

基于 YOLOv3 的水下小目标自主识别

袁利毫, 咎英飞, 钟声华, 祝海涛

(哈尔滨工程大学 船舶工程学院, 黑龙江哈尔滨 150001)

摘要 针对智能水下机器人作业时小目标自主识别的需求, 提出基于深度学习的 YOLOv3 算法, 通过对水下机器人实采数据进行神经网络权重训练, 实现对水下小目标物快速、精确的识别与分类, 从而解决在复杂的水下地形和未知作业环境中对水下目标识别问题。并分析算法学习率在水下海珍品数据集上对损失函数值的影响。实验结果表明, 基于 YOLOv3 算法的水下海珍品的目标检测具有强实时性与高准确率, 所有目标类别查准率高达 99%, 物体的查全率在 90% 以上, 可达 35 帧/秒的检测速率; 在网络训练过程中调整学习率有利于加速并降低损失函数值。

关键词 水下机器人; 深度学习; YOLOv3 算法; 水下小目标识别

中图分类号: U445.54+3

文献标志码: A

Small Underwater Target Recognition Based on YOLOv3

YUAN Li-hao, ZAN Ying-fei, ZHONG Sheng-hua, ZHU Hai-tao

(College of Shipbuilding Engineering, Harbin Engineering University, Harbin, Heilongjiang 150001, China)

Abstract For the demand of autonomous recognition for small objects in subsea operations by remote operated vehicles (ROVs), a YOLOv3 algorithm based on deep learning is proposed. ROVs are trained according to neural network weight of the data collected in ROV fetching competition, in order to identify and classify small underwater objects fast and precisely, which will solve the underwater-object-recognition problem in complex and unknown subsea operation environment. The effect of algorithm learning rate on loss functions in the seafood data set is analyzed. The results show that underwater object recognition and classification of seafood based on YOLOv3 algorithm is of high real time as well as precision. The precision ratio of all objects classified reaches up to ninety-nine percent, and the recall ratio is above ninety percent. The estimating speed can reach up to 35 frames per second (FPS). It also shows that adjusting learning rate during network training process can help decrease loss functions.

Key words remote operated vehicles (ROVs); deep learning; YOLOv3 algorithm; small underwater object recognition

0 引言

水下小目标检测是水下机器人作业智能化作业的关键问题。水下机器人进行水下目标识别的传统方法是依靠机器视觉与水下声呐, 普遍存在识别效率低, 误差大等问题。随着计算机技术不断发展, 深度学习成为当前研究目标检测与识别的热门领域。丁雪妍^[1]等人基于卷积神经网络研究水下图像增强算法, 用于提高水下目标识别的准确率; 赵增科^[2]等人提出一种新的深度神经网络 DCNN

用于水下目标识别。国外研究方面, Dirk^[3]等人开发了一套能够对水下疑似目标进行检测和标注的系统, 该系统可以去除背景对目标进行识别和跟踪; Medhdi^[4]等人提出一种基于纹理特征的水下电缆识别方法, 通过提取图像边缘信息, 从而找到目标物。

本文提出将 YOLOv3 目标检测算法应用于水下小目标检测, YOLOv3 是 Redmon J^[5]等人提出 CNN(卷积神经网络)的一种算法, YOLOv3 出色的性能使得它广泛应用在各行各业, 例如车辆实时检

基金项目: 国家重点研发计划(2018YFC0309400)

作者简介: 袁利毫(1982—), 男, 博士, 副教授, 主要从事大规模流场建模与可视化仿真技术研究。

测,行人检测与定位,视频内容检测等。本文使用采集的水下海珍品影像资料对YOLOv3深度神经网络进行权值训练。通过调整算法的超参数获得了不同的性能表现。实验结果表明,算法不仅满足水下作业要求实时性与准确率,并且海珍品目标识别率在90%以上。

1 基本理论

1.1 YOLOv3 算法

水下机器人智能作业对于目标检测的准确率和实时性有较高要求,常规的检测方法是基于图片检测,虽然准确率可以达到要求,但是检测速度较慢。本文采用的YOLOv3算法与其他算法相比具有强实时性和高准确率,其检测目标的流程如图1所示^[6]:

YOLOv3是YOLO系列最新的算法,相比早期

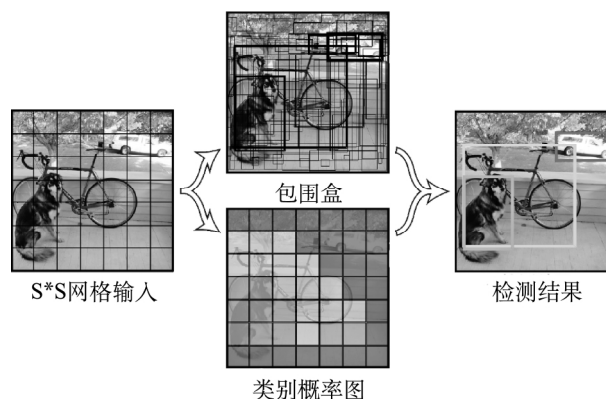


图1 YOLOv3 算法步骤

YOLOv2 算法^[5],YOLOv3 使用一个新的网络Darknet-53进行特征提取。Darknet-53融合了YOLOv2、Darknet-19以及其他新型残差网络,由 3×3 和 1×1 的卷积层组成。算法的网络结构如图2所示。

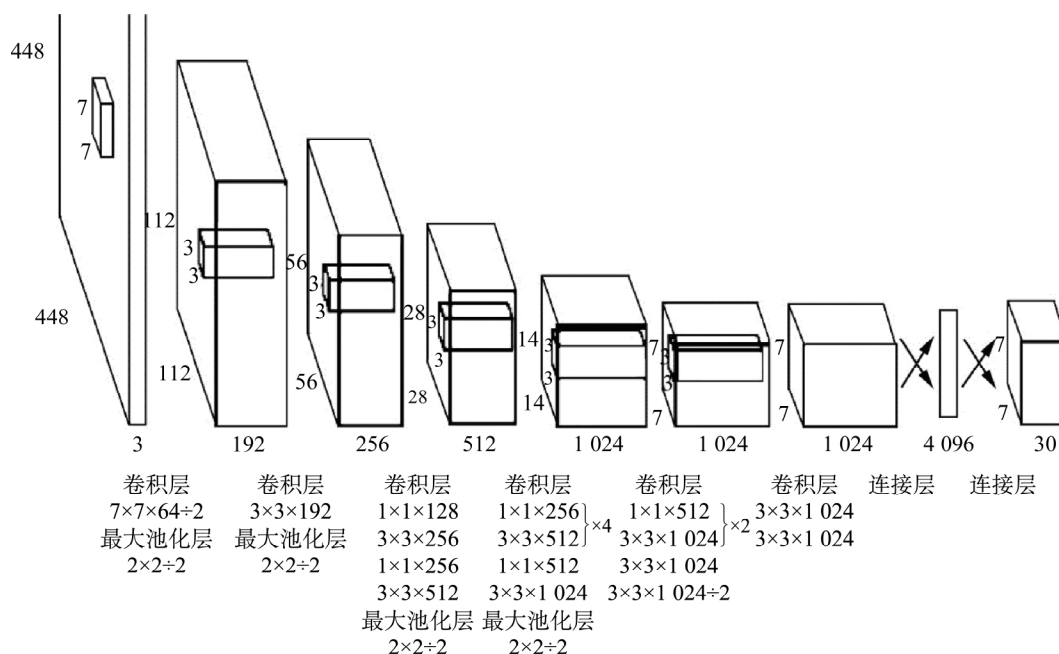


图2 YOLO 算法使用的卷积网络

Darknet-53网络在性能上要优于Darknet-19,在检测效率上同样优于ResNet-101和RetNet-152。如表1所示为Darknet-53在ImageNet数据上的测试结果^[7-8]。

YOLOv3算法将候选框提取、特征提取、目标分类等功能统一在一个深度神经网络中。YOLOv3在训练的过程中将图片划分为 $S \times S$ 的网格,如果标注物体的中心落入网格,那么该网格就负责检测

表1 ImageNet数据集测试结果

网络类型	正确率/%	正确率/%	每秒识别帧数/ $f \cdot s^{-1}$
Darknet-19	74.1	91.8	171
ResNet-101	77.1	93.7	53
ResNet-152	77.6	93.8	37
Darknet-53	77.2	93.8	78

物体。网格检测 n 个边界框及其置信度,以及 m 个类别概率。边界框的宽度和高度是以被检测物体的中心位置相对网格位置的偏移,网格的坐标值均被归一化。其中,置信度反映是否包含物体以及包含物体情况下位置的准确性,定义为:

$$P_{r(Object)} \times IOU_{pred}^{truth}, P_{r(Object)} \in [0, 1] \quad (1)$$

式中, $P_{r(Object)}$ 表示检测边界框包含目标物体的概率, IOU_{pred}^{truth} 为两个边界框的重叠度。最后计算每一个网格预测的类别概率 $C_i = P(Class_i | Object)$ 。

在网络训练误差中,不同大小的边界框权值相同。另外,误差的度量方式应该反映出不同边界框的小偏差相同。为了解决这个问题,回归的是边界框宽度和高度的平方根,而不是直接的宽度和高度。YOLOv3 在每个网格中预测出多个边界框。在训练网络时,对每个类别,只需要用一个边界框来负责该类别的回归。YOLOv3 的损失函数如下式:

$$\begin{aligned} Cost = & \varphi_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} [(x_i - \hat{x}_l)^2 + (y_i - \hat{y}_l)^2] \\ & + \varphi_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} [(\sqrt{w_i} - \sqrt{\hat{w}_l})^2 \\ & + (\sqrt{h_i} - \sqrt{\hat{h}_l})^2] + \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} (C_i - \hat{C}_l)^2 \\ & + \sum_{i=0}^{S^2} I_i^{obj} \sum_{c \in Classes} (P_i(c) - \hat{p}_l(c))^2 \end{aligned} \quad (2)$$

式中, I_{ij}^{obj} 表示目标出现在第 i 个网格单元; I_{ij}^{obj} 表示第 i 个网格单元中的第 j 个包围盒预测器; φ_{coord} 表示坐标误差权值,取 $\varphi_{coord} = 5$; φ_{noobj} 表示分类错误权值,取 $\varphi_{noobj} = 0.5$ 。

2 深度神经网络权重训练

2.1 数据集

本文从 2017 年獐子岛水下机器人目标抓取大

赛的影像资料中筛选 20 000 张海珍品照片。在训练集中,总共标记出 42 384 个正样本,分为海参、海胆、扇贝三个类别,如图 3(a)~(c)。同时还标记出了 5 421 个负样本,分为石头、海草、鱼三个类别,如图 3(d)~(f)。即训练集总共类别数为 5,再选取 2 000 张照片进行数据标注作为测试集。

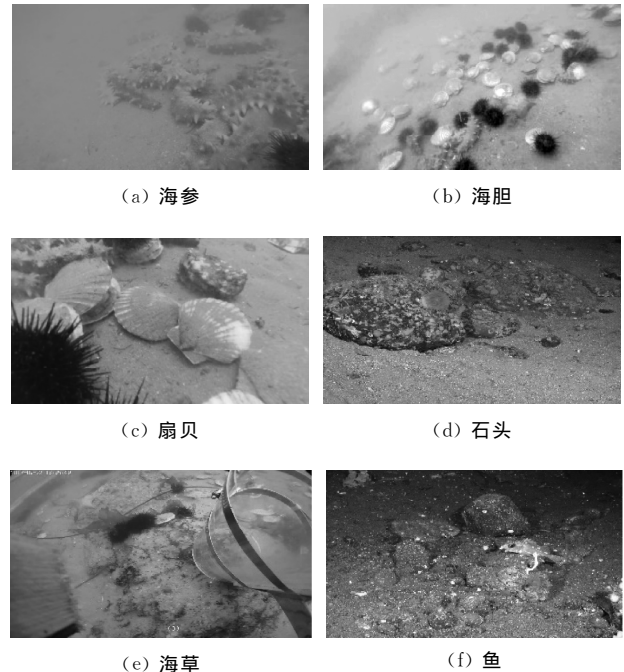


图 3 数据集分类

2.2 实验平台

本文使用 Darknet 框架中的 YOLOv3 算法进行实验,进行目标训练与检测的计算机配置为 CPU E5; GPU(GTX 1080Ti*2)和 64 G 内存。

2.3 网络训练过程

本文采用小批量梯度下降法和冲量在训练过程更新网络梯度信息。冲量有利训练过程加速收敛。在算法中对误差进行反向传播并更新参数,直到到损失函数逐渐收敛,算法训练需要设置的超参数初值如表 2 所示。

表 2 YOLOv3 算法参数

学习率	图片宽度	图片高度	缩放因子	最大迭代次数	图片数量	权值衰减
0.000 1	412	412	0.5	50 200	64	0.000 5

在算法训练初期,将学习率设置为 10^{-3} ,进行 1 000 次训练迭代。然后,修改学习率为 10^{-4} ,进行

500 次训练迭代。再将学习率修改为 10^{-5} 进行训练。利用在训练过程中动态调整学习率的方法,降

低并加速收敛损失函数值。最终,算法输出一个神经网络权重文件用于目标检测。

3 结果分析

本节分析学习率对算法损失函数值的影响,分别将设置学习率为 10^{-3} 与 10^{-4} ,分析网络训练过程中的损失函数值。图 4 是学习率恒为 10^{-3} 时训练过程中平均损失函数值与损失函数值历时曲线图。由图 4(a)可知,平均损失函数值在 0 到 2×10^4 次训练迭代期间呈逐渐下降趋势,在 2×10^4 次训练迭代后收敛至 $[3, 3.5]$ 区间。由图 4(b)可知算法的损失函数值总体呈下降趋势,最终在 $[2, 4.5]$ 区间振荡。

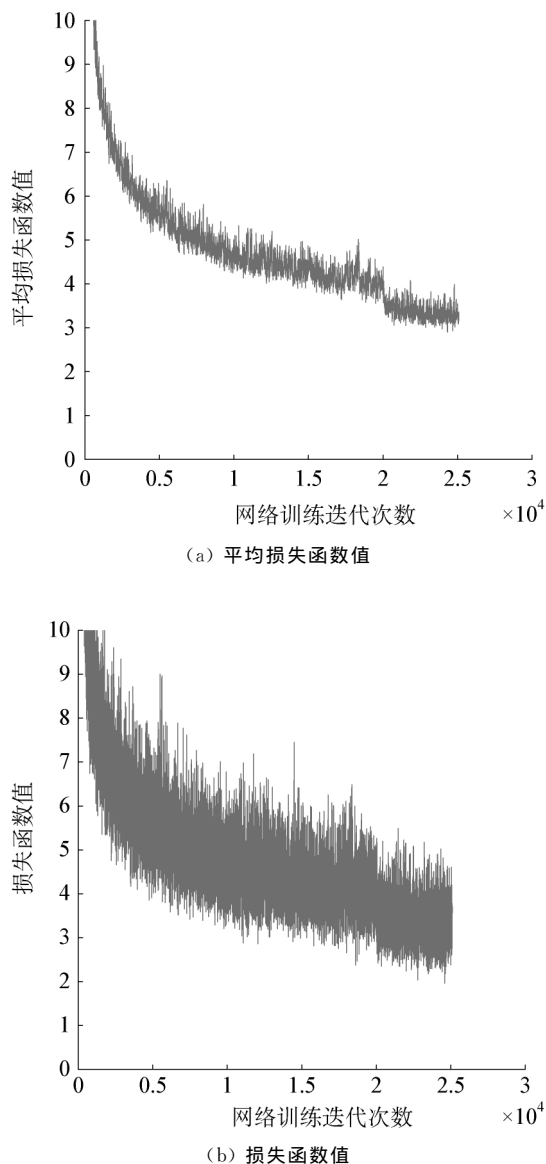


图 4 学习率 10^{-3} 的损失函数值历时图

图 5 是学习率恒为 10^{-4} 时训练过程中平均损失函数值与损失函数值的历时曲线图。由图可知,平均损失函数值与损失函数值的曲线整体呈下降趋势,在收敛过程中曲线振荡的幅度均要小于学习率恒为 10^{-3} 时。

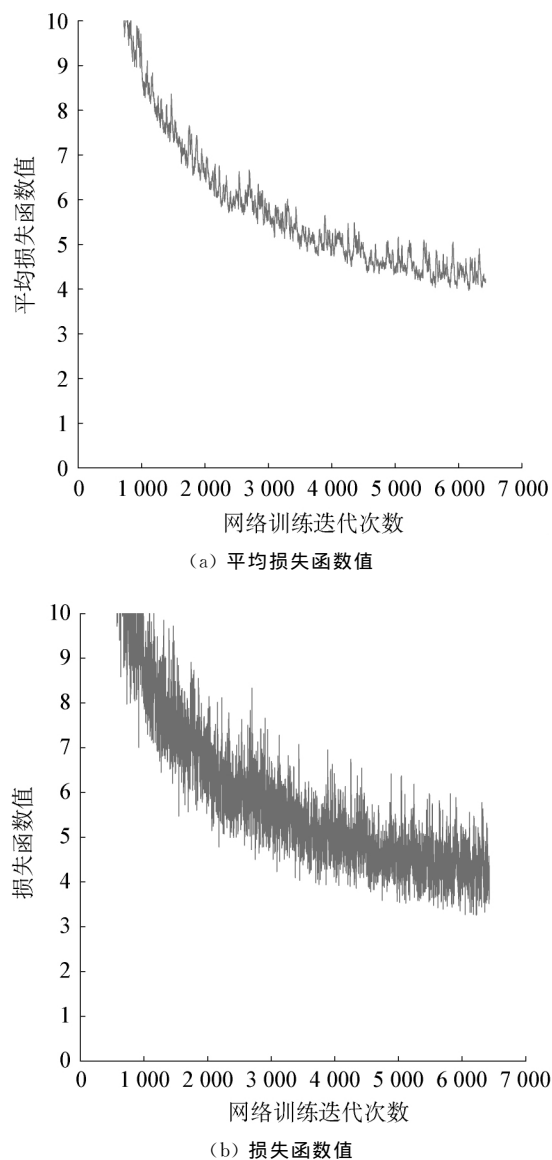
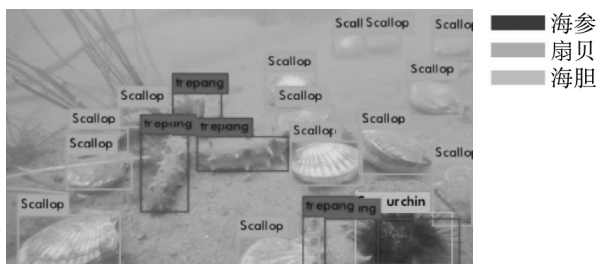
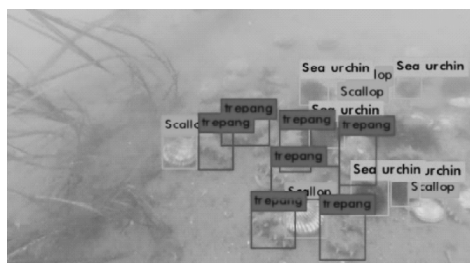


图 5 学习率 10^{-4} 的损失函数值历时图

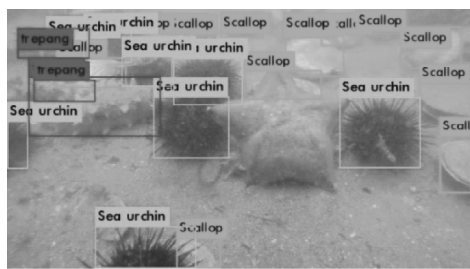
考虑到损失函数值振荡区间较小的因素,采用学习率恒为 10^{-4} 时输出的神经网络权重进行目标检测。我们将测试集数据分为目标稀疏与目标密集进行分析。图 6 为目标稀疏检测结果。图 6(a)中包含海参、海胆、扇贝三种目标,其中,个别目标被部分遮挡。算法成功检测并分类所有包含目标;图 6(b)中的部分目标物体的检测边界框重叠度较高,但算法仍能准确的识别并分类每个目标物。



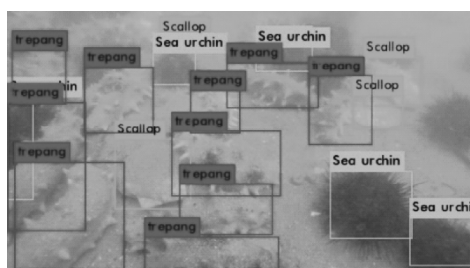
(a) 多目标稀疏检测结果



(b) 多目标稀疏检测结果



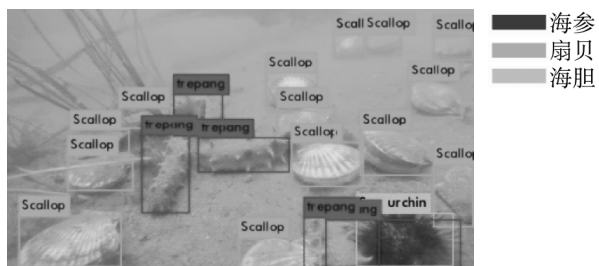
(c) 多目标稀疏检测结果



(d) 多目标稀疏检测结果

图 6 YOLOv3 稀疏目标检测结果

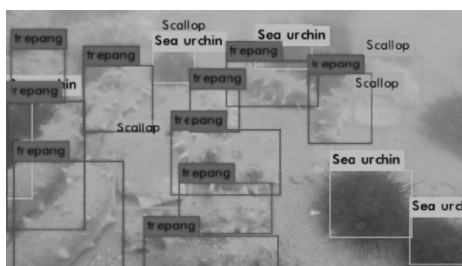
图 7 为密集目标检测结果,由图可知图片中大量目标物体的检测边界框重合度较高。在图 7(b)中部分目标未检测到,但是图内的物体的查全率仍



(a) 多目标稀疏检测结果



(c) 多目标稀疏检测结果



(d) 多目标稀疏检测结果



图 7 YOLOv3 密集目标检测结果

(b) 多目标稀疏检测结果

在 90% 以上。

通过实验结果得到测试集的正样本查准率、查全率以及每秒所能识别的图片帧数如表 3 所示:

表 3 测试集检测结果

类别	查准率/%	查全率/%	每秒识别帧数/ $f \cdot s^{-1}$
海参	99.9	91	35
海胆	99.9	92	35
扇贝	99.9	98	35

4 结 语

本文应用 YOLOv3 算法对采集的海珍品影像资料训练,得到可以快速、精准进行海珍品检测与分类的神经网络权重。验证了基于深度学习进行水下目标检测的可能性。实验结果表明,所有目标类别的查准率高达 99%,查全率均高于 90%,具有 35 帧/秒的检测速率。即使在目标物体相互重叠度较高时,算法也能稳定的进行检测与分类。最终对比了在海珍品数据集上不同学习率对损失函数值的影响,发现较小的学习率更利于损失函数值稳定

的收敛,且振荡区间较小。

参 考 文 献

- [1] 丁雪妍. 基于卷积神经网络的水下图像增强算法研究[D]. 大连: 大连海事大学, 2018.
- [2] 赵增科. 基于深度学习的水下目标识别[D]. 哈尔滨: 哈尔滨工程大学, 2017.
- [3] Walter D, Edgington D R, Koch C. Detection and tracking of object in underwater video [C]. IEEE Computer Society Conference on Computer Vision & Pattern Recognition, 2004: 544.
- [4] Fatan M, Daliri M R, Shahri A M. Underwater cable detection in the images using edge classification based on texture information [J]. Measurement, 2016, 91: 309.
- [5] Redon J, Farhadi A. YOLO9000: better faster stronger [C]. Computer Vision and Pattern Recognition, Honolulu, Hawaii, July, 2016: 6517.
- [6] Redon J, Farhadi A. YOLOv3: an incremental improvement [DB/OL]. [2018-04-08]. <https://arxiv.org/pdf/1804.02767.pdf>.
- [7] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Deep residual learning for image recognition [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770.
- [8] Liu Wei, Anguelov D, Erhan D, et al. SSD: single shot multibox detector [C]. European Conference on Computer Vision, 2016: 21.