

光电工程

Opto-Electronic Engineering

中文核心期刊 中国科技核心期刊
Scopus CSCD

基于ZYNQ的轻量化YOLOv5声呐图像目标检测算法及实现

赵冬冬, 谢墩翰, 陈朋, 梁荣华, 沈伊, 郭新新

引用本文:

赵冬冬, 谢墩翰, 陈朋, 等. 基于ZYNQ的轻量化YOLOv5声呐图像目标检测算法及实现[J]. 光电工程, 2024, 51(1): 230284.

Zhao D D, Xie D H, Chen P, et al. Lightweight YOLOv5 sonar image object detection algorithm and implementation based on ZYNQ[J]. *Opto-Electron Eng*, 2024, 51(1): 230284.

<https://doi.org/10.12086/oe.2024.230284>

收稿日期: 2023-11-21; 修改日期: 2024-01-10; 录用日期: 2024-01-10

相关论文

基于残差和注意力网络的声呐图像去噪方法

赵冬冬, 叶逸飞, 陈朋, 梁荣华, 蔡天诚, 郭新新

光电工程 2023, 50(6): 230017 doi: 10.12086/oe.2023.230017

基于级联稀疏查询机制的轻量化火灾检测算法

张小雪, 王雨, 吴思远, 孙帮勇

光电工程 2023, 50(10): 230216 doi: 10.12086/oe.2023.230216

基于改进YOLOv5s的无人机图像实时目标检测

陈旭, 彭冬亮, 谷雨

光电工程 2022, 49(3): 210372 doi: 10.12086/oe.2022.210372

基于YOLOv3和ASMS的目标跟踪算法

吕晨, 程德强, 寇旗旗, 庄焕东, 李海翔

光电工程 2021, 48(2): 200175 doi: 10.12086/oe.2021.200175

更多相关论文见光电期刊集群网站 

 | 光电工程
Opto-Electronic Engineering

<http://cn.ojournal.org/oe>



 OE_Journal



Website

DOI: 10.12086/oe.2024.230284

基于 ZYNQ 的轻量化 YOLOv5 声呐图像目标检测算法及实现

赵冬冬¹, 谢墩翰¹, 陈朋^{1*}, 梁荣华¹,
沈伊¹, 郭新新²

¹浙江工业大学计算机科学与技术学院, 浙江 杭州 310023;

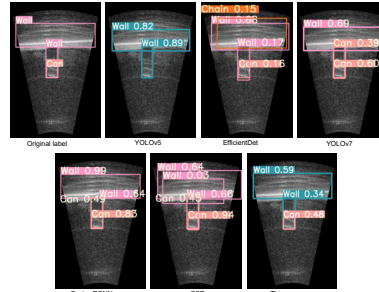
²中国科学院深海科学与工程研究所, 海南 三亚 572000

摘要: 针对声呐图像存在的模糊、样本量不足的现象, 本文提出一种基于 YOLOv5 的声呐图像目标检测改进算法。利用几何滤波、垂直翻转等方法, 对声呐图像数据集进行数据增强。添加融合注意力机制模块, 使算法更好地关注声呐图像小目标的特征。同时, 针对目前大多数目标检测算法运行在云端, 无法做到实时性声呐图像检测的问题, 本文利用替换轻量级网络和 NCNN 边缘移植技术, 同时在颈部网络中采用 GSConv 模块, 将算法成功移植到 ZYNQ 平台, 实现声呐图像的嵌入式端实时检测。实验表明, 本文提出的算法在降低了 56% 参数量的同时, 在 map50 和 map50-95 上分别提高 2.2% 和 2.5%。改进后的算法性能提升明显, 证明所提出的方法在轻量化声呐图像目标检测任务上具有一定的可行性与有效性。

关键词: 水下目标检测; YOLO; ZYNQ; 声呐图像; 深度学习; 轻量化

中图分类号: TP391

文献标志码: A



赵冬冬, 谢墩翰, 陈朋, 等. 基于 ZYNQ 的轻量化 YOLOv5 声呐图像目标检测算法及实现 [J]. 光电工程, 2024, 51(1): 230284
Zhao D D, Xie D H, Chen P, et al. Lightweight YOLOv5 sonar image object detection algorithm and implementation based on ZYNQ[J]. *Opto-Electron Eng*, 2024, 51(1): 230284

Lightweight YOLOv5 sonar image object detection algorithm and implementation based on ZYNQ

Zhao Dongdong¹, Xie Dunhan¹, Chen Peng^{1*}, Liang Ronghua¹, Shen Yi¹, Guo Xinxin²

¹College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, Zhejiang 310023, China;

²Institute of Deep-sea Science and Engineering, Chinese Academy of Sciences, Sanya, Hainan 572000, China

Abstract: To address the problems of blurring and insufficient sample size in sonar images, an improved sonar image target detection algorithm is proposed based on YOLOv5. The algorithm uses geometric filtering, vertical flipping, and other methods to enhance the sonar image dataset. The fusion attention mechanism module is added to make the algorithm better focus on the features of small targets in sonar images. At the same time, in response to the problem that most target detection algorithms currently run on the cloud and cannot achieve real-time sonar image detection, this paper uses lightweight network replacement and NCNN edge porting technology. It adopts the GSConv module in the neck network to successfully transplant the algorithm to the ZYNQ platform, realizing real-time detection of sonar images on the embedded end. After experiments, the algorithm proposed in this paper

收稿日期: 2023-11-21; 修回日期: 2024-01-10; 录用日期: 2024-01-10

基金项目: 国家自然科学基金资助项目 (62371421, 62001418, 62005245, U1909203); 浙江省自然科学基金资助项目 (LQ21F010011)

*通信作者: 陈朋, chenpeng@zjut.edu.cn.

版权所有©2024 中国科学院光电技术研究所

reduced the parameter quantity by 56%, increasing map50 and map50-95 by 2.2% and 2.5%, respectively. The algorithm's performance has significantly improved, proving that the method proposed has certain feasibility and effectiveness in lightweight sonar image target detection tasks.

Keywords: underwater target detection; YOLO; ZYNQ; sonar image; deep learning; lightweight

1 引言

地球上海洋总面积约为 3.6 亿平方千米, 超过地球表面面积的三分之二, 海洋资源对于地球上的生命来说至关重要。21 世纪, 人类进入了大规模开发利用海洋的时期。随着海洋强国战略任务的提出, 我们需要提高对海洋资源的开发以及利用能力, 而这需要大量精确的海洋环境数据, 以便进行水下目标定位、水下目标检测、海底资源探测^[1-3] 等任务。

水下目标检测技术是进行水下目标定位与识别的关键^[4]。水下环境复杂多变, 对水下数据的传输以及水下数据的可用性会造成很大影响, 水下目标的检测和识别^[5] 具有很大挑战性。如何提高检测和识别的速率以及准确度、降低水下数据计算以及通信成本、减小检测模型的参数数量和计算量, 这些都是亟待攻克的核心难题。水下目标检测的成像技术有许多种, 相较于光学成像等水下探测技术, 声呐成像具有探测距离远、穿透能力强的特点, 声呐能够在复杂多变的水下环境下稳定成像, 因而成为目前主流的水下目标检测手段^[6-7]。由于水声信道的复杂性, 以及声波本身传播的损失和散射等因素, 成像声呐直接采集得到的图像质量不高、分辨率低, 存在大量散斑噪声, 目标成像边缘不清晰^[8-9], 这也严重影响了声呐图像的后续处理^[10]。

传统的声呐图像目标检测算法包括基于数理统计、基于数学形态学和基于像素的方法^[11]。这些方法存在检测效果差、检测时间长等问题, 需要寻找更优的方法提高声呐图像的检测准确度以及缩短声呐图像的检测时间。近几年, 在目标检测领域, 卷积神经网络取得了飞速的发展, 越来越多的人开始尝试利用深度学习的方法来进行水下环境中的目标检测^[12]。Valdenegro^[13] 针对前视声呐图像的水下目标检测任务, 设计了一种卷积神经网络, 能够根据特定的阈值, 对区域建议进行目标物的置信度评估, 达到了很高的召回率。通过使用深度卷积神经网络, Williams^[14] 设计了一个 10 层卷积神经网络模型, 对合成孔径声呐 (SAS) 图像进行三个二元分类任务 (目标与非目标、

截锥体与岩石、地雷与截断圆锥体)。模型性能较传统的基于特征的分类方法有了明显的提升, 表明了深度学习在声呐图像目标检测上的巨大优势。Kim^[15] 等将 YOLO 目标检测算法应用于前视声呐图像的目标检测任务中。实验结果表明, YOLO 算法在离线验证时速度高达 107.7 FPS, 远超过传统的滑动窗口算法的 0.2 FPS。在真实海洋环境的前视声呐图像中, 算法帧率为 5 FPS, 证明了其在声呐图像目标实时检测方面的可行性。

通过卷积神经网络的逐层卷积、填充、池化等操作, 基于深度学习的目标检测算法, 可以从声呐图像中提取到更深层次的语义信息。虽然基于深度学习的目标检测算法性能较之传统算法有了显著提高^[16-17], 但由于网络所需参数量大, 因而基于深度学习的目标检测算法大多不能直接部署于嵌入式端, 无法达到水下设备实时进行目标检测的目的^[18]。由于声呐图像的特殊性, 声呐图像数据集较少且声呐图像成像质量不高, 如何利用小样本进行深度学习也成了进行声呐图像检测不可避免的难题^[19]。在声呐图像中, 细小目标和细长形状是常见的, 但目前的主流算法在对它们进行定位回归时, 检测精度有所不足^[20]。

本文针对细长、细小的声呐图像检测精度差的问题, 提出一种基于注意力机制的声呐图像小目标检测算法, 改善检测细长、细小的声呐图像的关键策略。针对水下设备内存不足以及设备运算能力有限的问题, 设计了一种声呐图像的轻量化检测方法, 通过替换轻量化网络及 NCNN 推理框架压缩移植模型, 以及在颈部网络中采用 GSConv 模块替换传统卷积, 来改善目标检测网络模型参数量太大、计算复杂的问题, 使其可部署于 ZYNQ 嵌入式平台, 达到实时检测声呐图像的目的。

2 面向声呐图像的轻量化 YOLOv5 目标检测算法

YOLOv5 是一种基于深度学习的单阶段目标检测算法。相比于 YOLOv4^[21], 它在速度与精度都有了显

著的提升。在目标检测任务中，特别是对小目标的检测中，表现出优秀的检测效果^[22-23]。对比后来提出的 YOLOv7 算法，YOLOv5 的训练和推理速度更快，并且内存占用也更少。这使得 YOLOv5 在某些应用场景中更占优势，例如在移动设备或者运行资源有限的系统中。由于水下声呐图像模糊、噪声大、目标小，同时嵌入式设备运行内存小，设备运算能力有限的情况，为了提升声呐图像小目标的检测精度和提高模型的推理速度并进一步降低网络参数量，本文在 YOLOv5 算法的基础上，设计了一种面向声呐图像的轻量化 YOLOv5 目标检测网络模型，其结构如图 1 所示。

2.1 深度可分离卷积

传统的 YOLOv5 算法采用 CSP-Darknet53 作为主干网络来进行图像特征的提取，虽然在检测效果上有较高的精度，但在嵌入式平台上，其计算量和内存占用太大，不适合低运算能力和存储能力的情况^[24]。

为了进一步减少参数量和运算量，本文采用深度可分离卷积^[25]代替原来的卷积模块。深度可分离卷

积包含逐通道卷积和逐点卷积。逐通道卷积用不同的卷积核对输入的不同通道单独做卷积，而逐点卷积是用 1×1 的卷积核对逐通道卷积所有通道的输出合并在一起。其结构如图 2 所示。

对于一个卷积点而言：假设有一个 3×3 大小的卷积层，输入通道为 16、输出通道为 32。要用 32 个 3×3 大小的卷积核在 16 个通道中分别做卷积，得到 32 个输出通道，参数数量为 $16 \times 32 \times 3 \times 3 = 4608$ 个。

如果应用深度可分离卷积结构块，先用 16 个 3×3 大小的卷积核在 16 个通道上做逐通道卷积，得到 16 个特征图谱。在融合操作之前，接着用 32 个 1×1 大小的卷积核在这 16 个特征图谱上做逐点卷积，参数数量为 $16 \times 3 \times 3 + 16 \times 32 \times 1 \times 1 = 656$ 个。可以看出，深度可分离卷积的应用可以显著地降低模型的参数量。

2.2 GSConv

在 YOLOv5 中，Neck 层是位于 Backbone 和 Head 之间的一组卷积层，用于提取特征金字塔。特征金字塔有助于模型对不同大小和比例的物体进行泛化。

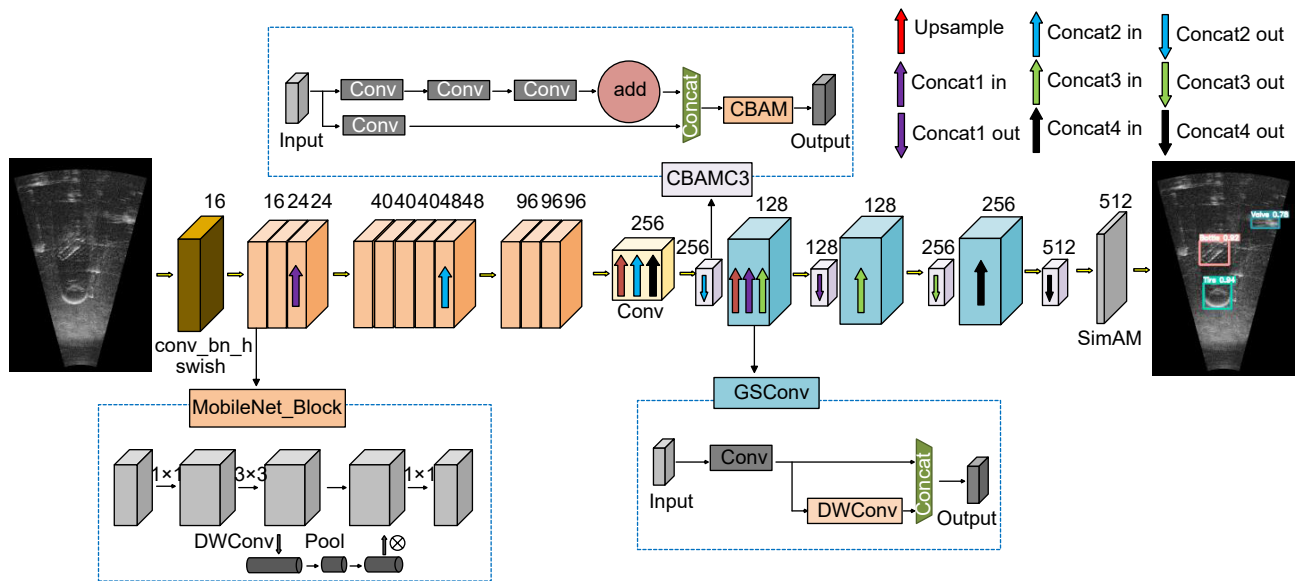


图 1 轻量化声呐图像 YOLOv5 目标检测网络模型

Fig. 1 Lightweight sonar image YOLOv5 object detection network model

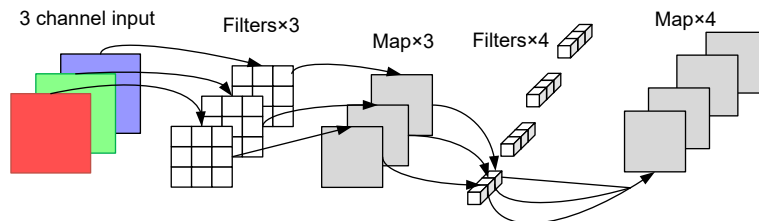


图 2 深度可分离卷积结构

Fig. 2 Depthwise separable convolution structure

Neck 层的作用是将不同层次的特征图进行融合, 以提高特征提取的效率和准确性。相较于普通卷积, 深度可分离卷积虽然显著减少了参数量, 但其最大缺陷在于一定程度上忽视了通道之间的关系, 产生了信息割裂。深度可分离卷积对输入通道的内容进行分层处理, 针对每一个通道进行相互分离的卷积操作, 最后依照其输入通道重新合并, 虽然在一定程度上降低了多通道所带来的计算量问题, 但是通道间的相关信息却丢失了。为了有效保留多通道信息, 增强提取到的语义信息, 让图像特征更好地表示图像的本质和特点^[26], 本文在 Neck 层采用 GSConv^[27] 代替传统卷积, 提高检测模型的精度, 同时减少模型的参数量, 而不额外增加计算量^[28]。

GSConv 模块的结构如图 3 所示。设输入和输出通道数分别为 C_1 和 C_2 。首先经过一个标准卷积, 通道数变为 $C_2/2$, 再经过一个深度可分离卷积, 通道数不变, 最后将两次卷积的结果进行拼接和混洗。最后的混洗操作, 能够将通道信息进行均匀打乱, 增强提取到的语义信息, 加强特征信息的融合, 提高图像特征的表达力。

GSConv 模块是由标准卷积、深度可分离卷积和 Shuffle 三部分卷积混合而成。该模块对标准卷积所生成的特征信息通过 Shuffle 混合策略渗透到深度可分离卷积生成的特征信息的每一模块中, 使得该方法的卷积计算接近于标准卷积的输出, 但降低了计算成本。通过添加深度可分离卷积层和 Shuffle 层, 增强了特征信息的非线性表达能力, 使得 GSConv 卷积更加适用于轻量化的模型检测器。

网络在 Neck 层进行特征融合时, 语义信息也会不断地向下传输, 当特征图的高宽和通道数被不断压缩和扩展时, 会导致部分语义信息的丢失, 从而影响到最后的预测。本文在网络的 Neck 层引入 GSConv

模块, 使用 GSConv 模块代替标准卷积来进行上采样和下采样, 降低模型的参数量和计算量, 并在最大程度保证采样效果, 保证训练精度。

2.3 融合注意力机制

对于声呐图像存在模糊且小目标物体易被忽略等现象, 研究发现在主干网络中添加注意力机制可以有效地剔除不相关信息的干扰, 更好关注小目标物体的特征^[29-30]。为了提高模型的精度, 本文中使用了三维无参注意力机制 SimAM 和卷积块注意力机制 CBAM 相结合的融合注意力机制。

2.3.1 CBAM 注意力机制

在 YOLOv5 算法改进中, 常见的是添加 SE (squeeze-and-excitation) 注意力机制模块^[31]。SE 注意力机制能够通过自适应地学习通道之间的关系来增强重要特征的表达力, 但是在声呐图像小目标检测问题上存在着一些不足。由于声呐图像中的目标尺寸较小, 其空间信息对于检测和定位至关重要, SE 注意力机制是将特征图压缩为全局描述向量, 这会导致空间信息的丢失, 不利于声呐图像小目标检测。因此本文引入 CBAM 注意力机制^[32], 更好地结合通道和空间注意力, 更好地适应小目标的特征表示需求。

卷积块注意力机制 (CBAM), 包括两个子模块: 通道注意力模块 (CAM) 和空间注意力模块 (SAM)。两个模块具有不同的功能, SAM 可以提取输入图像的主要特征信息, CAM 可以根据每个特征通道的重要性设置权重。CBAM 模块结构如图 4 所示。

在通道注意力模块中, 将输入的图像分别应用平均池化和最大池化, 输出的特征图通过只有一个隐藏层的多层感知器 (multi layer perceptron, MLP) 来生成通道注意力图。再通过元素相加计算通道注意力子模块。计算公式如下:

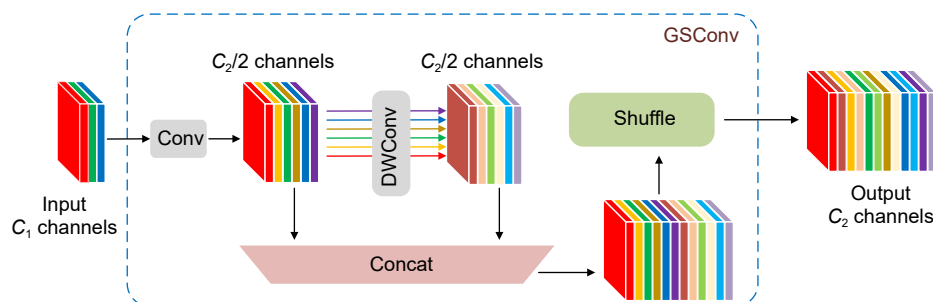


图 3 GSConv 结构图

Fig. 3 GSConv structure diagram

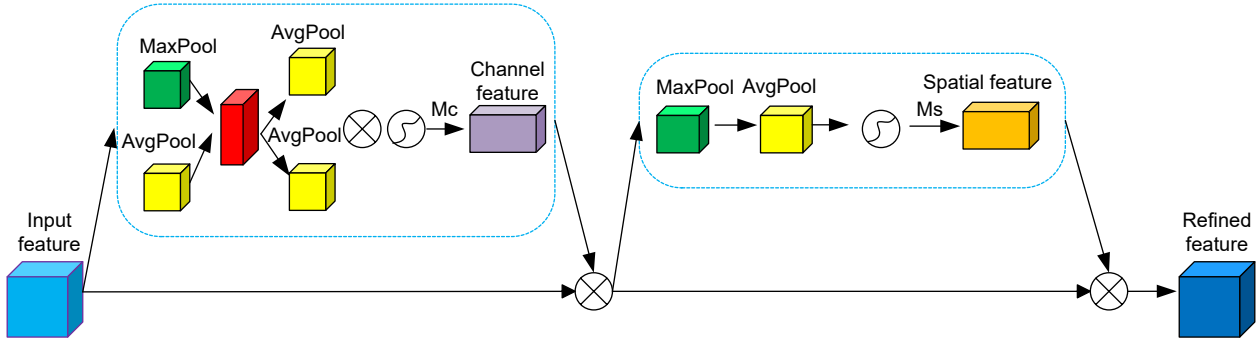


图 4 CBAM 模块结构图

Fig. 4 CBAM module structure diagram

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))), \quad (1)$$

其中： σ 是 Sigmoid 函数；MLP 为多层感知器； $\text{AvgPool}(F)$ 和 $\text{MaxPool}(F)$ 分别表示平均池化和最大池化； $M_c(F)$ 为通道注意力模块。

与通道注意力不同，空间注意力子模块为了更好地检测目标的位置信息，其沿通道轴同时进行平均池化和最大池化过程，并通过卷积层生成空间注意力图。计算公式如下：

$$M_s(F') = \sigma(f([\text{AvgPool}(F'), \text{MaxPool}(F')])), \quad (2)$$

其中： F' 为输入特征图； f 为 7×7 卷积运算； $\text{AvgPool}(F')$ 和 $\text{MaxPool}(F')$ 分别表示平均池化和最大池化； $M_s(F')$ 是空间注意力模块。

2.3.2 SimAM 注意力机制

由于声呐图像目标较小，其特征容易被忽略或丢失，本文在算法的 Head 部分因引入 SimAM 注意力机制^[33]，SimAM 的结构如图 5 所示。在实际声呐图像检测过程中，即使目标在图像中所占比例较小，依然可以通过评估每个神经元的重要性，从而提高局部感知能力以帮助网络更好地捕捉目标的语义特征，提高精度和成功率；同时，SimAM 机制通过调整特征图的尺寸和通道权重相互关系，加强相似的特征图之间的关联性，在不增加模型参数数量的前提下，实现模型的自适应特征选择。

该注意力机制的作用原理是通过能量函数来实现的，而能量函数的定义与神经科学领域中的空域抑制现象有关，活跃的神经元会抑制周围信息匮乏神经元的表现能力。因此，神经元的重要程度可通过能量函数来表达，最小能量可以通过式 (3) 得到：

$$e_i^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda}, \quad (3)$$

其中： t 代表输入特征 X 的目标神经元， $\hat{\mu}$ 和 $\hat{\sigma}^2$ 是除 t 外所有神经元计算所得的平均值和方差。对于一个目标神经元，如果它的输出与其他神经元的输出差异较大，那么它的能量越低，也就是它的重要性越高，因而可用 $1/e_i^*$ 来表示神经元的重要性。

接着，我们对特征进行增强以特征的表达能力：

$$\tilde{X} = \sigma\left(\frac{1}{E}\right) \odot X, \quad (4)$$

其中： \tilde{X} 表示增强后的特征图，是原始特征图 X 于 Sigmoid 函数处理后的能量的逐元素乘积， X 表示输入特征的原始图， E 表示每个神经元的能量。

本实验将 SimAM 注意力模块放在 Head 网络，此举可以在不额外增加参数的情况下，通过对空间位置特征和通道信息的全面评估来对重点神经元进行特别关注，最终实现在改善网络特征提取能力的同时加快网络训练速度。

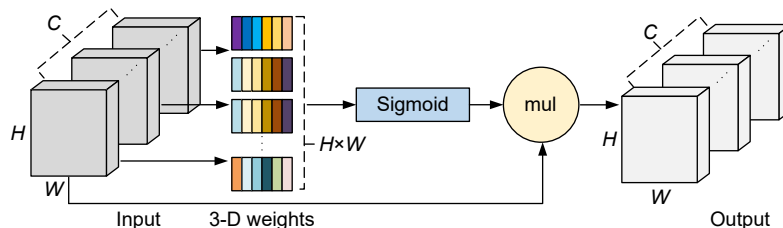


图 5 SimAM 结构图

Fig. 5 SimAM structure diagram

2.4 损失函数设计

因为声呐图像的特点, 图像数据集中各类别数量往往不平衡, 导致模型没有很好的泛化能力。本文借鉴 Focal-Loss 的计算思路, 将 YOLOv5 原始的 CIoU 损失函数优化为 Focal-CIoU^[34], 更好地处理小目标、困难样本和类别不平衡问题, 提高目标检测的精度和鲁棒性。它是一种动态缩放的交叉熵损失, 可以专注于难以分类的错误分类样本, 其计算公式为

$$L_{\text{Focal-CIoU}} = IoU^r L_{\text{CIoU}}, r = 0.5. \quad (5)$$

借助动态缩放因子, 可以动态降低训练中那些易分类样本的权重^[35], 从而让模型可以更快地关注那些复杂样本。对于分类不准确的样本, Focal-CIoU 的损失没有改变, 而对于分类准确的样本, 损失会变小^[36]。换言之, 增加了不准确分类样本在 Focal-CIoU 中的权重, 提高了分类的鲁棒性。

3 基于 ZYNQ 的声呐系统设计及目标检测算法实现

3.1 基于 ZYNQ 的声呐系统设计系统

为了验证本文所设计的声呐图像目标检测算法的

可行性, 设计了一种基于 ZYNQ 的声呐图像检测系统。本文所设计的声呐图像目标检测系统分为 PL 端和 PS 端两个部分, 系统设计图如图 6 所示。PL 端主要包括 Xilinx Artix-7 系列 FPGA, PS 端则主要两个 ARM Cortex A9 处理器组成。

本文所设计的声呐系统工作流程如图 7 所示, 主要包括两部分: PL 端产生 PWM 信号上传到发射驱动模块, 该模块通过升压激励发射换能器。接收换能器接收目标的反射和散射回波信号, 经过滤波、放大等信号调理后, 将信号传输至 AD 进行采集。随后, 系统对接收到的数据进行 DFT 处理, 并将处理结果上传至波束形成模块。PL 端通过波束形成算法每两秒约生成 40 张图, 根据式 (6) 计算每对图片的均方差 (MSE), 然后根据均方差的值选择差异最大的两张图传至 DDR; PS 端每次从 DDR 中读取两张声呐图像, PS 端的两个 CPU 利用转换好的模型同时进行声呐图像的实时目标检测。

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (I_1[i, j] - I_2[i, j])^2, \quad (6)$$

其中: m 和 n 分别是图像的高度和宽度。 $I_1[i, j]$ 和 $I_2[i, j]$ 分别是两个图像在像素位置 (i, j) 处的灰度值。

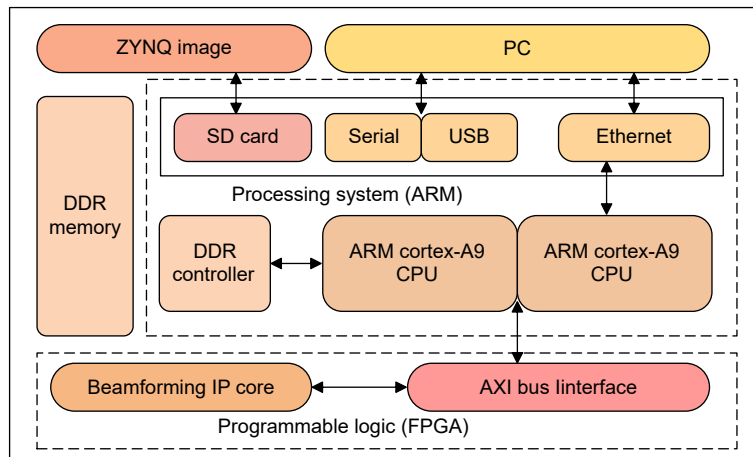


图 6 声呐系统结构图

Fig. 6 Diagram of the sonar system

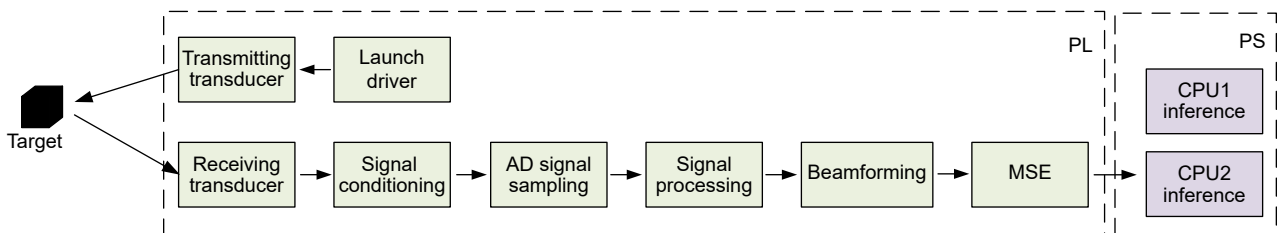


图 7 声呐系统工作流程图

Fig. 7 Sonar system workflow diagram

3.2 ZYNQ 端目标检测算法部署

PC 端所训练出的深度学习参数往往保存在指定的模型中, 并不能应用在所有硬件平台, 模型移植 ZYNQ 端首先需要经过参数提取与格式转换。如图 8 所示为本文算法部署到 ZYNQ 端的原理图, 深度学习的算法模型可基于 NCNN 框架实现高效快速的前向推理计算。声呐图像目标检测算法成功部署到 ZYNQ 端, 首先需要将 Pytorch 所训练的 pt 模型文件转化为开放神经网络交换模型 (open neural network exchange, ONNX), 并检查模型是否正确; 使用 onnx-simplifier 优化 ONNX 模型, 以提升模型的推理速度; 随后通过 onnx2ncnn 工具将 ONNX 模型转化为 NCNN 模型, 并利用 ncnnoptimize 工具对模型进行量化操作。

分别在 GPU 2080Ti、X86 CPU 和 ZYNQ7020 三个平台上进行实验, 以确定本文所设计的模型在不同平台上的计算量、内存占用、推理率、运行功耗和单位功耗帧速率指标, 所得的实验结果如表 1 所示。

将模型移植到 ZYNQ 平台时, 通过 ncnnoptimize 操作将卷积层替换为内积层, 并删除部分无用的连接, 减少了部分模型的计算量。更改的检测流程, 使得双核 CPU 更好的利用, 提高了推理速度, 减少了系统内存的占用, 同时也由于 ZYNQ7020 平台低功耗设计和 SoC 架构, 使得内存占用远远优于其他平台。对比模型的推理率, 显然, 2080Ti 的推断率是最好的, 然而功耗也异常高。本研究所使用的 ZYNQ7020 平台获得了高达 0.162 FPS/W 的优异性能, 从而验证了所设计的声呐系统比其他平台具有综合优势。

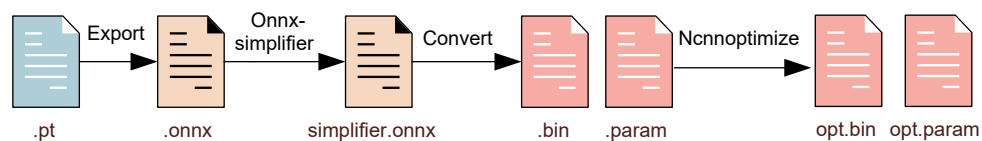


图 8 模型转换流程

Fig. 8 Model conversion process

表 1 不同设备运行结果

Table 1 Running results of different equipments

Equipment	FLOPs/G	Memory footprint/MB	Frame rate/FPS	Power/W	Frame rate/power/(FPS/W)
GPU	6.0	1401	36.4	250	0.146
CPU	6.0	2103	8.8	65	0.135
ZYNQ7020	5.8	90	0.6	3.7	0.162

4 实验结果与分析

4.1 声呐图像数据集

本文采用的数据集是使用 ARIS Explorer 3000 声呐在水箱和转盘中用前视声呐捕获的海洋废弃物数据 (vPropeller、Shampoo-bottle、Standing-bottle、Tire、Valve、Wall), 共有 1868 张图片。本文测试集按总数据 0.1 的比例划分, 训练集共 1718 张, 测试集共 150 张。数据集每个类别的分类如图 9 所示。

声呐图像目标检测神经网络的训练, 通常需要大量的水下数据支持。由于声呐图像受到设备、成本、环境等各种因素的限制, 导致声呐图像数据集的图片数量短缺。本文采用数据增强的方法, 对原始的训练集图片进行上下翻转、几何滤波等操作, 将原始的训练集图片数量扩大了三倍, 使其更好进行声呐图像上的卷积神经网络训练。扩充后训练集和验证集分别包含 5154 张和 150 张图片, 训练集图像数据增强的结果如图 10 所示。

4.2 目标检测算法性能

4.2.1 对比实验

为了进一步验证本文所提出的算法在声呐图像目标检测的优势, 将本文算法与其他常见目标检测算法进行了对比, 对比实验结果如表 2 所示。

从实验结果可以看出, 本文所提出的算法在 Map50 指标上高于其他算法, 在模型参数量上也优于其他算法。本文算法在检测精度和模型大小上均具有出色的能力, 更能适应嵌入式端声呐图像目标检测任务。

为了比较本文算法与其他算法在不同类别上的实

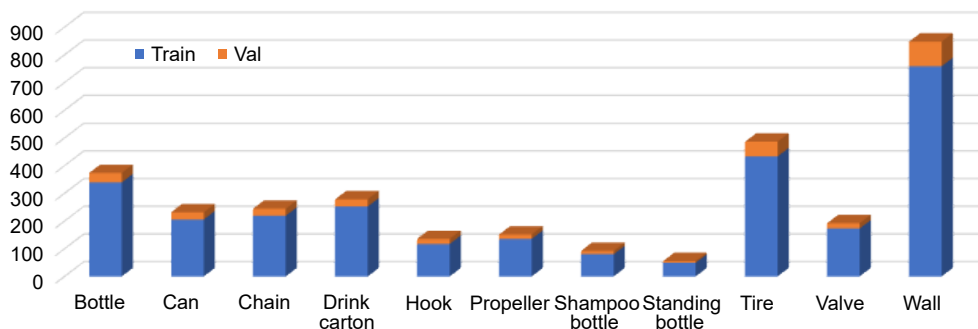


图 9 数据集中的图片数量分布

Fig. 9 Distribution of the number of images in the dataset

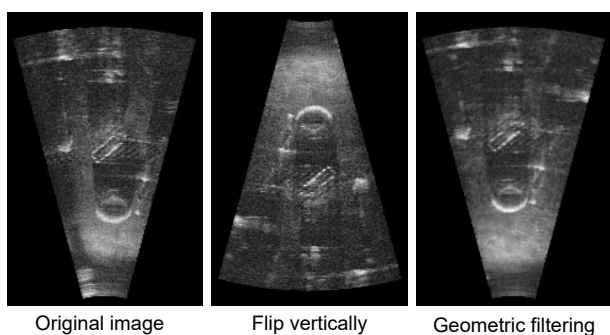


图 10 数据增强图像

Fig. 10 Data-enhanced images

际检验效果, 对数据集上的测试结果进行了可视化展示, 比较结果如图 11 所示。

从图中可以看出, 本文算法的检测框图与实际标注文件的目标最为紧密, 而且在各类别的检测情况中, 错检、漏检的现象都要优于其他算法。并且, 本文算法在 Can 和 Drink-carton 这些小目标类别上的检测效果优于其他算法, 更好地关注了声呐图像细小目标的特征。

4.2.2 消融实验

为了直观地观察不同的改进模块对声呐图像检测模型检测性能的影响, 本节采用消融实验进行验证。具体而言, 在原有的 YOLOv5s 模型中分别进行加入 SimAM、CBAM、混合注意力机制、改进的损失函数、替换主干网络的卷积为深度可分离卷积以及替换颈部网络的卷积为 GSConv 来保证在相同数据集、相同训练轮次下检测效果的比较。

实验结果如表 3 所示。当分别加入 SimAM 和 CBAM 时, Map50 和 Map50-95 均有所提高。将混合注意力机制加入到网络后, Map50 和 Map50-95 均比单独添加要高, 提高了目标检测的精度指标。当使用了改进的损失函数后, Map50 和 Map50-95 有了显著

表 2 对比试验

Table 2 Comparative experiments

Algorithm	Map50	Params(M)
YOLOv5s	0.912	7.2
EfficientDet	0.834	3.9
YOLOv7	0.890	37
Faster-RCNN	0.896	25.6
SSD	0.912	51
This paper	0.933	3.2

提高。当替换了 YOLOv5s 的主干网络的卷积为深度可分离卷积以及替换颈部网络的卷积模块后, 模型的参数量大大减少, Map50 和 Map50-95 也有了提高。当同时应用上述改进模块集成生成最终目标检测模型时, Map50 和 Map50-95 分别达到 0.934 和 0.708, 检测精度达到最高。

4.2.3 经典数据集对比实验

为证明改进模型也同样适用于其它图像的目标检测, 本文补充了对于经典数据集 VOC2007 的检测结果, 将 YOLOv5s 和本文算法进行对比实验, 结果如表 4 所示。

实验结果表明, 在经典数据集上, 本文算法相较于 YOLOv5s 算法, 本文算法的 Map50 提升不足 1%, 远不如声呐数据集上提升明显, 且 Map50-95 由于更多关注小目标特征反而有所下降, 证明了本文所改进模型在声呐图像目标检测上的有效性。

5 结论

本文针对声呐图像背景噪声大、小目标检测精度低等问题, 设计了一种面向声呐图像的轻量化 YOLOv5 目标检测算法。首先利用替换主干网络为轻

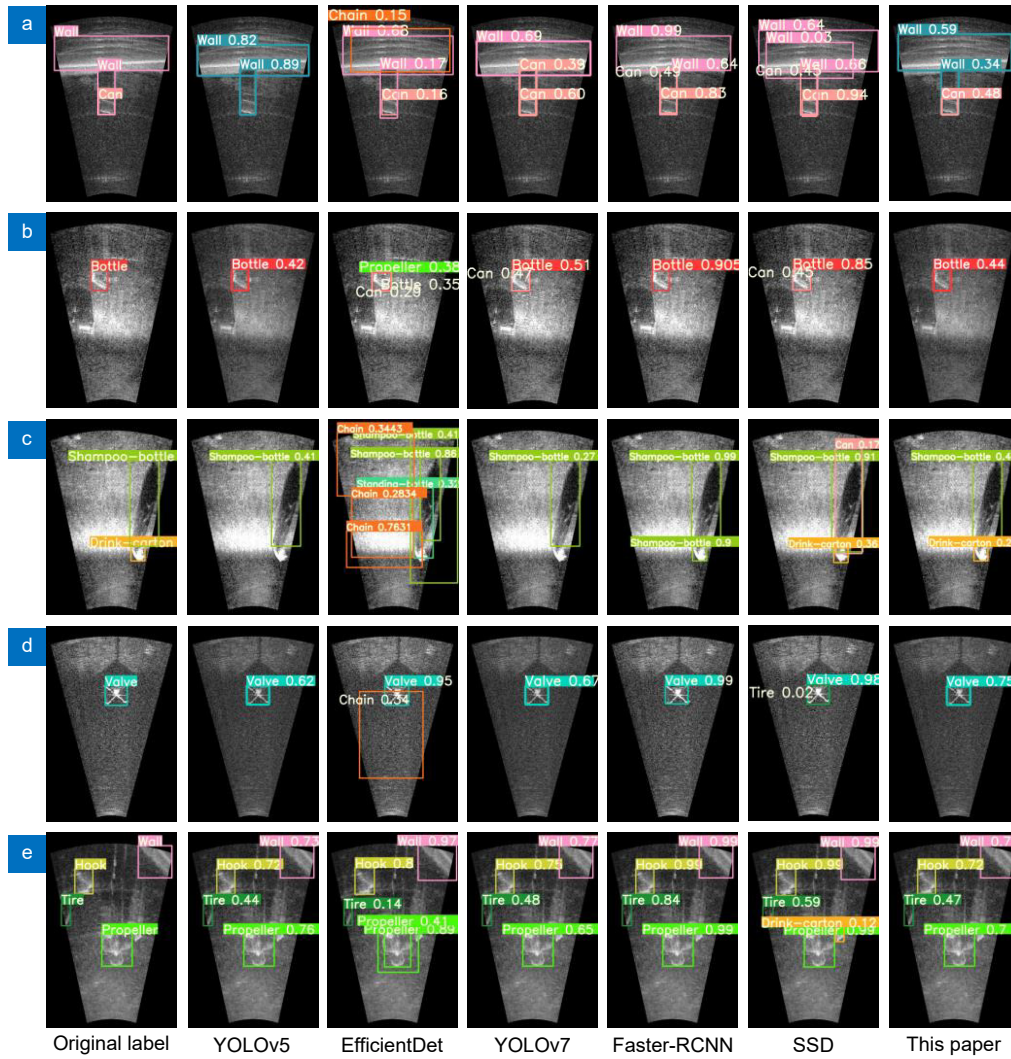


图 11 不同算法图像检测结果对比图

Fig. 11 Different algorithm image detection result comparison

表 3 消融实验

Table 3 Ablation experiments

Algorithm	Precision	Recall	Map50	Map50-95	Params(M)
YOLOv5s	0.896	0.905	0.912	0.685	7.2
YOLOv5s+SimAM	0.892	0.911	0.922	0.700	7.0
YOLOv5s+CBAM	0.896	0.927	0.918	0.697	7.1
YOLOv5s+SimAM+CBAM	0.892	0.912	0.925	0.701	7.1
YOLOv5s+Focal-CIOU	0.899	0.891	0.930	0.700	7.2
YOLOv5s+DWConv	0.880	0.911	0.919	0.694	3.6
YOLOv5s+GSCConv	0.897	0.906	0.923	0.699	6.2
This paper	0.907	0.894	0.933	0.712	3.2

表 4 经典数据集对比实验

Table 4 Comparison experiment of classic datasets

Algorithm	Precision	Recall	Map50	Map50-95
YOLOv5s	0.678	0.53	0.583	0.353
This paper	0.661	0.519	0.59	0.339

量级网络, 大大降低了模型的运算量, 然后利用融合注意力机制和改进损失函数的方法, 更好地提取小目标特征, 提升模型的检测精度, 最后利用 NCNN 前向推理框架, 将训练好的模型成功部署到 ZYNQ 端, 实现了声呐图像的实时检测。实验证明, 本文所提出的方法在声呐图像小目标检测任务中具有较高的检测准确度, 可应用到水下声呐图像实时检测任务中。

参考文献

- [1] Guo G, Wang X K, Xu H P. Review on underwater target detection, recognition and tracking based on sonar image[J]. *Control Decis*, 2018, **33**(5): 906–922.
郭戈, 王兴凯, 徐慧朴. 基于声呐图像的水下目标检测、识别与跟踪研究综述[J]. *控制与决策*, 2018, **33**(5): 906–922.
- [2] Petrich J, Brown M F, Pentzer J L, et al. Side scan sonar based self-localization for small autonomous underwater vehicles[J]. *Ocean Eng*, 2018, **161**: 221–226.
- [3] Huang S W, Chen E, Guo J. Efficient seafloor classification and submarine cable route design using an autonomous underwater vehicle[J]. *IEEE J Oceanic Eng*, 2018, **43**(1): 7–18.
- [4] Park C, Kim Y, Lee H, et al. Development of a 2 MHz sonar sensor for inspection of bridge substructures[J]. *Sensors (Basel)*, 2018, **18**(4): 1222.
- [5] Chen P, Cai X W, Zhao D D, et al. Despeckling for side-scan sonar images based on adaptive block-matching and 3D filtering[J]. *Opto-Electron Eng*, 2020, **47**(7): 190580.
陈朋, 蔡旭伟, 赵冬冬, 等. 基于自适应BM3D的侧扫声呐图像散斑降噪[J]. *光电工程*, 2020, **47**(7): 190580.
- [6] Abu A, Diamant R. A statistically-based method for the detection of underwater objects in sonar imagery[J]. *IEEE Sensors J*, 2019, **19**(16): 6858–6871.
- [7] Negahdaripour S. Application of forward-scan sonar stereo for 3-D scene reconstruction[J]. *IEEE J Oceanic Eng*, 2020, **45**(2): 547–562.
- [8] Shang Z G, Zhao C H, Wan J. Application of multi-resolution analysis in sonar image denoising[J]. *J Sys Eng Electron*, 2008, **19**(6): 1082–1089.
- [9] Jin Y, Ku B, Ahn J, et al. Nonhomogeneous noise removal from side-scan sonar images using structural sparsity[J]. *IEEE Geosci Remote Sens Lett*, 2019, **16**(8): 1215–1219.
- [10] Wang Z, Zhang S W, Huang W Z, et al. Sonar image target detection based on adaptive global feature enhancement network[J]. *IEEE Sensors J*, 2022, **22**(2): 1509–1530.
- [11] Mignotte M, Collet C, Pérez P, et al. Markov random field and fuzzy logic modeling in sonar imagery: application to the classification of underwater floor[J]. *Comput Vis Image Underst*, 2000, **79**(1): 4–24.
- [12] Köhntopp D, Lehmann B, Kraus D, et al. Classification and localization of naval mines with superellipse active contours[J]. *IEEE J Oceanic Eng*, 2019, **44**(3): 767–782.
- [13] Valdenegro-Toro M. End-to-end object detection and recognition in forward-looking sonar images with convolutional neural networks[C]//2016 IEEE/OES Autonomous Underwater Vehicles (AUV), 2016: 144–150. <https://doi.org/10.1109/AUV.2016.7778662>.
- [14] Williams D P. Underwater target classification in synthetic aperture sonar imagery using deep convolutional neural networks[C]//2016 23rd International Conference on Pattern Recognition (ICPR), 2016: 2497–2502. <https://doi.org/10.1109/ICPR.2016.7900011>.
- [15] Kim J, Yu S C. Convolutional neural network-based real-time ROV detection using forward-looking sonar image[C]//2016 IEEE/OES Autonomous Underwater Vehicles (AUV), 2016: 396–400. <https://doi.org/10.1109/AUV.2016.7778702>.
- [16] Liu G X, Liu S Y, Wu J F, et al. Machine vision object detection algorithm based on deep learning and application in banknote detection[J]. *China Measur Test Technol*, 2019, **45**(5): 1–9.
刘桂雄, 刘思洋, 吴俊芳, 等. 基于深度学习的机器视觉目标检测算法及在票据检测中应用[J]. *中国测试*, 2019, **45**(5): 1–9.
- [17] Harikrishnan J, Sudarsan A, Sadashiv A, et al. Vision-face recognition attendance monitoring system for surveillance using deep learning technology and computer vision[C]//2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN), 2019: 1–5. <https://doi.org/10.1109/ViTECoN.2019.8899418>.
- [18] Li B Q, Huang H N, Liu J Y, et al. Synthetic aperture sonar underwater multi-scale target efficient detection model based on improved single shot detector[J]. *J Electron Inf Technol*, 2021, **43**(10): 2854–2862.
李宝奇, 黄海宁, 刘纪元, 等. 基于改进SSD的合成孔径声呐图像水下多尺度目标轻量化检测模型[J]. *电子与信息学报*, 2021, **43**(10): 2854–2862.
- [19] Chen L, Zhang J L, Peng H, et al. Few-shot image classification via multi-scale attention and domain adaptation[J]. *Opto-Electron Eng*, 2023, **50**(4): 220232.
陈龙, 张建林, 彭昊, 等. 多尺度注意力与领域自适应的小样本图像识别[J]. *光电工程*, 2023, **50**(4): 220232.
- [20] Ge X Y, Wei N Y, Zhou H K, et al. Research on small underwater target detection technology based on side-scan sonar[J]. *Digit Ocean Underwater Warf*, 2023, **6**(2): 155–161.
葛锡云, 魏柠阳, 周宏坤, 等. 基于侧扫声呐的水下小目标检测技术研究[J]. *数字海洋与水下攻防*, 2023, **6**(2): 155–161.
- [21] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[Z]. arXiv: 2004.10934, 2020. <https://doi.org/10.48550/arXiv.2004.10934>.
- [22] Singh A, Kalaichelvi V, Karthikeyan R. Performance analysis of object detection algorithms for robotic welding applications in planar environment[J]. *Int J Comput Integr Manuf*, 2023, **36**(7): 1083–1108.
- [23] Chen X, Peng D L, Gu Y. Real-time object detection for UAV images based on improved YOLOv5s[J]. *Opto-Electron Eng*, 2022, **49**(3): 210372.
陈旭, 彭冬亮, 谷雨. 基于改进YOLOv5s的无人机图像实时目标检测[J]. *光电工程*, 2022, **49**(3): 210372.
- [24] Ge Z, Liu S T, Wang F, et al. YOLOX: exceeding YOLO series in 2021[Z]. arXiv: 2107.08430, 2021. <https://doi.org/10.48550/arXiv.2107.08430>.
- [25] Ying B Y, Xu Y C, Zhang S, et al. Weed detection in images of carrot fields based on improved YOLO v4[J]. *Traitem Signal*, 2021, **38**(2): 341–348.
- [26] Huang L, Yang Y, Yang C Y, et al. FS-YOLOv5: lightweight infrared rove target detection method[J]. *Comput Eng Appl*, 2023, **59**(9): 215–224.
黄磊, 杨媛, 杨成煜, 等. FS-YOLOv5: 轻量化红外目标检测方法[J]. *计算机工程与应用*, 2023, **59**(9): 215–224.
- [27] Li H L, Li J, Wei H B, et al. Slim-neck by GSConv: a better design paradigm of detector architectures for autonomous vehicles[Z]. arXiv: 2206.02424, 2022. <https://doi.org/10.48550/arXiv.2206.02424>.
- [28] Feng Z Q, Xie Z J, Bao Z W, et al. Real-time dense small object detection algorithm for UAV based on improved

- YOLOv5[J]. *Acta Aeronaut Astronaut Sin*, 2023, **44**(7): 327106.
奉志强, 谢志军, 包正伟, 等. 基于改进YOLOv5的无人机实时密集小目标检测算法[J]. *航空学报*, 2023, **44**(7): 327106.
- [29] Zhu X K, Lyu S C, Wang X, et al. TPH-YOLOv5: improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]//*Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops*, 2021: 2778–2788. <https://doi.org/10.1109/ICCVW54120.2021.00312>.
- [30] Niu Z Y, Zhong G Q, Yu H. A review on the attention mechanism of deep learning[J]. *Neurocomputing*, 2021, **452**: 48–62.
- [31] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 7132–7141. <https://doi.org/10.1109/CVPR.2018.00745>.
- [32] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[C]//*Proceedings of the 15th European Conference on Computer Vision*, 2018: 3–19. https://doi.org/10.1007/978-3-030-01234-2_1.
- [33] Yang L X, Zhang R Y, Li L D, et al. SimAM: a simple, parameter-free attention module for convolutional neural networks[C]//*Proceedings of the 38th International Conference on Machine Learning*, 2021: 11863–11874.
- [34] Zheng Z H, Wang P, Liu W, et al. Distance-IoU loss: faster and better learning for bounding box regression[C]//*Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence*, 2020: 12993–13000. <https://doi.org/10.1609/aaai.v34i07.6999>.
- [35] Yuan M Y, Song Y L, Zhang C, et al. Underwater object detection based on GA-RetinaNet[J]. *Comput Syst Appl*, 2023, **32**(6): 80–90.
袁明阳, 宋亚林, 张潮, 等. 基于GA-RetinaNet的水下目标检测[J]. *计算机系统应用*, 2023, **32**(6): 80–90.
- [36] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//*Proceedings of the 2017 IEEE International Conference on Computer Vision*, 2017: 2999–3007. <https://doi.org/10.1109/ICCV.2017.324>.

作者简介



赵冬冬(1990-), 男, 博士, 副教授, 博士生导师, 主要研究方向为图像以及信号处理。

E-mail: zhaodd@zjut.edu.cn



谢墩翰(1999-), 男, 硕士研究生, 主要研究方向为图像处理、嵌入式系统程序设计。

E-mail: 2112112026@zjut.edu.cn



【通信作者】陈朋(1981-), 男, 博士, 教授, 博士生导师, 主要研究方向为图像处理、嵌入式系统设计。

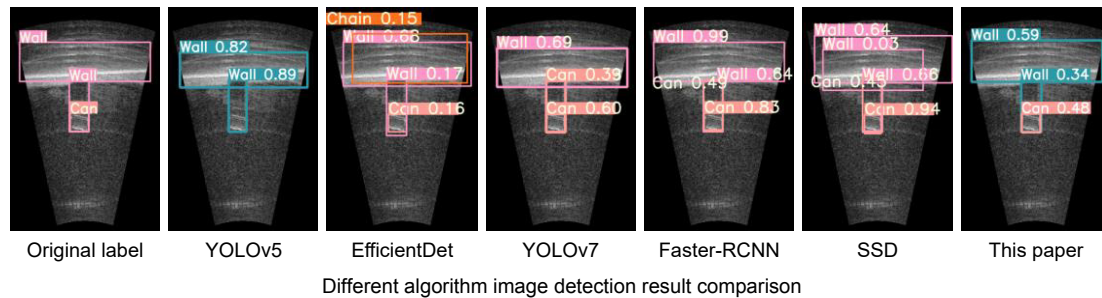
E-mail: chenpeng@zjut.edu.cn



扫描二维码, 获取PDF全文

Lightweight YOLOv5 sonar image object detection algorithm and implementation based on ZYNQ

Zhao Dongdong¹, Xie Dunhan¹, Chen Peng^{1*}, Liang Ronghua¹, Shen Yi¹, Guo Xinxin²



Overview: In the 21st century, mankind has entered a period of large-scale development and utilization of the ocean. With the strategic mission of becoming a maritime power, we need to improve our ability to develop and utilize marine resources, which requires a large amount of accurate marine environment data to carry out tasks such as underwater target detection and seabed resource detection. Underwater target detection technology is crucial for underwater target positioning and underwater resource exploration. Underwater target detection can be achieved through different imaging technologies, but sonar is currently the most commonly used detection method because it can operate reliably in low visibility conditions. Due to the complexity of underwater acoustic channels, as well as the loss and scattering of sound waves themselves, the image quality and resolution obtained by imaging sonar are low, accompanied by a large amount of speckle noise and unclear edges, which also seriously affects the follow-up of sonar images. Currently, most sonar image target detection algorithms run on the cloud and cannot achieve real-time sonar image detection. This paper proposes a lightweight YOLOv5 sonar image target detection algorithm based on ZYNQ to achieve real-time detection of small target images on the embedded side of sonar equipment. First, geometric filtering, vertical flipping, and other methods are used to perform data enhancement on the sonar image dataset, adding a fusion attention mechanism module allows the algorithm to better focus on the characteristics of small targets in sonar images. At the same time, in order to solve the problem that most target detection algorithms currently run in the cloud and cannot achieve real-time sonar image detection, this paper uses replacement lightweight networks and NCNN edge-end transplantation technology, and uses the GSConv module in the neck network to convert the algorithm successfully ported to ZYNQ platform. The sonar image detection system is independently designed. The PL side uses the wave velocity formation algorithm to generate images, and the PS side realizes the embedded side real-time detection of sonar images. After experiments, the algorithm proposed in this article reduced the calculation amount by 56%, while map50 and map50-95 increased by 2.2% and 2.5%, respectively. The performance of the improved algorithm has been significantly improved, proving that the method proposed in this article has certain feasibility and effectiveness in lightweight sonar image target detection tasks.

Zhao D D, Xie D H, Chen P, et al. Lightweight YOLOv5 sonar image object detection algorithm and implementation based on ZYNQ[J]. *Opto-Electron Eng*, 2024, 51(1): 230284; DOI: 10.12086/oe.2024.230284

Foundation item: Project supported by National Natural Science Foundation of China (62371421, 62001418, 62005245, U1909203), and Zhejiang Provincial Natural Science Foundation of China (LQ21F010011)

¹College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, Zhejiang 310023, China; ²Institute of Deep-sea Science and Engineering, Chinese Academy of Sciences, Sanya, Hainan 572000, China

* E-mail: chenpeng@zjut.edu.cn