

基于深度学习的循环水养殖鳗鲡(*Anguilla*) 计数研究*

李 凯^{1,2} 江兴龙^{1,2} 陈尔康^{2,3} 陈 彭^{2,3} 许志扬^{1,2} 林 茜^{1,2}

(1. 集美大学水产学院 福建厦门 361021; 2. 鳗鲡现代产业技术教育部工程研究中心 福建厦门 361021; 3. 集美大学海洋与信息工程学院 福建厦门 361021)

摘要 鳗鲡(*Anguilla*)作为我国优质水产养殖种类, 精准掌握其数量对高效养殖有重要意义。为实现对循环水养殖鳗鲡的准确计数, 提出了一种基于深度学习的改进 Faster RCNN 模型。针对检测目标即鳗鲡头部尺寸小的问题, 选择在特征提取网络 ResNet50 中加入 FPN 结构来作为模型的骨干网络, 以提取并融合多尺度的特征; 针对原模型锚框都是基于人工经验设置的, 并不适用于鳗鲡数据集的问题, 使用 k -means 聚类算法对训练集中标注的鳗鲡头部检测框进行聚类分析, 获得了适合鳗鲡数据集的 15 种不同尺度的锚框; 针对图像中存在鳗鲡头部重叠的问题, 选择使用 Soft-NMS 算法替代原 NMS 算法对 RPN 部分生成的候选框进行筛选, 以减少模型对鳗鲡重叠部分的漏检情况。试验结果表明: 改进后的 Faster RCNN 模型对鳗鲡头部的检测精度(mAP^{0.5})高达 96.5%, 较原 Faster RCNN 模型(Backbone 为 ResNet50)显著提升了 14%, 与 SSD300 和 YOLOV3 模型相比分别显著提升了 24.9%和 15%; 在鳗鲡计数上, 利用改进后的 Faster RCNN 模型检测结果进行计数, 计数准确率达到 90%以上, 提升了模型对鳗鲡的检测识别能力。

关键词 鳗鲡计数; 深度学习; Faster RCNN 模型; FPN 结构; k -means 聚类算法; Soft-NMS 算法
中图分类号 Q959.9; S965 doi: 10.11693/hyhz20211200332

鳗鲡(*Anguilla*)是我国常见的淡水养殖鱼类之一, 因其肉质细嫩、营养丰富, 且具有很高的经济价值, 而成为我国的一种高档水产品(张婷婷等, 2019)。20 世纪 70 年代初期我国的福建、广东等地开始养殖鳗鲡, 至 20 世纪 80 年代中期之后迅速发展, 现今我国已经成为养鳗业的世界第一大国(叶翠等, 2019)。而在鳗鲡养殖过程中, 掌握鳗鲡的准确数量有助于科学投喂饲料、成活率计算以及合理控制养殖密度, 对实现鳗鲡精准高效养殖有重要意义。

传统的鱼类计数方法多为人工计数法, 由养殖场工人将鱼从养殖池中捞出, 逐条进行计数后再放回养殖池中, 这类方法存在耗费人力, 计数误差大, 效率低等缺点(黄玲等, 2012), 而且在捞鱼计数过程

中会使鱼产生应激反应, 甚至造成鱼体损伤, 不利于鱼的后续生长发育。在 20 世纪 80 年代之后我国陆续出现了一些现代化的计数器, 比如多通道式、电桥式、光电式等鱼苗计数器(王宝舫, 1988; 李伯珩等, 1991; 张康德, 1997), 但是容易受到鱼苗大小和通道管径大小的影响, 导致计数误差大, 而且设备价格高昂。随着计算机和图像处理技术的发展, 基于图像处理的计数方式逐渐成为一种重要的鱼类计数方法。Zhang 等(2021)通过对采集的鱼苗图像使用二值化、膨胀、腐蚀等一系列图像处理方法, 提取出鱼苗的特征, 然后使用连通面积算法和细化算法对图像中的鱼苗进行计数, 对比两种算法之后发现细化算法对于鱼苗重叠度较高的图像的准确率远

* 国家重点研发计划“特色鱼类精准高效养殖关键技术集成与示范”, 2020YFD0900102 号。李 凯, 硕士研究生, E-mail: 924140205@qq.com

通信作者: 江兴龙, 博士, 教授, E-mail: xinlongjiang@jmu.edu.cn

收稿日期: 2021-12-19, 收修改稿日期: 2022-01-26

高于连通面积算法,为密集鱼苗计数提供了一种新方法。但是这种方法对采集的鱼苗图像有很高的要求,比如图像前景和背景对比度要明显,并且图像要清晰、噪声少,而在真实养殖的情况下很难满足这些要求。王文静等(2016)拍摄一个具有稳定流速的水箱,鱼苗会随水流平稳地流过水箱,通过这种方法来采集鱼苗图像,对采集到的图像进行局部阈值分割以提取出鱼苗目标,然后对图像中不重叠的区域进行计数,最后累加计算鱼苗总数。这种方法的计数误差控制在 15%以内,但是易受水流速大小、鱼苗大小和数量等因素的影响。王硕等(2015)利用常值演化法实现了对图像中黏连重叠鱼苗的准确计数,但是这种方法可能存在鱼体和鱼尾分开演化成不同目标的问题,也很容易受到外部杂质的影响,造成计数结果偏多的情况。

目前,深度学习技术在鱼类检测方面也得到了应用(段青玲等,2018)。以目标检测算法为例,典型的有 Faster RCNN (Ren *et al.*, 2015)、YOLO (Redmon *et al.*, 2016)、SSD (Liu *et al.*, 2016)等,它们在 ImageNet、COCO 等大型数据集上都有很好的表现,因此利用深度学习技术对鱼类进行检测和计数成为了研究学者们的重要研究课题之一。Mandal 等(2018)利用 Faster RCNN 模型对水下视频中的鱼类进行检测和识别,克服了水下图像模糊、背景复杂、光照不均等因素带来的挑战,最终模型得到的准确率达到 82.4%,相比以往提出的方法准确率更高。Labao 等(2019)提出了一种基于深度学习的海洋鱼类检测系统,用于对水下鱼类进行检测和计数,该系统的算法以 Faster RCNN 模型结构为模板,由一系列基于区域的卷积神经网络组成,并且这些卷积神经网络通过 LSTM 网络以级联的形式连接,最终系统得到的计数准确率约为 70%。崔榛(2020)提出了一种改进的 YOLOV3 模型,实现了对红鳍东方鲀的检测和计数,试验结果表明改进后的 YOLOV3 模型有效降低了对鱼类密集区域的漏检情况,最终模型的计数准确率为 92.5%。深度学习模型的网络结构更深,并且它所提取到的特征是神经网络经过大量的学习得到,不易受鱼类大小、图像噪声、光照不均等因素的影响,相较于图像处理等方法具有更强的鲁棒性(Robust)。综上,本研究将针对鳗鲡体型和姿态的特殊性,在原 Faster RCNN 模型基础上进行改进,以提升模型对鳗鲡的检测识别能力,最后基于模型的检测结果对图像中的鳗鲡进行计数。

1 图像采集与标注

1.1 图像采集

实验图像均采集于集美大学水产试验场的鳗鲡循环水养殖车间。经过实验发现若拍摄过程中使用普通光源进行补光,循环水养殖系统中的养殖鳗鲡会因突然的光线刺激而四处躲避,这是因为鳗鲡的生活习性为昼伏夜出(陈学豪,2003),对普通光线很敏感;当使用红外灯进行补光时,通过长时间观察后发现红外光线不会对养殖鳗鲡活动产生影响,因此本文拍摄用的相机选择近红外工业相机(MER-232-48NIR)并使用红外灯补光。图像采集和处理由红外相机、红外灯、计算机完成。整个图像采集系统如图 1 所示,红外相机固定于养殖桶上方中央位置,为避免水面反光,选择把两个红外灯分别安装在离相机较远的两侧,相机和计算机通过一根长 4 m 的千兆双绞线相连。图像采集过程中为避免增氧设备产生的气泡对图像处理造成影响,选择将增氧设备短暂关停,把每次图像采集时间都严格控制在 30 min 以内,以避免鳗鲡缺氧。

试验选择体重在 150~950 g 之间的南美洲鳗鲡为研究对象,总共采集了 1 000 张图像,图像中鳗鲡的数量在 13~45 条之间。

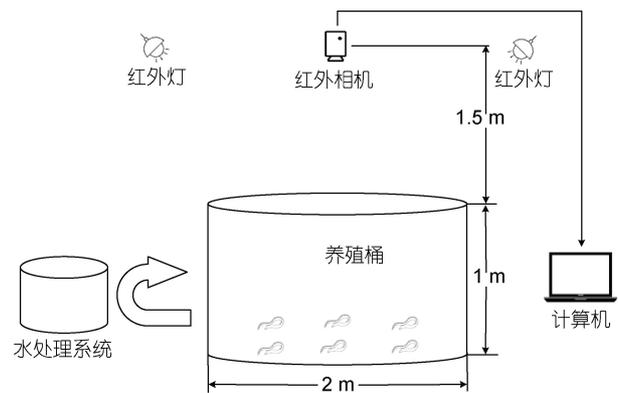


图 1 图像采集系统

Fig.1 The image acquisition system

1.2 图像标注

图像标注工具选择 labelImg, 每张图像的标注信息都会存储在一个 xml 文件中。由于鳗鲡体型呈棍棒型且姿态多变,若对整条鱼进行标注如图 2,会出现一个矩形框中包含多条鱼的情况,存在太多冗余信息,容易对模型检测效果造成不利影响。因此参照汪梦婷等(2020)的方法,只对特征明显的鱼头部进行标注如图 3。

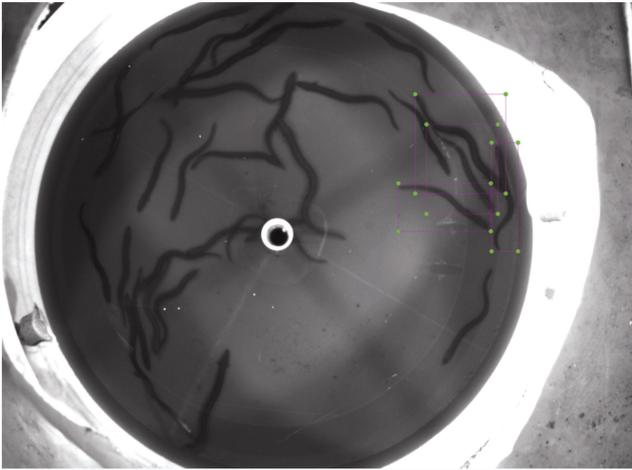


图2 标注整条鱼(方框示例: 标注四条鱼)

Fig.2 The whole-fish labeling (four fish are labeled in this case)

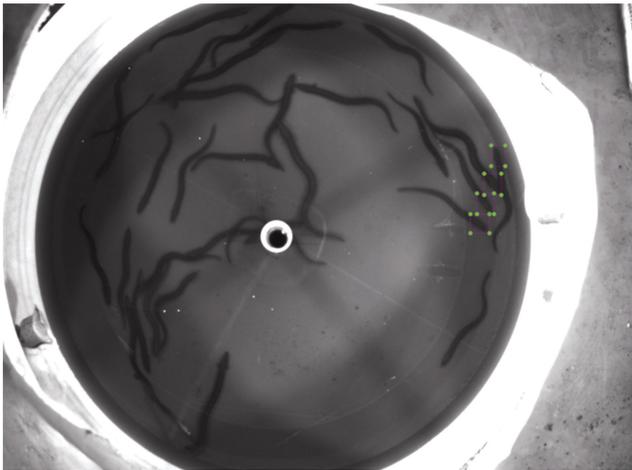


图3 标注鱼头部(方框示例: 标注四条鱼)

Fig.3 The fish-head labeling (four fish are labeled in this case)

2 Faster RCNN 模型及改进

2.1 Faster RCNN 模型

Faster RCNN 是 Girshick 团队继 RCNN (Girshick *et al*, 2014)、Fast RCNN (Girshick, 2015)之后又一力作, 相较于前两个模型, 它的主要贡献在于用 RPN 网络代替了非常耗时的 Selective Search 算法来生成候选框, 这一改进大大提高了整个模型的检测速度, 使其在 GPU 上检测一张图片仅需 0.2 s, 与此同时模型的检测精度也获得了提升。

Faster RCNN 模型结构如图 4 所示, 它将目标检测大致分为 3 个步骤: 首先, 待检测图像输入到模型的骨干网络如 ResNet (He *et al*, 2016)中提取特征, 此特征图被用于后续的 RPN 网络和全连接层; 之后经过 RPN 网络后获得一系列候选框, 使用 NMS 算法对这些候选框进行筛选; 然后将得到的不同尺寸的特征图通过 ROI Pooling 层统一缩放至固定尺寸, 最后让特征图通过后续的全连接层进行类别预测和边界框回归预测。

2.2 Faster RCNN 模型改进

由 1.2 节图像标注可知, 鳗鲡数据集有两个特点, 一是检测目标即鳗鲡头部尺寸较小, 二是图像中存在部分鳗鲡重叠的情况。如果利用原 Faster RCNN 模型进行检测会存在以下两个弊端:

(1) 对小尺度目标检测效果差。原 Faster RCNN 模型只在骨干网络最后一层的特征图上进行预测, 因为骨干网络通常选择层数较深的网络(如 ResNet50 等), 在网络最后一层的特征图尺寸小, 包含更多的是图像的高级语义特征, 损失了图像的细节特征, 所以会导致模型对小尺度目标即鳗鲡头部检测效果差。

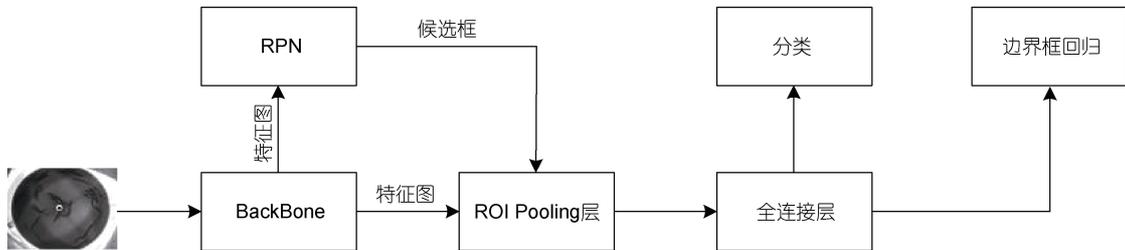


图4 Faster RCNN 模型结构

Fig.4 Structure of the Faster RCNN model

(2) 对鳗鲡重叠区域容易出现漏检的情况。原 Faster RCNN 模型中的 NMS (非极大值抑制)算法有个很大的缺点, 就是会将相邻的两个高置信度的检测框中的一个直接去除, 从而导致模型出现漏检的情况。

本文针对上述两个弊端, 提出改进的 Faster RCNN 模型, 如图 5 所示。鳗鲡图像输入 ResNet50+FPN 骨干网络中提取多尺度特征并融合, 利用 *k*-means 聚类算法获取适合鳗鲡头部检测的锚框 (anchors), 经过 RPN 网络后获得一系列候选框, 使用

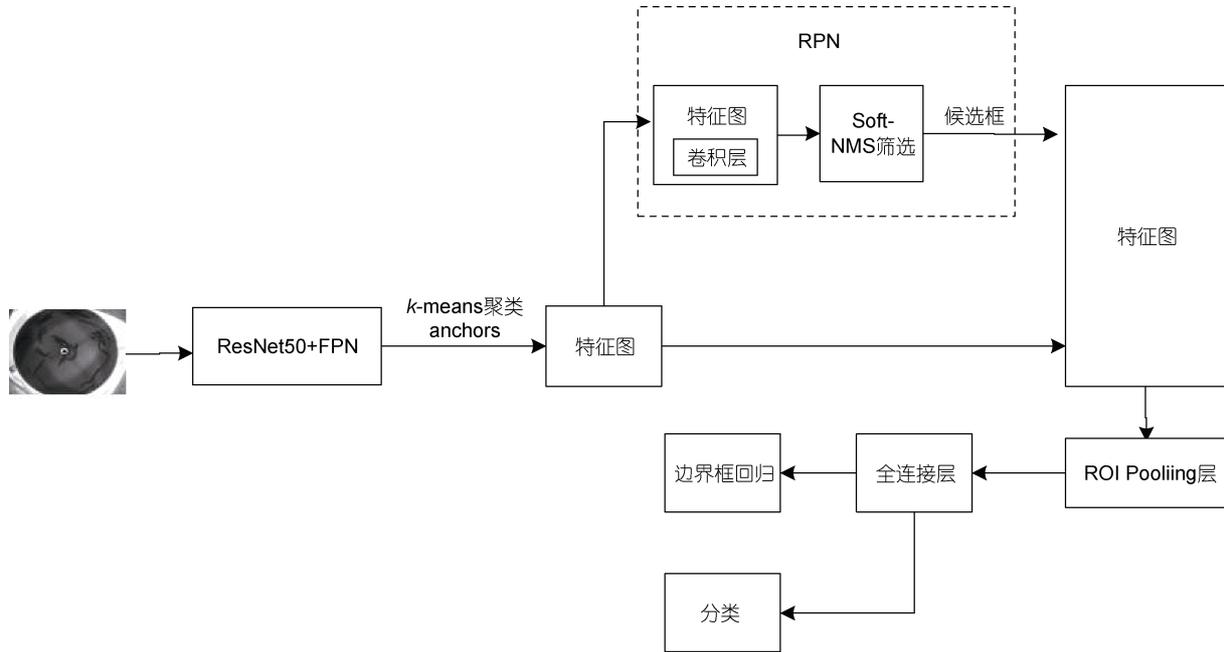


图 5 改进的 Faster RCNN 模型
 Fig.5 The improved Faster RCNN model

Soft-NMS 算法对这些候选框进行筛选, 之后将得到的不同尺寸的特征图通过 ROI Pooling 层统一缩放至固定尺寸, 最后特征图经过一系列全连接层进行类别预测和边界框回归位置预测。接下来将对改进的 Faster RCNN 模型作具体介绍。

2.2.1 加入 FPN 的多尺度特征提取网络 在 Faster RCNN 模型的特征提取网络即骨干网络中加入 FPN (Lin *et al*, 2017)结构, 能够将网络深层特征和浅层特征相融合, 使得模型在检测过程中不会损失图像的细节特征, 以提升模型对小尺度目标的检测能力。因此本文选择在 ResNet50 网络中嵌入 FPN 结构来作为模型的骨干网络, 如图 6 所示, 分为 ResNet50 和 FPN 两个部分, 下面将对此网络作具体分析。

ResNet50 部分包含 Conv1、Conv2_x、Conv3_x、Conv4_x 以及 Conv5_x 五个卷积模块, C2、C3、C4 以及 C5 分别表示 ResNet50 后四个卷积模块的最后一层输出的特征图, 这四层特征图的维度分别为 256 维、512 维、1024 维和 2048 维。各层特征图输入 FPN 部分后先经过一个 1×1 卷积层(卷积核大小为 1×1, 维度为 256 维, 步长为 1), 将维度统一变为 256 维以减少运算量。M5 层由 C5 层经过上述 1×1 卷积层得到, M5 特征层进行 2 倍上采样后与经过 1×1 卷积层的 C4 层执行相加操作得到 M4 层, 用同样的方法依次可以得到 M3 层和 M2 层。M2、M3、M4 以及 M5 层特征图都分别经过一个 3×3 卷积层(卷积核大小为 3×3, 维

度为 256 维, 步长为 1), 作用是消除深层特征和浅层特征融合时产生的混叠效应, 然后依次可以得到 P2、P3、P4 以及 P5 层特征, P6 层特征由 P5 层经过一个 1×1 的最大池化层(卷积核大小为 1×1, 维度为 256 维, 步长为 2)得到。

最终得到的 P2、P3、P4、P5 以及 P6 这五个特征层分别用来预测不同尺度的目标, 为此分别设置了不同尺寸的锚框, 对应关系如表 1 所示, 锚框的尺寸设置五种(32², 64², 128², 256², 512²), 宽高比同为 (0.5, 1, 2)。

2.2.2 k-means 算法聚类生成锚框 在 Faster RCNN 模型中, 通过生成的锚框进行目标检测。在执行检测任务时, 模型检测的准确度和速度依赖于锚框的尺寸和数量, 所以应当根据不同数据集的特点合理设置锚框, 而不是简单地沿用原模型设置的锚框。在基于 ResNet50+FPN 的 Faster RCNN 模型中, 15 种不同尺寸的锚框都是基于人工经验设置的, 对于本文数据集中鳗鲡头部尺寸小的情况并不完全适用, 因此本文通过使用 k-means (Kanungo *et al*, 2004) 算法聚类生成适合鳗鲡数据集的锚框尺寸, 以提升模型对鳗鲡头部检测的准确度。

k-means 聚类算法的输入样本是鳗鲡数据集中真实标注框的宽、高, 在算法统计样本数据之前需要提前指定聚类簇个数 k (人为指定)。之后算法将相似度高的标注框分配至同一个指定簇中, 每个标注框只

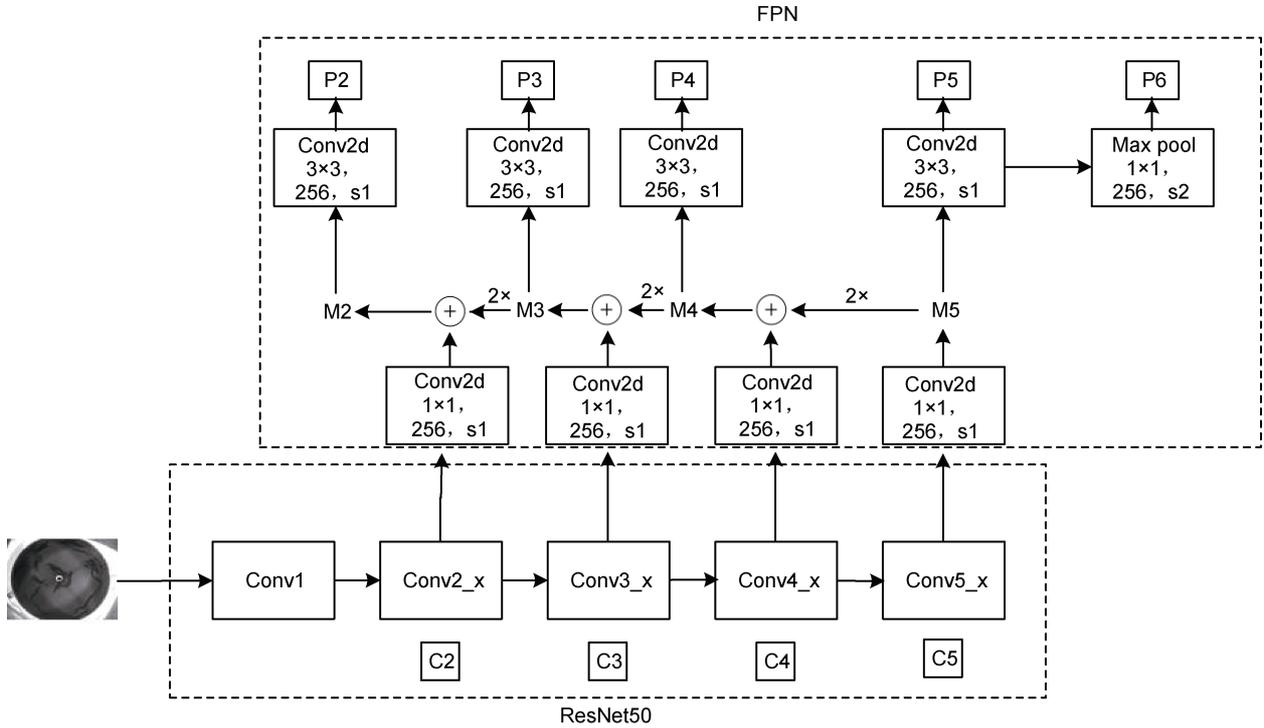


图 6 ResNet50 + FPN 网络
Fig.6 The ResNet50 + FPN model

表 1 锚框和特征层对应关系

Tab.1 Correspondence between anchors and characteristic layers

| 项目 | P2 特征层 | P3 特征层 | P4 特征层 | P5 特征层 | P6 特征层 |
|--------|-----------------|-----------------|------------------|------------------|------------------|
| 锚框的尺寸 | 32 ² | 64 ² | 128 ² | 256 ² | 512 ² |
| 锚框的宽高比 | (0.5, 1, 2) | (0.5, 1, 2) | (0.5, 1, 2) | (0.5, 1, 2) | (0.5, 1, 2) |

会分配至其中一个簇，标注框与簇中心的距离最小，算法会不断地迭代更新，直至聚类簇中心不再改变或达到所设定的迭代次数为止。通过计算 k 个簇中全部标注框宽、高的均值就能得到适合鳗鲡数据集的锚框尺寸。

在聚类过程中，本文使用的距离度量公式如式(1)所示：

$$d(\text{box}, \text{center}) = 1 - \text{IOU}(\text{box}, \text{center}), \quad (1)$$

式(1)中，box 代表真实标注框，center 代表聚类中心框，IOU(box, center)代表聚类中心框与真实标注框的交并比， $d(\text{box}, \text{center})$ 代表聚类中心框和真实标注框的距离。

值得注意的是，通常在模型的数据预处理阶段会限制输入网络中的图像大小，比如本文的模型中就在保持图像宽高比不变的情况下将输入图像的最

大边长缩放至 1 000 像素，同时图像中的标注框也会进行缩放，因此在聚类前需要使用同样的处理方式去缩放鳗鲡标注框，这样聚类获得的锚框才是实际网络训练时匹配的锚框。

取聚类簇个数 $k=15$ 对鳗鲡标注数据进行聚类，同时统计所有聚类中心框的宽高及其比例，结果如图 7 和表 2 所示。根据聚类结果将锚框尺寸重新设置为(14², 18², 24², 28², 32²)，宽高比设置为{0.6, 1.2, 1.8}。

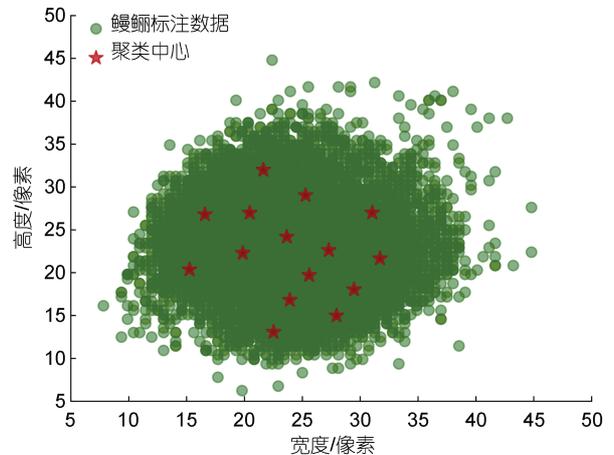


图 7 鳗鲡标注数据聚类分析
Fig.7 Clustering analysis of eel labeling data

表 2 聚类中心的宽高及其比例

Tab.2 Width and height and their ratio in the cluster center

| 编号 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 宽度 | 22 | 15 | 23 | 27 | 16 | 20 | 25 | 29 | 19 | 24 | 27 | 21 | 32 | 25 | 30 |
| 高度 | 13 | 20 | 16 | 14 | 26 | 22 | 19 | 17 | 26 | 24 | 22 | 31 | 21 | 30 | 27 |
| 宽高比 | 1.7 | 0.8 | 1.4 | 1.9 | 0.6 | 0.9 | 1.3 | 1.7 | 0.7 | 1.0 | 1.2 | 0.7 | 1.5 | 0.8 | 1.1 |

注: 表中宽度和高度的单位均为像素

2.2.3 Soft-NMS 算法 在目标检测模型中 NMS 算法发挥着重要作用, 它能够有效去除大量重复的检测框, 从而降低模型的误检率。算法的主要流程是将检测框按照置信度大小进行排序, 筛选出置信度最高的框, 之后计算其余所有检测框和该框的重叠度, 倘若重叠度的值大于或等于所设置的阈值, 就将此检测框的置信度分数置为 0。NMS 算法的分数重置函数如公式(2)所示:

$$S_i = \begin{cases} S_i & \text{IOU}(M, b_i) < N_t \\ 0 & \text{IOU}(M, b_i) \geq N_t \end{cases} \quad (2)$$

式中, N_t 为人为设置阈值, 本文设置为 0.5; S_i 为第 i 个检测框对应的得分; M 为当前得分最高的检测框; b 为所有检测框的集合, b_i 为其中的第 i 个检测框; IOU 为 M 和 b_i 的重叠度。

从式(2)中可以看出, NMS 算法直接将得分最高的框 M 的重叠度大于或等于所设阈值的相邻检测框的分数置 0, 所以当图像中出现鳗鲡较为重叠的情况时就容易导致漏检目标。因此本文选择使用 Soft-NMS (Bodla *et al.*, 2017)算法来代替原 NMS 算法, 该算法会根据相邻检测框重叠度的大小对其分数进行衰减, 而不是简单地将其分数置 0。若检测框与最高得分框 M 的重叠度很大, 则这个检测框的分数会变得很低; 若检测框与最高得分框 M 的重叠度很小, 则这个检测框的分数基本不受影响。Soft-NMS 算法的分数重置函数如式(3)所示:

$$S_i = \begin{cases} S_i & \text{IOU}(M, b_i) < N_t \\ S_i(1 - \text{IOU}(M, b_i)) & \text{IOU}(M, b_i) \geq N_t \end{cases} \quad (3)$$

3 实验与结果分析

3.1 实验环境、参数及计数方法设置

本文实验基于 PyTorch 框架, 使用硬件环境配置为: 处理器为 Intel 酷睿 i9-9900X, 显卡为 NVIDIA RTX2080Ti, 显存为 11 G; 使用软件环境配置为: 系统为 Ubuntu18.04.1, 运行内存为 128 G, Python 版本为 3.8, CUDA 版本为 11.2。

实验图像均为个人采集, 采集地点为集美大学

水产试验场的鳗鲡循环水养殖车间。总共采集了 1 000 张鳗鲡图像, 通过 labelImg 软件将鳗鲡数据集标注为 VOC 数据格式, 并对数据集进行随机划分, 其中 70%用作训练, 30%用作测试。

在 Faster RCNN 模型训练阶段: 模型总训练迭代次数设置为 35 000 次; 模型优化器选择带动量的 SGD 算法, 其中动量因子参数设置为 0.9; 权重衰减参数设置为 0.000 5; 模型学习率参数初始化为 0.005, 并且模型每训练迭代 1 750 次就衰减为原来的 0.33 倍。

本文的鳗鲡计数方法参考自张银松等(2019)对昆虫的计数方法, 首先人为设定置信度阈值 N (本文设定为 0.7), 然后遍历图像当中所有的检测框, 当某一检测框的得分 S 大于阈值 N 时, 就认为检测到了目标, 此检测框参与统计计数; 当某一检测框的得分 S 小于或等于阈值 N 时, 就认为没有检测到目标, 此检测框不参与统计计数, 计数方法如式(4)所示:

$$S_i = \begin{cases} 1 & S_i > N \\ 0 & S_i \leq N \end{cases} \quad (4)$$

3.2 Faster RCNN 模型改进后的训练结果

改进后的 Faster RCNN 模型每迭代一次就记录一下模型总损失值, 将训练集数据输入 COCO 预训练模型进行训练, 模型训练时的总损失和学习率变化如图 8 所示。从图 8 中可知, 模型训练总迭代次数为 35 000 次, 训练过程中模型总损失下降很快, 在迭代 5 000 次之前有持续下降趋势, 超过 5 000 次之后基本趋于稳定, 证明模型已经达到收敛。

模型每训练一个 epoch (训练集中所有图像都训练过一次)就通过测试集进行评估, 记录下此时模型的 mAP 值(IOU 阈值为 0.5 时的平均精度均值)。模型

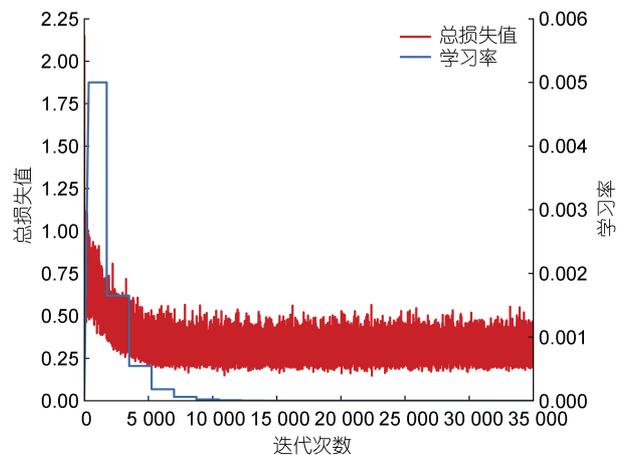


图 8 总损失和学习率变化图

Fig.8 The total loss and the variant of learning rate

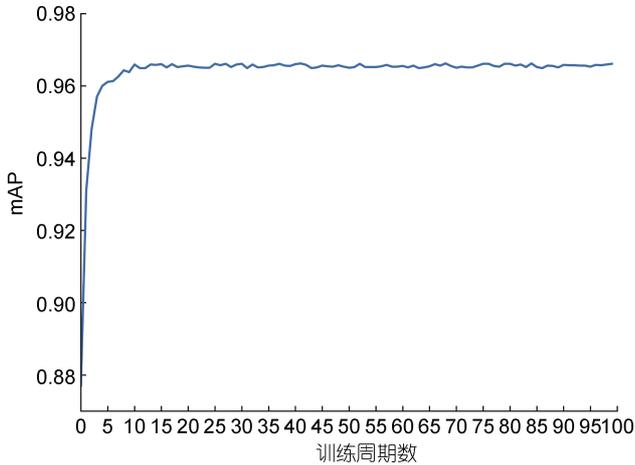


图9 测试集上 mAP 变化图

Fig.9 mAP changes on the test set

在训练阶段的 mAP 如图 9 所示。从图 9 中可知, 模型共训练了 100 个 epoch, 训练时 mAP 值上升很快, 在第 5 个 epoch 时模型检测精度就达到 96% 以上, 在第 10 个 epoch 之后模型总损失值虽然还略有下降, 但是 mAP 值已经基本趋于稳定。

3.3 模型检测结果比较与分析

使用不同的 Faster RCNN 改进模型在相同的鳊数据上进行实验, 同时在其测试集上比较各个模型的 mAP 值, 实验时的实验环境和所设置的参数均保持一致。使用不同改进策略后的模型的实验结果如表 3 所示。SSD 和 YOLO 都是目标检测领域中典型的单阶段检测模型, 具有检测速度快、精度高等优点, 本文使用改进后的 Faster RCNN 模型对比 SSD300 模型以及 YOLOV3 模型的实验结果如表 4 所示。

表 3 不同改进模型的性能比较

Tab.3 Comparison in performance of different models that are improved

| 模型 | 骨干网络 | mAP ^{0.5} /% | mAP ^{0.75} /% | mAP ^{medium} /% | 时间/s |
|----------------------------------|--------------|-----------------------|------------------------|--------------------------|-------|
| Faster RCNN | MobilenetV2 | 73.2 | 19.7 | 31.8 | 0.095 |
| Faster RCNN | ResNet50 | 82.5 | 29.4 | 39.2 | 0.112 |
| Faster RCNN | ResNet50+FPN | 95.5 | 64.2 | 58.1 | 0.123 |
| Faster RCNN +k-means | ResNet50+FPN | 96.5 | 66.7 | 59.4 | 0.123 |
| Faster RCNN+ k-means+Soft-NMS | ResNet50+FPN | 96.5 | 68.5 | 60.1 | 0.136 |

注: mAP^{0.5} 的 IOU 阈值为 0.5; mAP^{0.75} 的 IOU 阈值为 0.75; mAP^{medium} 的 IOU 阈值为 0.5:0.05:0.95, 检测面积限制在 32²~96² 像素之间。表 4 同

表 4 改进后的 Faster RCNN 模型与其他模型性能比较

Tab.4 The performance of the improved Faster RCNN model compared with other models

| 模型 | 骨干网络 | mAP ^{0.5} /% | mAP ^{0.75} /% | mAP ^{medium} /% | 时间/s |
|-------------------|--------------|-----------------------|------------------------|--------------------------|-------|
| SSD300 | ResNet50 | 71.6 | 21.7 | 32.3 | 0.035 |
| YOLOV3 | DarkNet53 | 81.5 | 27.5 | 38.4 | 0.082 |
| 改进 Faster RCNN 模型 | ResNet50+FPN | 96.5 | 68.5 | 60.1 | 0.136 |

从表 3 中可以看出, Backbone 为 MobilenetV2 (Sandler *et al*, 2018) 网络时, 模型的 mAP^{0.5} 为 73.2%, mAP^{0.75} 为 19.7%, mAP^{medium} 为 31.8%, 检测速度为 0.095 s/图。当 Backbone 更改为 ResNet50 网络时, 模型各 mAP 值都有较大的提升, 其中 mAP^{0.5} 提高了 9.3%, mAP^{0.75} 提高了 9.7%, mAP^{medium} 提高了 7.4%, 检测速度有所下降, 多耗时 0.017 s。笔者对 Backbone 为 VGG16 (Simonyan *et al*, 2015) 网络时也进行过实验, 结果发现模型各 mAP 值都几乎为 0, 分析原因可能是由于 VGG16 缺少像 ResNet50 (图 10) 中的瓶颈结构以及 MobilenetV2 中的倒残差结构 (图 11), 网络深层的输出中没有融合浅层特征, 损失了小目标的特

征, 因此使得模型的检测效果很差, 也反映出了浅层特征对于小目标检测非常重要。当 Backbone 为 ResNet50 结合 FPN 结构时, 模型的检测精度得到了极其显著的提升, 与 Backbone 为 ResNet50 模型相比, mAP^{0.5} 提升了 13%, mAP^{0.75} 提升了 34.8%, mAP^{medium} 提高了 18.9%, 检测速度仅多耗时 0.011 s。

从表 3 中还可以看出在 Backbone 为 ResNet50+FPN 网络基础上, 使用 k-means 算法聚类生成的锚框之后, 模型的 mAP^{0.5} 为 96.5%, mAP^{0.75} 为 66.7%, mAP^{medium} 为 59.4%, 检测速度为 0.123 s/图, 在保持检测速度不变的情况下提升了模型的检测精度, 其中 mAP^{0.5} 提升了 1%, mAP^{0.75} 提升了 2.5%, mAP^{medium}

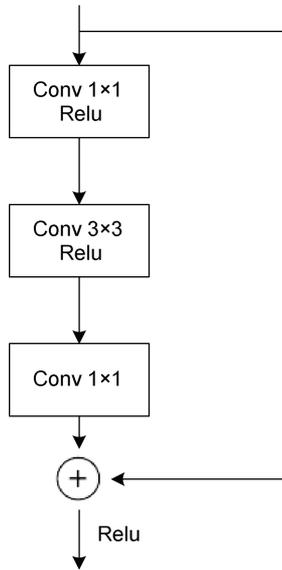


图 10 瓶颈结构

Fig.10 The bottleneck structure

提升了 1.3%。接着使用 Soft-NMS 算法之后, 模型的检测精度又有了进一步的提升, 其中 $mAP^{0.75}$ 提升了 1.8%, mAP^{medium} 提升了 0.7%, 检测速度多耗时 0.013 s。

从表 4 中可以看出改进后的 Faster RCNN 模型与 SSD300 模型和 YOLOV3 模型相比检测速度慢但是模型各检测精度均有显著提升, 其中 $mAP^{0.75}$ 的提升尤为显著。与 SSD300 模型相比, 改进后的模型 $mAP^{0.5}$ 提升了 24.9%, $mAP^{0.75}$ 提升了 46.8%, mAP^{medium} 提升了 27.8%。与 YOLOV3 模型比较, 改进后的模型 $mAP^{0.5}$ 提升了 15%, $mAP^{0.75}$ 提升了 41%, mAP^{medium} 提升了 21.7%。

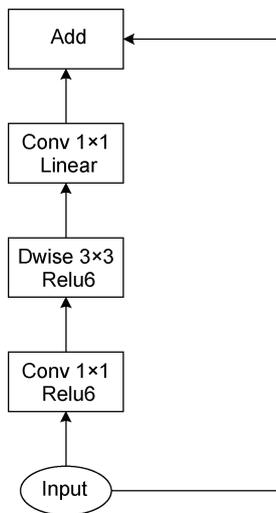


图 11 倒残差结构

Fig.11 The inverted residual structure

综上所述, 本文对 Faster RCNN 模型的不同改进策略均显著有效, 改进后的模型对鳗鲡头部识别能力更强、检测效果更好。

改进后的 Faster RCNN 模型对鳗鲡头部的检测结果如图 12 所示, 两张图片中鳗鲡的数量分别为 30 条和 40 条, 用来测试模型效果图片均来自鳗鲡数据集中的测试集。

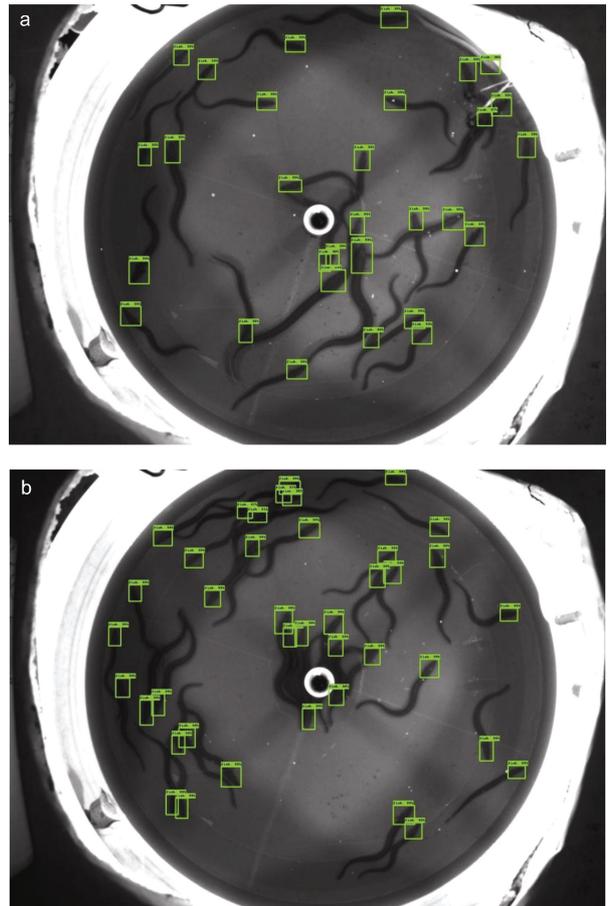


图 12 改进后的模型检测效果

Fig.12 The detection result of improved model

注: a. 30 条鱼; b. 40 条鱼

3.4 鳗鲡计数结果与分析

本文是基于模型的检测结果对图像中的鳗鲡进行计数的, 即模型的检测效果越好, 计数结果就越准确, 因此选择使用改进后的 Faster RCNN 模型即 Faster RCNN (ResNet50+FPN)+k-means+Soft-NMS 模型进行计数。

首先从测试集当中选取 9 张图像, 各图像中的鳗鲡数量都不同, 然后使用训练好的改进 Faster RCNN 模型进行检测计数, 最后将模型计数结果与图像中鳗鲡真实数量进行比较, 结果如表 5 所示。可以看出,

当图像中鳗鲡数量中较少时,模型计数结果准确无误差;随着鳗鲡数量的增加模型误差也在增大,其中最大误差为-4,表示模型漏计4条;从整体来看,模型的计数准确率都在90%以上,而模型的误差都表现为漏检。通过对模型检测后的图像进行分析后发现,漏检原因主要有三个:一是图像中鳗鲡较为密集时容易出现部分鳗鲡头部遮挡严重甚至完全遮挡的情况,导致模型漏检;二是模型未检测出鳗鲡头部,导致漏检。三是模型检测框中虽然包含鳗鲡头部,但是该检测框的置信度分数低于所设的计数阈值 N (0.7),此时该检测框不参与统计计数,从而导致模型漏检。因此未来的工作中将融合深度学习领域的分割算法如Unet (Ronneberger *et al.*, 2015)、Mask RCNN (He *et al.*, 2017)等辅助计数,还将扩充鳗鲡数据集,以提高模型的泛化能力和检测精度。

表5 鳗鲡计数结果
Tab.5 The eel counting results

| 图片 | 真实值 | 改进 Faster RCNN 模型计数 | 改进 Faster RCNN 模型误差 |
|----|-----|------------------------|------------------------|
| 1 | 13 | 13 | 0 |
| 2 | 15 | 15 | 0 |
| 3 | 20 | 19 | -1 |
| 4 | 25 | 24 | -1 |
| 5 | 31 | 28 | -3 |
| 6 | 36 | 36 | 0 |
| 7 | 40 | 37 | -3 |
| 8 | 41 | 38 | -3 |
| 9 | 45 | 41 | -4 |

注:表中表示鳗鲡数量的单位为条,模型误差为检测值与真实值的差值

4 结论

本文将目标检测模型 Faster RCNN 应用于循环水养殖鳗鲡的检测和计数,针对原 Faster RCNN 模型在鳗鲡数据集上存在的弊端,对模型做出了改进。其中针对模型对小尺度目标检测效果差的问题,用 ResNet50+FPN 网络作为模型的骨干网络,提取并融合了多尺度的特征,以提升模型对鳗鲡头部的检测能力;接着用 k -means 算法聚类生成适合鳗鲡数据集的15种不同尺度的锚框,进一步提升模型的检测能力;针对图像中存在鳗鲡头部重叠的问题,选择使用 Soft-NMS 算法替代原 NMS 算法对 RPN 部分生成的候选框进行筛选,以减少模型对鳗鲡重叠部分的漏检情况。

实验结果表明,本文对 Faster RCNN 模型的各改进策略均显著有效。最终改进后的 Faster RCNN 模型的 $mAP^{0.5}$ 、 $mAP^{0.75}$ 、 mAP^{medium} 分别为 96.5%、68.5%、60.1%,相较于原模型(Backbone 为 ResNet50) $mAP^{0.5}$ 提升了 14%, $mAP^{0.75}$ 提升了 39.1%, mAP^{medium} 提升了 20.9%。此外,与 SSD300 和 YOLOV3 模型相比,改进后的 Faster RCNN 模型各检测精度均有显著提升,其中在 $mAP^{0.75}$ 上的提升尤为显著,较 SSD300 提升了 46.8%,较 YOLOV3 提升了 41%。在鳗鲡计数上,利用改进后的 Faster RCNN 模型的检测结果进行计数,计数准确率达到 90%以上。综上,本文改进后的 Faster RCNN 模型在循环水养殖鳗鲡计数中有重要应用价值。

参 考 文 献

- 王文静,徐建瑜,杜秋菊,2016. 基于计算机视觉的鱼苗自动计数系统研究[J]. 渔业现代化,43(3): 34-38, 73.
- 王宝航,1988. 多通道式鱼苗计数器概论[J]. 渔业机械仪器(1): 26-28.
- 王硕,范良忠,刘鹰,2015. 基于计算机视觉的大菱鲆鱼苗计数方法研究[J]. 渔业现代化,42(1): 16-19.
- 叶翠,钟传明,池宝兴,2019. 福建鳗鲡产业发展形势分析[J]. 中国水产(4): 56-61.
- 李伯珩,邵诗国,尹文庆,等,1991. 鱼苗计数器的试验研究[J]. 南京农业大学学报,14(4): 116-120.
- 汪梦婷,袁飞,程恩,2020. 鱼类目标的密度估计模型[J]. 哈尔滨工程大学学报,41(10): 1545-1552.
- 张银松,赵银娣,袁慕策,2019. 基于改进 Faster-RCNN 模型的粘虫板图像昆虫识别与计数[J]. 中国农业大学学报,24(5): 115-122.
- 张康德,1997. 鱼(虾)苗光电计数器[J]. 企业技术开发(12): 13-14.
- 张婷婷,赵峰,张涛,等,2019. 中国鳗鱼产业发展及其资源保护建议[J]. 渔业信息与战略,34(4): 235-243.
- 陈学豪,2003. 鳗鲡养殖技术之一: 美洲鳗鲡的生物学特性及养殖技术特点[J]. 中国水产(6): 38-39.
- 段青玲,刘怡然,张璐,等,2018. 水产养殖大数据技术研究进展与发展趋势分析[J]. 农业机械学报,49(6): 1-16.
- 黄玲,胡波,曹乃文,2012. 基于图像处理的鱼苗计数方法[J]. 湖北农业科学,51(9): 1880-1882.
- 崔榛,2020. 基于计算机视觉的红鳍东方鲀计数系统[D]. 大连: 大连海洋大学: 31-32.
- BODLA N, SINGH B, CHELLAPPA R, *et al.*, 2017. Soft-NMS—Improving object detection with one line of code [C] // Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 5562-5570.
- GIRSHICK R, 2015. Fast R-CNN [C] // Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE: 1440-1448.
- GIRSHICK R, DONAHUE J, DARRELL T, *et al.*, 2014. Rich feature hierarchies for accurate object detection and

- semantic segmentation [C] // Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE: 580-587.
- HE K M, GKIOXARI G, DOLLÁR P, *et al*, 2017. Mask R-CNN [C] // Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 2980-2988.
- HE K M, ZHANG X Y, REN S Q, *et al*, 2016. Deep residual learning for image recognition [C] // Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE: 770-778.
- KANUNGO T, MOUNT D M, NETANYAHU N S, *et al*, 2004. A local search approximation algorithm for k -means clustering [J]. *Computational Geometry*, 28(2/3): 89-112.
- LABAO A B, NAVAL P C JR, 2019. Cascaded deep network systems with linked ensemble components for underwater fish detection in the wild [J]. *Ecological Informatics*, 52: 103-121.
- LIN T Y, DOLLÁR P, GIRSHICK R, *et al*, 2017. Feature pyramid networks for object detection [C] // Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE: 936-944.
- LIU W, ANGUELOV D, ERHAN D, *et al*, 2016. SSD: Single shot MultiBox detector [C] // 14th European Conference on Computer Vision. Amsterdam, the Netherlands: Springer: 21-37.
- MANDAL R, CONNOLLY R M, SCHLACHER T A, *et al*, 2018. Assessing fish abundance from underwater video using deep neural networks [C] // 2018 International Joint Conference on Neural Networks (IJCNN). Rio de Janeiro, Brazil: IEEE: 1-6.
- REDMON J, DIVVALA S, GIRSHICK R, *et al*, 2016. You only look once: Unified, real-time object detection [C] // Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE: 779-788.
- REN S Q, HE K M, GIRSHICK R, *et al*, 2015. Faster R-CNN: Towards real-time object detection with region proposal networks [C] // Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, Canada: MIT Press: 91-99.
- RONNEBERGER O, FISCHER P, BROX T, 2015. U-Net: Convolutional networks for biomedical image segmentation [C] // 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany: Springer: 234-241.
- SANDLER M, HOWARD A, ZHU M L, *et al*, 2018. MobileNetV2: Inverted residuals and linear bottlenecks [C] // Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE: 4510-4520.
- SIMONYAN K, ZISSERMAN A, 2015. Very deep convolutional networks for large-scale image recognition [Z]. arXiv:1409.1556.
- ZHANG J F, PANG H T, CAI W M, *et al*, 2021. Using image processing technology to create a novel fry counting algorithm [J]. *Aquaculture and Fisheries*, <https://doi.org/10.1016/j.aaf.2020.11.004>.

AUTO-COUNTING THE EEL *ANGUILLA* IN RECIRCULATING AQUACULTURE SYSTEM VIA DEEP LEARNING

LI Kai^{1,2}, JIANG Xing-Long^{1,2}, CHEN Er-Kang^{2,3}, CHEN Peng^{2,3}, XU Zhi-Yang^{1,2}, LIN Qian^{1,2}

(1. Fisheries College, Jimei University, Xiamen 361021, China; 2. Engineering Research Center of the Modern Technology for Eel Industry, Ministry of Education, Xiamen 361021, China; 3. School of Ocean Information Engineering, Jimei University, Xiamen 361021, China)

Abstract Eel is an economically valuable species in China. To realize accurate and efficient auto-counting of eel cultured in recirculating aquaculture system, an improved Faster RCNN model by deep learning was proposed. Because the head size of eel is small, how to recognize eel individual correctly became an issue of this study. To solve the issue, the FPN structure was added to the feature extraction network ResNet50 as the backbone network of model to extract and fuse multi-scale features. However, the original model anchors are based on artificial experience, which is not suitable for eel dataset process. Therefore, the eel-head detection boxes marked in the training dataset were clustered and analyzed in *k*-means clustering algorithm, and 15 different scales of anchors suitable for the eel dataset were obtained. In addition, to avoid eel head overlap in the images, Soft-NMS algorithm instead of the original NMS algorithm was applied to select candidate boxes generated by RPN to reduce the missed detection of eel due to the image overlap. Results show that the detection accuracy of the improved Faster RCNN model for eel-head counting reached 96.5%, which is 14%, 24.9%, and 15% higher than that of the original Faster RCNN (the backbone of ResNet50), SSD300, and YOLOV3 models, respectively. With the improved Faster RCNN model, the correct counting accuracy could reach >90%, by which the eel detection and identification in recirculating aquaculture system could be enhanced considerably.

Key words eel counting; deep learning; Faster RCNN; FPN; *k*-means; Soft-NMS