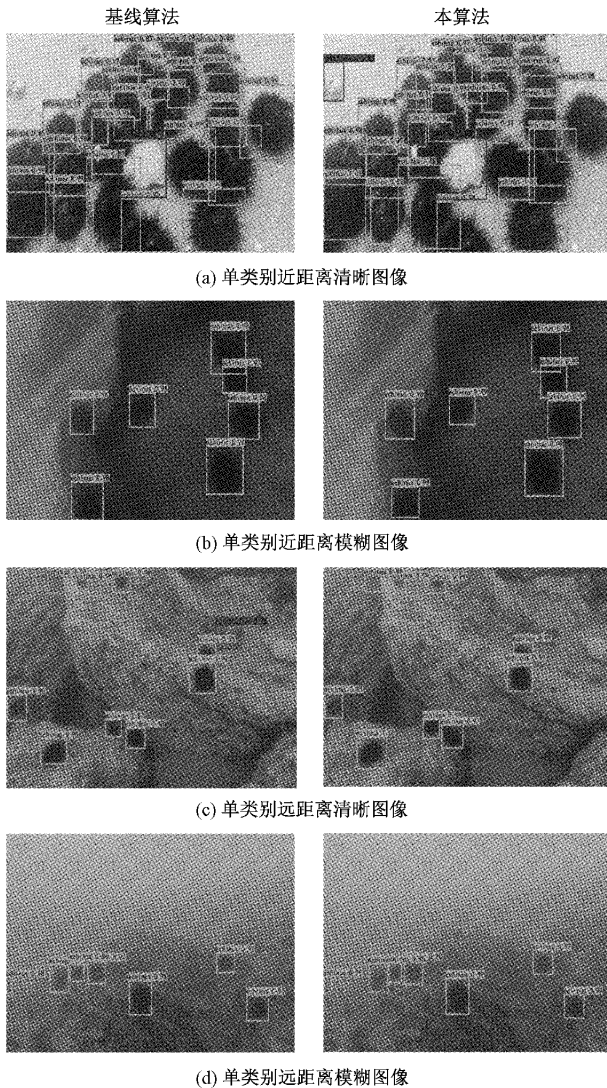


图 8 单类别目标检测结果

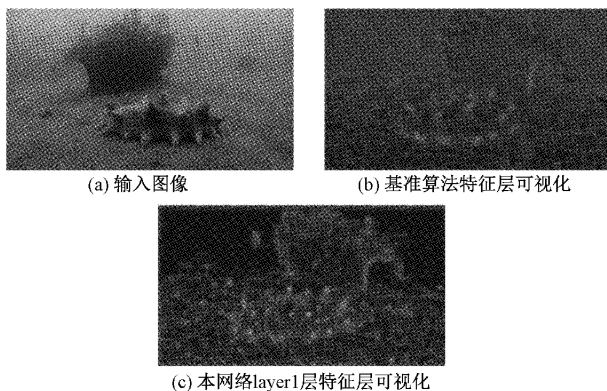


图 9 特征层可视化

征层可视化,其中颜色越鲜亮的部分表示对该部分在特征提取任务中的贡献越大,相比于基准的特征图可视化结果,可看出改进后的网络对输入图像中的目标的高频特征得到了有效的增强,进而提高了对目标的检测能力。

频率分析如图 10 所示。图 10(a)和(b)是将对应的输出特征层进行傅里叶变换,并将平均结果可视化,图像的四个角落为高频信息,图像中间为低频信息。由两图对比可以看出,原始 layer1 的特征层输出的高频分量低于本文中改进的网络的高频分量,从频率上证明了本网络对高频分量起到了放大作用。

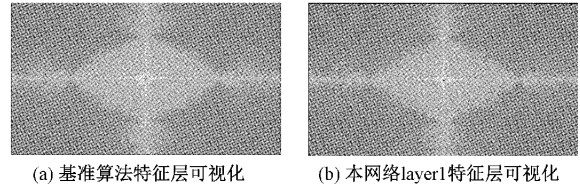


图 10 输出特征层进的傅里叶变换

4 结 论

本文针对水下目标与背景对比度不高的问题,提出一种高频增强卷积,并将其融入浅层主干网络中。针对小目标的问题,将改进的主干网络与 FPN 相结合,满足实际应用的检测精度要求,证明了本文所提算法对复杂水下环境的有效性和实用性,为水下目标检测任务提供了一种新的思路。但本文未考虑目标聚集与遮挡的问题,这是未来改进的一个方向。

参考文献

- [1] SUN X, SHI J, LIU L, et al. Transferring deep knowledge for object recognition in Low-quality underwater videos[J]. Neurocomputing, 2018 (275): 897-908.
- [2] 李鹏, 马味敏. 融合特征点与环状域检测的水下声呐目标匹配[J]. 电子测量与仪器学报, 2019, 33(11): 1020-1127.
- [3] QIAO X, BAO J, ZENG L, et al. An automatic active contour method for sea cucumber segmentation in natural underwater environments [J]. Computers and Electronics in Agriculture, 2017(135): 134-142.
- [4] CHEN Z, WANG H B, XU L Z, et al. Visual-adaptation-mechanism based underwater object extraction[J]. Optics & Laser Technology, 2014(56): 119-130.
- [5] LIU H, XU Q, LIU S L, et al. Evaluation of body weight of sea cucumber Apostichopus japonicus by computer vision[J]. Chinese Journal of Oceanology and Limnology, 2015(1): 114-120.
- [6] GUO C, LI Z, QIAO X, et al. Image segmentation of underwater sea cucumber using GrabCut with saliency map[J]. Transactions of the CSAM, 2015, 46(S1): 47-152.
- [7] FANG P, ZHENG M, FEI L A, et al. S-FPN: A

- shortcut feature pyramid network for sea cucumber detection in underwater images[J]. *Expert Systems with Applications*, 2021(182):1-13.
- [8] LIN T, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]. *Computer Vision and Pattern Recognition*, 2017:936-944.
- [9] 王红梅, 王晓鸽, 王晓燕. 基于深度学习的复杂背景下目标检测研究[J]. *控制与决策*, 2021:1-8.
- [10] 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 (ICASSP)*, IEEE, 2020, 2020:1-9.
- [11] 贾振卿, 刘雪峰. 基于 YOLO 和图像增强的海洋动物目标检测[J]. *电子测量技术*, 2020, 43(14):84-88.
- [12] PAN T S, HUANG H C, LEE J C, et al. Multi-scale ResNet for real-time underwater object detection[J]. *Signal Image and Video Processing*, 2020:1-9.
- [13] FAN B, CHEN W, CONG Y, et al. Dual refinement underwater object detection network [C]. *European Conference on Computer Vision*. Springer, Cham, 2020:275-291.
- [14] 张明华, 龙腾, 宋巍, 等. 基于改进 Cascade R-CNN 和图像增强的夜晚水下鱼类目标检测[J]. *农业机械学报*, 2021:1-9.
- [15] DING X, GUO Y, DING G, et al. Acnet: Strengthening the kernel skeletons for powerful cnn via asymmetric convolution blocks[C]. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019: 1911-1920.
- [16] LIU J J, HOU Q, CHENG M M, et al. Improving convolutional networks with self-calibrated convolutions [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 10096-10105.
- [17] LI D, HU J, WANG C, et al. Involution: Inverting the inherence of convolution for visual recognition[C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 12321-12330.
- [18] CHEN Y P, FAN H Q, XU B, et al. Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution [C]. *IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, 2019:3434-3443.
- [19] CHEN L, ZHANG H, XIAO J, et al. Sca-cnn: Spatial and channel-wise attention in convolutional networks for image captioning[C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 5659-5667.
- [20] JIE H, LI S, GANG S, et al. Squeeze-and-excitation networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 42(8):2011-2023.
- [21] QIN Z, ZHANG P, WU F, et al. Fcanet: Frequency channel attention networks [C]. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021: 783-792.
- [22] WANG Q, WU B, ZHU P, et al. Supplementary material for 'ECA-Net: Efficient channel attention for deep convolutional neural networks[C]. *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Seattle, WA, USA. 2020: 13-19.
- [23] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. *European Conference on Computer Vision*. Springer, Cham, 2016: 21-37.
- [24] CAI Z, FAN Q, FERIS R S, et al. A unified multi-scale deep convolutional neural network for fast object detection [C]. *European Conference on Computer Vision*. Springer, Cham, 2016: 354-370.
- [25] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C]. *Proceedings of International Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE Press, 2016: 770-778.
- [26] GE Z, LIU S, WANG F, et al. YOLOx: Exceeding yolo series in 2021 [J]. *ArXiv Preprint*, 2021, ArXiv:2107.08430.
- [27] REDMON J, FARHIDI A. YOLO9000: Better, faster, stronger [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2017: 6517-6525.
- [28] REDMON J, FARHADI A. YOLOv3: An incremental improvement[J]. *ArXiv Preprint*, 2018, ArXiv:1804.02767.
- [29] KONG T, SUN F C, YAO A B, et al. RON: Reverse connection with objectness prior networks for object detection[C]. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2017: 5936-5944.
- [30] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. *Proceedings of European Conference on Computer Vision*,

- Heidelberg: Springer, 2016: 21-37.
- [31] ZHANG S F, WEN L Y, BIAN X, et al. Single-shot refinement neural network for object detection [C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 4203-4212.
- [32] DAI J, LI Y, HE K, et al. R-FCN: Object detection via region-based fully convolutional networks [J]. Advances in Neural Information Processing Systems, 2016, 29.
- [33] GIDARIS S, KOMODAKIS N. Object detection via a multi-region and semantic segmentation-aware CNN model [C]. Proceedings of IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2015: 1134-1142.
- [34] 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.
- [35] LI C, GUO C, REN W, et al. An underwater image enhancement benchmark dataset and beyond[J]. IEEE Transactions on Image Processing, 2019, 29: 4376-4389.
- [36] WANG P Q, CHEN P F, YUAN Y, et al. Understanding convolution for semantic segmentation[C]. Proceedings of IEEE Winter Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2018: 1451-1460.

作者简介

乔美英, 博士, 副教授, 主要研究方向为目标识别、故障诊断、MWD 数据分析与处理。

E-mail: qiaomy@hpu.edu.cn

赵岩, 硕士研究生, 主要研究方向为计算机视觉, 目标检测。

E-mail: zhaoyanjoin@gmail.com

史建柯, 硕士研究生, 主要研究方向为计算机视觉, 目标检测。

E-mail: abettertomorrow2@foxmail.com