



武汉大学学报(信息科学版)

Geomatics and Information Science of Wuhan University

ISSN 1671-8860, CN 42-1676/TN

《武汉大学学报(信息科学版)》网络首发论文

题目: 侧扫声纳检测沉船目标的改进 YOLOv5 法
作者: 汤寓麟, 边少锋, 翟国君, 刘敏, 张卫东
DOI: 10.13203/j.whugis20210353
网络首发日期: 2021-09-01
引用格式: 汤寓麟, 边少锋, 翟国君, 刘敏, 张卫东. 侧扫声纳检测沉船目标的改进 YOLOv5 法. 武汉大学学报(信息科学版). <https://doi.org/10.13203/j.whugis20210353>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

侧扫声纳检测沉船目标的改进 YOLOv5 法

汤寓麟¹, 边少锋¹, 翟国君², 刘敏³, 张卫东⁴

¹ 海军工程大学电气工程学院, 湖北 武汉, 430033

² 海军海洋测绘研究所, 天津, 300061

³ 91001 部队, 北京, 100841

⁴ 军委联合参谋部战场环境体系论证中心, 北京, 100088

摘要:针对基于 YOLOv3 模型的侧扫声纳沉船目标检测方法存在小目标漏警率高、模型权重大、检测速度未能满足实时性需求等问题, 引入 YOLOv5 算法并根据数据集特点, 提出基于 YOLOv5 模型的侧扫声纳海底沉船目标检测方法。在 YOLOv5 基础框架下尝试 YOLOv5a、YOLOv5b、YOLOv5c、YOLOv5d、YOLOv5s、YOLOv5m、YOLOv5l 和 YOLOv5x 八种不同深度和宽度的模型结构, 并选择最优的结构, 使用 GA+K (Genetic Algorithm+K-mean) 算法优化检测框, 通过 CIOU_Loss 对损失函数进行改进。实验结果表明, 改进的 YOLOv5a 模型在 AP_{0.5} 和 AP_{0.5-0.9} 较原始模型提高了 0.3% 和 0.6%, 较 YOLOv3 模型有了全面大幅提升, 其中 AP_{0.5} 和 AP_{0.5-0.9} 分别提高了 4.2% 和 6.1%, 检测速度达到 426 帧/秒, 提升了几乎一倍, 更加益于实际应用和工程部署。

关键词: 侧扫声纳沉船目标; 检测方法; YOLOv5 模型; CIOU_Loss; Genetic Algorithm+K-mean 算法

侧扫声纳对海底沉船的识别和探测是海上搜救和障碍物核查的重要组成部分^[1-5]。传统的识别和检测方法包括: 人工判读和模式识别。众所周知, 人工判读存在效率低、耗时长、主观不确定性大和经验依赖性强等问题, 而模式识别通过人工提取灰度、纹理、边缘和形状等特征, 经过图像和数据处理后设计分类器进行判别^[6-12], 文献[13]研究了基于灰度直方图和几何特征的声纳图像目标识别方法; 文献[14]通过属性直方图提取目标声影特征, 再使用模糊聚类和 BP 神经网络对目标进行识别; 文献[15]将水平集、不变矩、SVM 等方法组合用于水下目标识别; 文献[16]使用简单的 BP 神经网络对水下图像目标进行分类识别, 人工选取特征后送入神经网络进行分类训练; 文献[17]提出一种基于 DBN 结构用于水下目标识别的深度学习框架; 文献[18]提出基于 SVM 算法和 GLCM 的侧扫声纳影像分类研究; 尽管这些方式能很好的适用于简单或者特征明显的目标判别, 但是针对复杂目标时受声纳图像质量、特征提取算法模型的针对性等影响, 存在特征设计难度大、运算大量冗余以及泛化性和鲁棒性差等问题。

近年来, 随着计算机视觉领域的蓬勃发展, 基于深度学习的侧扫声纳图像海底沉船目标识别与检测相关技术研究正在开展, 尤其是以卷积神经网络代替人工判读和模式识别。图像识别解决了图像中目标的有无问题, 文献[19]在与经典机器学习 SVM 算法对比后提出以改进的 VGG-16 为框架的卷积神经网络迁移学习识别方法, 完成了侧扫声纳海底沉船的自动识别并取得明显优于传统方式的精度和效率; 目标检测在图像识别的基础上完成了目标定位的任务, 文献[20]利用 Faster R-CNN 网络实现了侧扫声纳海底沉船目标的自动检测。但是针对该模型存在结构复杂、训练和检测效率低等问题, 文献[21]又提出了基于迁移学习的改进 YOLOv3 模型的侧扫声纳沉船目标检测方法, 虽然在一定程度上提高了训练和检测效率, 但是仍存在小目标漏警率高, 检测速度无法满足实时性要求等问题。

针对 YOLOv3 存在的问题以及面向工程应用的现实需求, 本文拟构建基于 YOLOv5 模型的侧扫声纳海底沉船自动检测方法, 通过尝试 8 种不同深度和宽度的结构选择最适合侧扫声纳沉船数据集的模

项目资助: 国家自然科学基金 (41974005, 41971416, 42074074)

第一作者: 汤寓麟, 博士研究生, 主要研究方向为水下目标检测和计算机视觉。494592292@qq.com

通讯作者: 边少锋, 博士, 教授。sfbian@sina.com

型, 并通过 CIoU_Loss 进行损失函数的改进和基于 GA+K (Genetic Algorithm+K-mean) 的检测框优化算法对选择的模型进行改进, 以期增强小目标检测能力、提升检测效率和减小模型权重, 达到工程应用的目标。

1. YOLOv5 基础模型

YOLOv5 模型结构和 YOLOv3 模型一样由输入、Backbone、Neck 和 Prediction 四部分组成, 具体模型结构如图 1 所示。

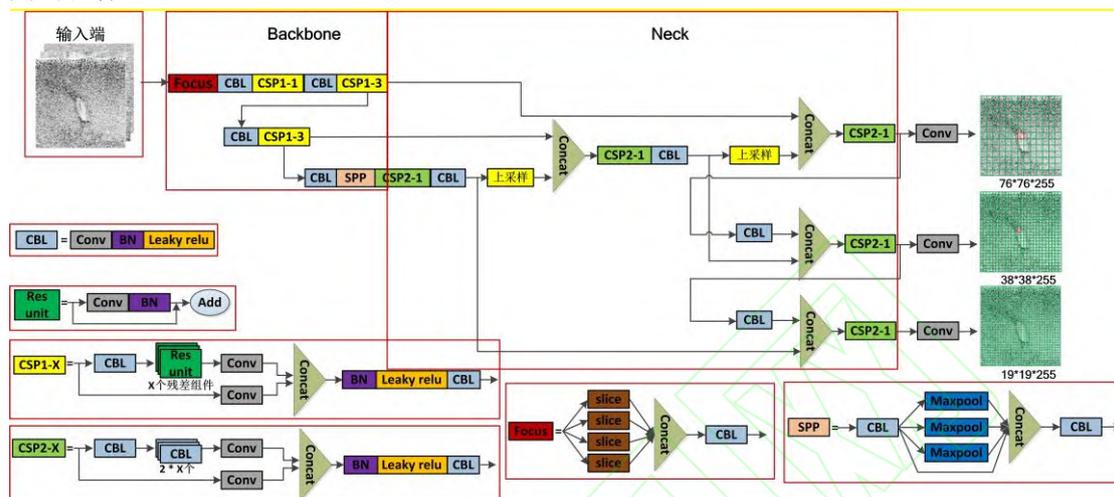


图 1 YOLOv5 模型结构

Fig.1 YOLOv5 Model Structure

输入端即表示输入的图片, YOLOv5 网络在该阶段对输入的图片进行一个图像的预处理。首先将输入的图像缩放到网络输入的大小, 本文网络输入大小为 608×608 , 然后再进行归一化等操作。

YOLOv5 较 YOLOv3 在数据增强方面采用了 Mosaic 方式, 将四张图片进行翻转、缩放、色域变化等操作, 在随机剪裁后进行拼接, 从而形成一张新的图片作为数据集进行模型的训练。这种方式的优势体现在丰富检测数据集, 尤其是随机缩放增加了小目标的数量, 增强模型的鲁棒性, 并且在 BN (Batch Normalizing) 计算时同时计算四张图片的数据, 极大的提高了模型训练的效率。本文的侧扫声纳沉船数据集目标中心点在 x 和 y 轴的分布如图 2 所示。

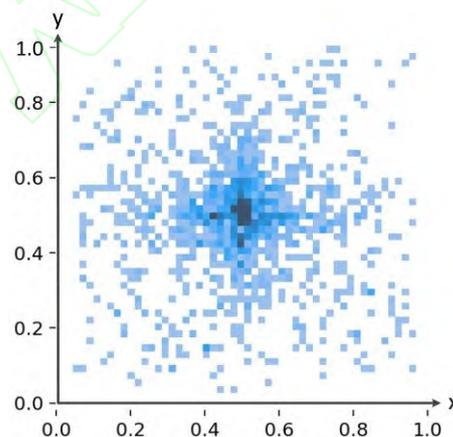


图 2 检测目标在图上分布情况

Fig.2 The Distribution of Objects on the Graph

从图中可以看出, 沉船目标主要集中在图片的中央位置, 且大多为小尺寸的目标。而 Mosaic 数据增强方法丰富了目标的位置分布情况且在一定程度上放大了小尺寸目标, 从而在提高模型训练效率的同时将提高模型的泛化能力。部分 Mosaic 训练图片如图 3 所示。

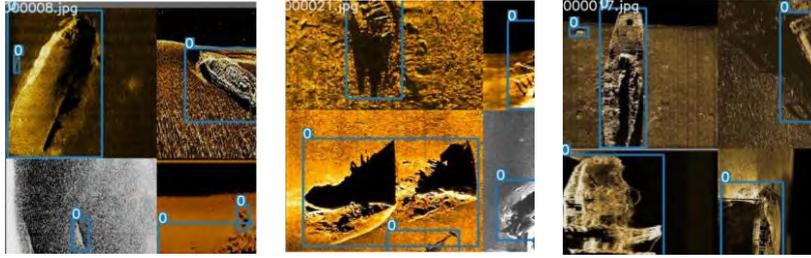


图3 部分 Mosaic 数据增强后的训练集

Fig.3 Part of the Mosaic Enhanced Training Set

YOLOv5 的主干网络主要由一个 Focus、四个 CBL 和三个 CSPDarknet53 和一个 SPP 结构组成，用以提取目标的通用特征。

YOLOv5 将 CSP (Cross Stage Partial Network) 跨阶段局部网络结构作为主干网络的基础，解决了大型卷积神经网络框架 Backbone 中网络优化的梯度信息重复的问题。

Focus 结构的作用是在图片进行下采样的过程中避免图片特征信息的丢失，从而在使用卷积操作时提取到更加充分的特征。Focus 模块在模型中的关键作用是在图片没有进入 backbone 之前进行切片操作，即选取每隔一个像素的值进行下采样，在没有丢失信息的情况下得到四张特征互补的图片，并将 W、H 信息集中到通道空间，输出空间扩充了四倍，原先的三通道变成了十二通道，从而在卷积操作后得到没有信息丢失的二倍下采样特征图，切片操作如图 4 所示。

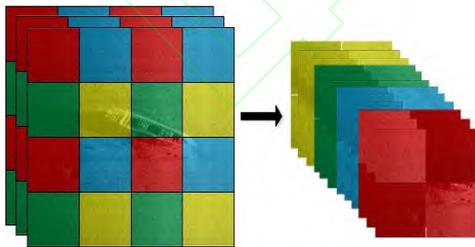


图4 Focus 模块切片操作示意图

Fig.4 Schematic Diagram of Slicing Operation of Focus Module

Neck 位于 Backbone 和 Prediction 之间，目的是进一步丰富特征的多样性以达到提升模型鲁棒性的目的。YOLOv5 在 FPN (Feature Pyramid Networks) 的基础上采用 PAN (Path Aggregation Network)

结构进行特征融合，具体如图 5 所示。

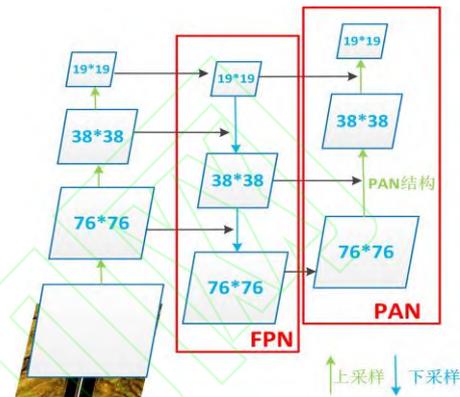


图5 FPN+PAN 结构图

Fig.5 FPN+PAN Structure

从图中可以看出 FPN 是由上而下，通过上采样将深层特征与底层特征进行融合得到预测的特征图，但是这个操作仅仅是将高层的强语义特征进行向下的传递，虽然增强了模型对图像特征的学习能力，但是却可能丢失一些定位特征。为此，在 FPN 之后添加了 PAN，通过自下而上的传达强定位特征，与 FPN 达到互补的效果，通过组合这两个模块，从不同的主干层对不同的检测层进行参数的聚合，从而全面的提升了模型的鲁棒性和学习性能。

Prediction 即为输出端，作用为完成目标检测结果的输出。

2. 模型改进策略

2.1 GA+K 先验框改进算法

为了获得更加适合本数据集的 anchors，提升模型的识别和定位精度，考虑到 YOLOv5 模型随机设置的 9 个 anchors 未必能很好的适应本实验数据集的目标，另外模型对训练数据进行了随机数据增强，增加了目标的多样性，因此，

对所有 anchors 进行了 GA+K 算法的优化，GA+K 算法流程图如图 6 所示。

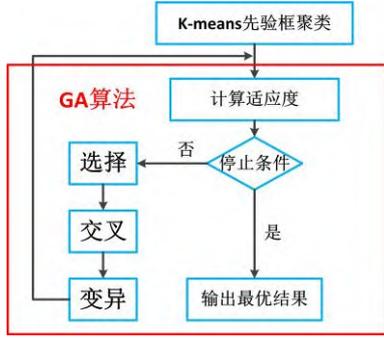


图 6 GA+K 算法流程图

Fig.6 GA+K Algorithm Flow Chart

GA+K 算法流程图包括 K-means 聚类算法和 GA 两大部分，首先对随机的 9 个 anchors 采用以交并比为距离度量的 K-means 聚类算法进行优化，并进行适应度的计算，并计算 bpr (best possible recall)，经多次试验验证，当 bpr 数值为 0.98 时先验框效果最好，若 bpr 大于 0.98 即输出最优结果；若 bpr 小于 0.98，则利用 GA (Genetic Algorithm) 算法进行迭代优化，进行选择、交叉和变异操作寻找最优适应度值以获得最优结果。

其中 K-means 聚类公式如式 (1) (2) 所示。

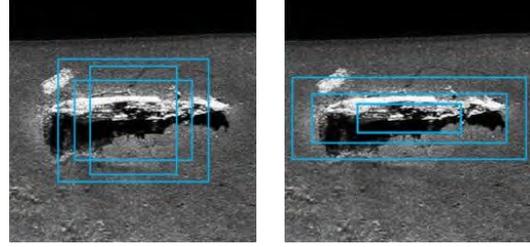
$$d(b,o)=1-IO(b,o) \quad (1)$$

$$IOU(b_{an},b_{gt})=\frac{b_{an} \cap b_{gt}}{b_{an} \cup b_{gt}} \quad (2)$$

式中， $d(b,o)$ 为先验框 b 和聚类中心 O 之间的距离； b_{an} 为先验框； b_{gt} 为标框。

适应度函数为度量的加权组合，其中 $AP_{0.5}$ 和 $AP_{0.5-0.9}$ 权重占比分别为 10% 和 90%。在计算 bpr 时主要考虑的是 9 个 anchors 和 gt (ground truth) 框宽高之间的差距。bpr 为所有 gt 框选择宽高比例值满足 threshold 的匹配度最高的平均值，其中 threshold 为设置的宽高比例阈值。gt 框与 anchor 的长宽比值为 $\frac{w_{gt}}{w_{an}}$ 和 $\frac{h_{gt}}{h_{an}}$ ，值可能大于 1 或者小于 1，本文通过取倒数的方式统一为小于或者等于 1。

经过 GA+K 算法改进的先验框和传统的先验框对比如图 7 所示，可以看出经过改进后的先验框更加符合沉船数据集目标特点。



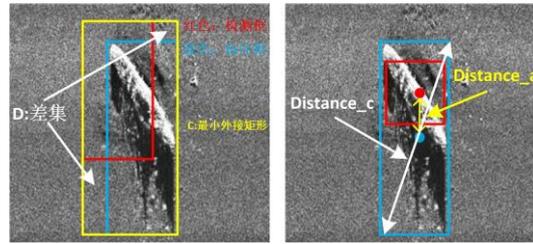
(a) 传统先验框 (b) 改进后先验框

图 7 改进前后先验框对比图

Fig.7 Comparison Diagram of Traditional and Improved Anchors

2.2 CIUO_Loss 损失函数改进

为达到提高模型定位精度的目的，考虑到 YOLOv5 模型采用 GIOU_Loss 作为回归损失函数，该损失函数虽然针对 IOU=0 和 IOU 相同但检测框位置或者大小不一样情况增加了相交尺度的衡量方式，但是当检测框面积相同但在标注框内部不同位置或者在相同位置但长宽比不一样时，此时检测框和标注框的差集 D 相同，GIOU_Loss 无法区分相对位置关系，因此，本文在考虑重叠面积的基础上增加了检测框和标注框中心点距离和长宽比因素的影响。GIOU 和 CIUO 原理图如图 8 所示，CIUO_Loss 作为损失函数，如式 (3) 所示。



(a) GIOU 原理图 (b) CIUO 原理图

图 8 改进前后损失函数原理图

Fig.8 Schematic Diagram of Traditional and Improved Loss

$$CIUO_Loss=1-(IOU-\frac{Distance_a^2}{Distance_c^2}-\delta\tau)$$

$$\partial = \frac{\tau}{(1-IOU)+\tau} \quad (3)$$

$$\tau = \frac{4}{\pi} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w^p}{h^p} \right)^2$$

式中, $Distance_a$ 为标注框和预测框中心点的坐标, $Distance_c$ 为最小外接矩形的对角线距离, ∂ 是权重函数, τ 是度量宽高比的一致性, w^p 为预测框宽、 h^p 为预测框的高、 w^{gt} 为标注框的宽、 h^{gt} 为标注框的高。

3. 实验与分析

为设计出在侧扫声纳海底沉船目标上检测性能更优的 YOLOv5 模型, 本文首先构建了八种不同宽度和深度的模型, 分别为 YOLOv5a、YOLOv5b、YOLOv5c、YOLOv5d、YOLOv5s、YOLOv5m、YOLOv5l 和 YOLOv5x, 并与 YOLOv3 进行比对实验。然后, 通过实验挑选出性能最优的模型结构, 并通过 CIou_Loss 来改进损失函数、GA+K 算法进行检测框的优化以达到更优的检测性能。

本文的实验数据集由各海道测量部门和国内生产厂家使用 Klein3000、EdgeTech4200、Yellowfin 和海卓系列等国内外主流侧扫声纳仪器设备, 在我国黄渤海、东海和南海等区域实测得到的 1 200 张侧扫声纳沉船图像组成。实验均基于 Pytorch 框架用 Python 语言实现, 实验环境: Windows10 操作系统; CPU 为 Intel(R) Core(TM) i9-10900X@3.70 GHz; GPU 为 NVIDIA GeForce RTX 3090, 24 GB 内存。

本文首先将数据集根据 1.1 节进行 Mosaic、图像旋转、图像加噪等数据增强操作后, 通过程序将数据集按照 0.8: 0.2 随机分成训练集和测试集, 并设置了超参

数具体如表 1 所示。

表 1 训练超参数

Tab.1 Training Hyper parameter

训练参数	参数值
动量	0.937
训练步数	1 200
warm-up 步数	3
warm-up 动量	0.8
权重衰减	0.000 5
学习率	0.01

本实验在正式开始训练之前进行了 3 steps 小学习率的 warm-up 训练, 目的是使用小学习率的预热训练使模型慢慢趋于稳定, 等相对稳定之后再使用预设的学习率进行训练, 使模型训练收敛速度变得更快, 模型效果更佳。

3.1 基础模型性能比对分析

在深度学习模型中, 结构越复杂, 深度越深, 检测效果越好。但是相应的模型参数也就越多, 训练效率越低, 并且过于复杂的模型在对小样本数据的训练上未必能够达到最优的检测效果, 甚至会产生过拟合的情况, 另外越复杂的模型权重越高, 检测时需要更多的浮点运算, 不利用工程部署。为此, 为了设计出性价比更高的模型, 根据侧扫声纳沉船数据集的特点, 考虑到训练效率和检测精度等综合因素, 本文分别尝试 YOLOv5a、YOLOv5b、YOLOv5c、YOLOv5d、YOLOv5s、YOLOv5m、YOLOv5l 和 YOLOv5x 八种基础模型结构, 八种模型的各项指标如表 2 所示。

表 2 八种模型指标参数

Tab.2 Index Parameters of Eight models

模型	深度	宽度	层数	参数数量	GFLOPS
YOLOv5a	0.67	0.50	391	9 373 302	22.7
YOLOv5b	0.55	0.55	373	11 429 950	27.4
YOLOv5c	0.50	0.63	355	14 857 277	35.7
YOLOv5d	0.40	0.75	293	16 183 005	37.2
YOLOv5s	0.33	0.50	283	7 276 605	17.1

YOLOv5m	0.67	0.75	391	21 375 645	51.4
YOLOv5l	1.00	1.00	499	46631350	114.1
YOLOv5x	1.33	1.25	607	87244374	217.3

其中 Depth 表示模型深度比、Width 表示模型的宽度比，八种模型按照 s、a、b、c、d、m、l 和 x 顺序结构越来越复杂，模型参数越来越多，每秒 10 亿次的浮点运算 (GFLOP) 越高，相应的对 GPU 的性

能需求就越强。

实验根据表 1 的超参数设置对 8 种深度的 YOLOv5 模型和 YOLOv3 模型在侧扫声纳沉船图像训练集上进行了 1 200 steps 的训练，训练结果如图 9 所示。

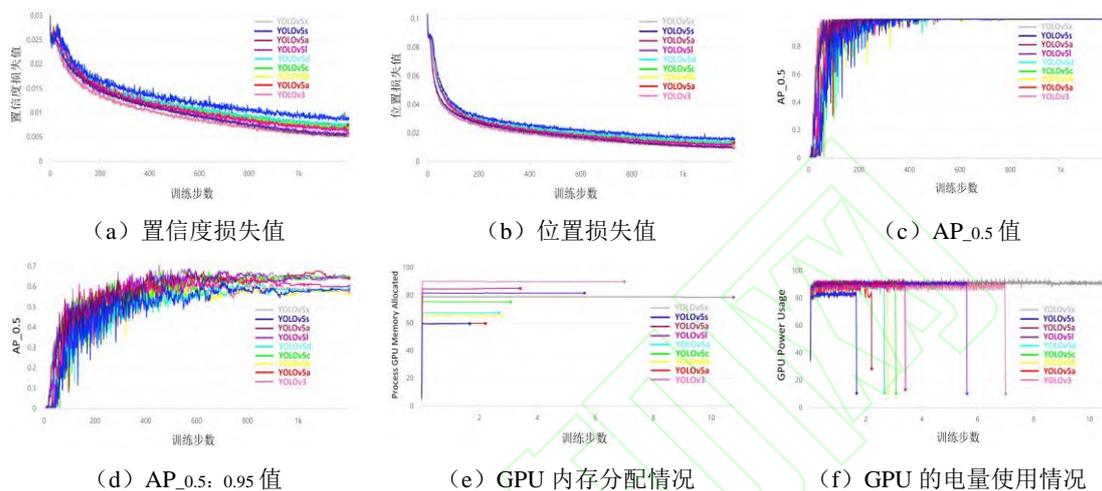


图 9 八种模型训练过程对比图

Fig.9 Comparison Diagram of Training Process of Eight Models

模型损失值由置信度和位置损失组成，分别如图 9 (a) (b) 所示，八种模型损失值均随着训练次数的增加而不断减小并逐渐趋于稳定，并且收敛速度相当。其中，YOLOv5a 模型无论是置信度损失还是位置损失均最低。

8 种模型在 IOU 设置为 0.5 时的 AP_{0.5} 值如图 9 (c) 所示，均在训练 500 之后趋于稳定达到了 1，为了更好的比较模型之间的检测性能，本实验比较了模型在 IOU 阈值从 0.5 至 0.95，步长为 0.05 的情况下的平均 AP 值，即 AP_{0.5: 0.95}，如图 9 (d) 所示，8 个模型在训练了大约 900 steps 后趋于稳定，其中 YOLOv5a 模型达到最高的 0.6461。

在推进工程应用的目的下，为了更好的评估各模型在训练时对计算机性能需求，本实验对 GPU 分配内存和 GPU 的电量使用情况进行了实时记录，分别如图 9

(e) (f) 所示。由图可知模型的结构越复杂，参数越多，相应的对计算机的性能需求越高，训练所花费的时间越长，成本越高。其中 YOLOv3 模型相比其他模型对 GPU 的内存需求最高，达到 90%，训练时间为 7h22min；模型结构最复杂的 YOLOv5x 对 GPU 的电量使用需求最高，且训练时间最长；而结构最简单的 YOLOv5s 无论是对 GPU 内存需求还是 GPU 电量使用需求均最低，相应的训练时间也最短；YOLOv5a 和 YOLOv5s 对 GPU 的内存需求一样仅为 59%，几乎仅为 YOLOv3 模型训练耗时 1/3 不到，大大减少了对计算机性能的硬性要求。

为了更好的评估模型检测性能，将训练完成的 8 个 YOLOv5 和 YOLOv3 模型在测试集上进行检测，检测的各项指标如表 3 所示。

表 3 八种模型在测试集上各项性能指标

Tab.3 Performance Indicators of the Eight Models on the Test Set

Model	准确率/%	召回率/%	FPS	AP _{0.5%}	AP _{0.5-0.95%}	检测速度/ms	模型权重/M
-------	-------	-------	-----	--------------------	-------------------------	---------	--------

YOLOv3	85.5	77.4	226	77.9	46.1	8.12	120.5
YOLOv5s	87.8	83.2	467	81.1	49.1	3.6	14.1
YOLOv5a	89.6	83.9	430	81.8	51.6	3.7	18.6
YOLOv5b	89.4	81.9	411	79.4	48.2	4.7	22.6
YOLOv5c	88	80.6	406	81.2	50.7	5.0	28.8
YOLOv5d	88.3	83.2	385	80.2	50.6	4.8	31.3
YOLOv5m	89.6	83.2	355	78.7	48.7	5.2	41.4
YOLOv5l	89	78.7	274	78.6	49	7.0	91.5
YOLOv5x	87.7	83.2	173	80.6	49.9	10.6	170.9

由表 3 可知，模型结构越复杂、参数越多，则相应的训练时间越长，训练完成后模型的权重就越大，而相对模型每秒的检测帧数越少，从而导致检测速度越慢，不利于工程部署和应用。并且越复杂的模型结构在不考虑效率的前提下也未必能达到最好的检测效果，甚至效果还不如结构相对简单的其他模型，比如结构最为复杂的 YOLOv5x 模型在检测性能上仅优于 YOLOv3 模型。而 YOLOv5a 模型在所有模型结构复杂程度上仅略比 YOLOv5s 复杂，但是在测试集上准确率和召回率却达到最高的 0.896 和 0.839， $AP_{0.5}$ 和 $AP_{0.5:0.95}$ 达到最高的 0.825 和 0.516，并且 FPS 达到了 432，相较 YOLOv3 模型几乎提升了 100%，且 18.6 M 的模型权重只有 YOLOv3 模型权重的 1/10。

综合图 9 和表 3 可知，YOLOv5 模型

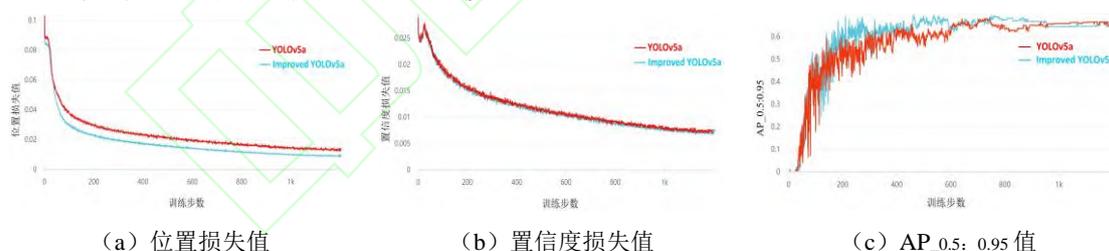


图 10 两种模型训练过程对比图

Fig.10 Comparison Diagram of Training Process Between Two Models

由图 10(a)(b)可知，改进 YOLOv5a 模型和 YOLOv5 基础模型一样，损失值随着训练步数的增加而不断下降，并在 1000 steps 后趋于稳定，模型达到拟合。无论是位置损失值还是置信度损失值，改进的模型均低于基础模型，其中改进模型的位置损失函数明显低于基础 YOLOv5a 模型，证明改进策略使模型具有更高的定位精度。 $AP_{0.5:0.95}$ 值如图 10(c) 所示，改

除了结构最为复杂的 YOLOv5x 模型，其他模型无论是在检测性能还是模型训练效率上都全面超越了 YOLOv3，其中 YOLOv5a 模型无论是在准确率、召回率、 $AP_{0.5}$ 还是 $AP_{0.5:0.95}$ 都达到最高，虽然其在 FPS、检测速度和最终的模型权重上略大于结构最简单的 YOLOv5s 模型，但是 0.1 ms 的检测速度差异几乎对模型性能没有影响，因此在全面综合评估后，选择 YOLOv5a 作为最佳的检测模型。

3.2 改进模型检测效果与性能评估

为进一步提升模型的检测性能，本文对选择的 YOLOv5a 模型通过 CIUO_Loss 进行损失函数改进并使用 GA+K 算法进行检测框改进。改进的 YOLOv5a 和 YOLOv5a 基础模型在训练时的损失值和 $AP_{0.5:0.95}$ 如图 10 所示。

改进模型在整个训练过程中 AP 值提升的效率更高，在训练 200 steps 后整体高于基础模型，最终达到 66.08%，相较基础模型提高了 1.36%。

为了更好的评估模型性能，将训练完成后的 YOLOv5a 和改进 YOLOv5a 模型在测试集上进行实验，两个模型的性能对比如表 4 所示。

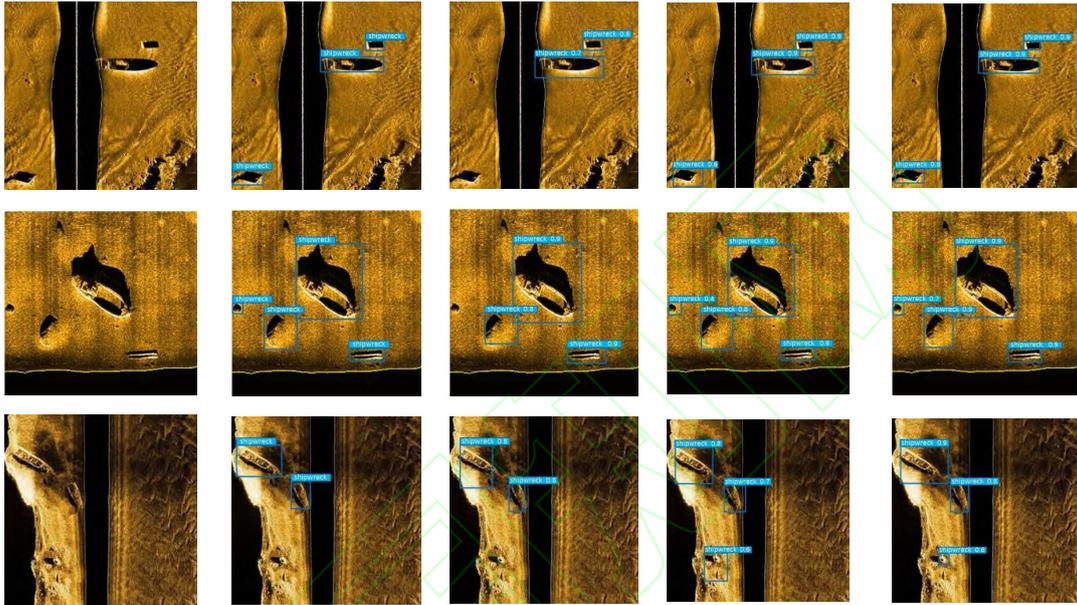
表 4 两种模型测试集性能对比

Tab.4 Comparison of the Performance of the Two Models in the Testing Set

	准确率/%	召回率/%	AP _{0.5} /%	AP _{0.5-0.9} /%	FPS
YOLOv5a	89.6	83.9	81.8	51.6	430
Improved YOLOv5a	90.8	85.5	82.1	52.2	426

从表 4 可以看出,改进的 YOLOv5a 较 YOLOv5a 模型在 AP_{0.5} 上提高了 0.3%, AP_{0.5: 0.95} 提高了 0.6%, Precision 提高了 1.2%, Recall 提高了 1.6%, 证明

CIOU_Loss 损失函数的改进和 GA+K 算法对检测框的优化在不影响检测速度的前提下,提高了模型的定位精度和检测精度。



(a) 测试样本原图 (b) 标注图 (c) YOLOv3 (d) YOLOv5a (e) 改进 YOLOv5a

图 11 三种模型检测效果对比

Fig.11 Comparison of Detection Results of the Three Models

图 11 所示为改进的 YOLOv5a 模型与基础模型、YOLOv3 模型的部分沉船目标检测效果对比。从 (a) 至 (e) 依次是测试样本原图、标注图、YOLOv3 检测图、YOLOv5a 检测图和改进的 YOLOv5a 检测图。由 (c) 可知, YOLOv3 模型基本可以准确的识别出大部分沉船目标,但是第一、二张图存在小尺度目标漏警的情况;与 YOLOv3 模型相比, YOLOv5a 模型很好的检测出了小尺度目标,同时在检测的置信度和定位精度上都了一定的提高。改进的 YOLOv5a 模型较基础的 YOLOv5a 模型具有更高的检测精度和定位精度。虽然在第三张图的检测中, YOLOv5a 模型存在虚警的情况,错误的将左下角的礁石目标检测成海底沉船目标,但是在实际的海底沉船搜救任务中,将类似沉船目标错误检测

成沉船目标的现实意义要远远超于对海底沉船目标的遗漏,也侧面的证明 YOLOv5a 较 YOLOv3 模型具有更好的全场景理解能力和更好的鲁棒性和泛化性。

4. 结语

针对基于 YOLOv3 模型的侧扫声纳海底沉船检测方法存在小目标漏警率高、模型权重大以及检测速度无法满足实时性需求的问题,本文引入了 YOLOv5 算法,根据沉船数据集特点设计了 YOLOv5a 模型,通过 CIOU_Loss 对损失函数进行改进,使用 GA+K 算法的对检测框进行优化。实验结果表明, YOLOv5 模型在检测性能上较 YOLOv3 模型有了全面提升,其中改进 YOLOv5a 模型在测试集中 AP_{0.5} 和 AP_{0.5-0.9} 达到 82.1% 和 52.2%, 较基础

YOLOv5a 模型提高了 0.3% 和 0.6%，较 YOLOv3 算法提高了 4.2% 和 6.1%；改进模型检测速度达到 426 帧/秒，较 YOLOv3 算法提升了几乎一倍，并且模型权重仅为 18.6 M，在保持了检测性能的同时兼顾了检测速度和模型权重，使侧扫声纳海底沉船的自动检测技术在工程应用上更具有现实价值。

参考文献

- [1] ZHAO Jianhu, LI Juanjuan, LI Meng. Progress and Future Trend of Hydrographic Surveying and Charting [J]. Journal of Geomatics, 2009, 34(04): 25-27. (赵建虎,李娟娟,李萌.海洋测量的进展及发展趋势[J].测绘信息与工程, 2009, 34(04): 25-27.)
- [2] WANG Jiu, CAO Jingtao, LU Bingwu, HE Bangtao. Underwater Target Detection Project Equipment Application and Development Trend[J]. China Water Transport, 2016(11):43-44. (王久,曹敬涛,卢秉武,何邦涛.水下目标探测项目设备应用和发展趋势[J].珠江水运,2016(11):43-44.)
- [3] CHEN Zhengrong,WANG Zhenghu. Research on Underwater Target Detection Using Side-scan Sonar and Multibeam Sounding System[J]. Hydrographic Surveying and Charting, 2013,33(04):51-54. (陈正荣,王正虎.多波束和侧扫声纳系统在海底目标探测中的应用[J].海洋测绘,2013,33(04):51-54.)
- [4] WANG Jiu, ZHOU Jian. Comprehensive Application of Side-scan Sonar and Multi-beam System in Shipwreck Survey[J]. China Water Transport, 2010,10(08):35-37. (王久,周健.侧扫声纳和多波束系统在失事沉船扫测中的综合应用[J].中国水运(下半月刊),2010,10(08):35-37.)
- [5] LIU Chen. The Comparative Analysis of Multi-beam Sounding System, Side-Scan Sonar and Magnetometer in the Wreck Detection[D]. China University of Geosciences (Beijing), 2015. (刘陈.多波束系统、侧扫声纳与磁力仪在海底沉船探测中的比较分析[D].中国地质大学(北京),2015.)
- [6] ZHAO Jianhu, WANG Xiao, ZHANG Hongmei. An Automatic Bottom Extracting Method for Side-scan Sonar Image. Geomatics and Information Science of Wuhan University, 2017, 42(12): 1797-1803. (赵建虎,王晓,张红梅.侧扫声呐图像海底线自动提取方法研究[J].武汉大学学报(信息科学版), 2017, 42(12): 1797-1803.)
- [7] YANG Fanlin, LIU Jingnan, ZHAO Jianhu. Sidescan Sonar Image Preprocessing Techniques Based on Data Fusion[J]. Geomatics and Information Science of Wuhan University, 2004(05): 402-406. (阳凡林,刘经南,赵建虎.基于数据融合的侧扫声呐图像预处理[J].武汉大学学报(信息科学版),2004(05):402-406.)
- [8] ZHU Dianyao, BIAN Hongyu. Research of Side-scan sonar target auto detection[J]. Journal of Jilin University (Information Science Edition), 2008, 26(06): 627-631. (朱殿尧,卞红雨.侧扫声纳目标自动探测研究[J].吉林大学学报(信息科学版),2008,26(06):627-631.)
- [9] LI Haibin, TENG Huizhong, SONG Haiying, DUAN Dongfang, HUANG Yi. Technology on the extraction of seabed target based on high resolution Side-scan sonar[J]. Hydrographic Surveying and Charting, 2010, 30(06): 71-73. (李海滨,滕惠忠,宋海英,段东方,黄毅.基于侧扫声纳图像海底目标物提取方法[J].海洋测绘, 2010, 30(06): 71-73.)
- [10] WANG Xiao. Research on Precise Processing of Side Scan Sonar Image and Object Recognition Methods[D].Wuhan University, 2017. (王晓.侧扫声呐图像精处理及目标识别方法研究[D].武汉大学, 2017.)
- [11] ZHAO Jianhu, WANG Xiao, ZHANG Hongmei, HU Jun, JIAN Xiaomin. The Neutrosophic Set and Quantum-behaved Particle Swarm Optimization Algorithm of Side-scan Sonar Image Segmentation[J]. Acta Geodaetica et Cartographica Sinica, 2016,

- 45(08): 935-942+951. (赵建虎,王晓,张红梅,胡俊,简晓敏.侧扫声呐图像分割的中性集合与量子粒子群算法[J]. 测绘学报, 2016, 45(08): 935-942+951.)
- [12] ZHAO Jianhu, WANG Aixue, WANG Xiao, GUAN Yongxian. A Segment Mosaic Method for Side scan Sonar Stripe Images Using Corresponding Features[J]. Geomatics and Information Science of Wuhan University,2013,38(09):1034-1038. (赵建虎,王爱学,王晓,关永贤.侧扫声纳条带图像分段拼接方法研究[J].武汉大学学报(信息科学版),2013,38(09):1034-1038.)
- [13] YAN Fanlin, DU Zhixing, WU Zixing, LI Jiabiao, CHU Fengyou. Object Recognizing on Sonar Image Based on Histogram and Geometric Feature[J]. Marine Science Bulletin, 2006(05):64-69. (阳凡林,独知行,吴自银,李家彪,初凤友.基于灰度直方图和几何特征的声纳图像目标识别[J].海洋通报,2006(05):64-69.)
- [14] ZHU Dianyao. Underwater acoustic image processing and pattern recognition[D]. Harbin Engineering University, 2008. (朱殿尧.水声图像处理与模式识别[D].哈尔滨工程大学, 2008.)
- [15] XU Yuanjun. Research on the Key Technologies of Detection Underwater[D]. Dalian Maritime University, 2011. (续元君.水下目标探测关键技术研究[D].大连海事大学,2011.)
- [16] CHEN Qiang. Research-based Underwater Acoustic Images Underwater Target Recognition[D]. Harbin Engineering University, 2012.(陈强.基于水声图像水下目标识别的技术研究[D].哈尔滨工程大
- 学,2012.)
- [17] SURAJ Kamal, SHAMEER K, MOHAMMED P R, SASEENDRAN Pillai, et al. Deep learning architectures for underwater target recognition[C]. Proceeding of Sympol 2013, Kochi, India, 23-25.
- [18] GUO Jun, MA Jinfen, WANG Aixue. Study of Side-scan sonar image classification based on SVM and gray level Co-Currence Matrix[J]. Geomatic&Spatial Information Technology, 2015, 38(03): 60-63. (郭军,马金凤,王爱学.基于SVM算法和GLCM的侧扫声纳影像分类研究[J].测绘与空间地理信息, 2015, 38(03): 60-63.)
- [19] TANG Yulin, JIN Shaohua, BIAN Gang, et al. The Transfer Learning with Convolutional Neural Network Method of Side-scan Sonar to Identify Wreck Images[J]. Acta Geodaetica et Cartographica Sinica,2021, 50(2): 260-269. (汤寓麟,金绍华,边刚,等.侧扫声呐识别沉船影像的迁移学习卷积神经网络法[J].测绘学报,2021, 50(2): 260-269.)
- [20] TANG Yulin, JIN Shaohua, BIAN Gang, ZHANG Yonghou, LI Fan. Wreckage Target Recognition in Side-scan Sonar Images Based on an Improved Faster R-CNN Model[C]. International Conference on Big Data & Artificial Intelligence & Software Engineering 2020:348-354.
- [21] TANG Yulin, JIN Shaohua, BIAN Gang, ZHANG Yonghou. Shipwreck Target Recognition in Side-scan Sonar Images by Improved YOLOv3 Model Based on Transfer Learning[J]. IEEE Access, vol. 8, pp. 173450-173460.

Improved YOLOv5 Method for Detecting Shipwreck Target with Side-scan Sonar

TANG Yulin¹, BIAN Shaofeng¹, ZHAI Guojun², LIU Min³, ZHANG Weidong⁴

¹ College of Electrical Engineering, Naval Univ. of Engineering, Wuhan Hubei 430033

² Naval Institute of Oceanographic Surveying and Mapping, Tianjin 300061

Abstract:In view of the problems of high miss-alarm rate of small targets, heavy model weight, and detection speed that fails to meet real-time requirements in side-scan sonar shipwreck detection method based on the YOLOv3 model. The paper introduces the YOLOv5 algorithm and proposes a model based on YOLOv5 according to the characteristics of the side-scan sonar shipwreck dataset. Try YOLOv5a, YOLOv5b, YOLOv5c, YOLOv5d, YOLOv5s, YOLOv5m, YOLOv5l and YOLOv5x under the basic framework of YOLOv5 with eight different depth and width model structures. Choose the best structure use GA+K (Genetic Algorithm+K-mean) algorithm to optimize the detection frame, and improve the loss function through CIOU_Loss. The experimental results show that the improved YOLOv5a model is 0.3% and 0.6% higher than the original model in AP_{0.5} and AP_{0.5-0.9}, and has a substantial improvement compared with the YOLOv3 model, in which AP_{0.5} and AP_{0.5-0.9} are improved by 4.2% and 6.1% and the detection speed reaches 426 frames per second, which is almost doubled that of YOLOv3, which is more conducive to practical applications and engineering deployment.

Key words: Side-scan Sonar Shipwreck Target; Detection Method; YOLOv5 Model; CIOU_Loss; Genetic Algorithm+K-mean Algorithm

First author: TANG Yulin, PhD, specializes in underwater target detection and computer vision. E-mail:494592292@qq.com

Corresponding author: BIAN Shaofeng, PhD, professor. E-mail:sfbian@sina.com

Foundation support: The National Natural Science Foundation of China (41974005, 41971416, 42074074)