文章编号:1671-4598(2024)09-0086-08 DOI:10.16526/j.cnki.11-4762/tp.2024.09.013 中图分类号:TP391.41 文献标识码:A

通道空间深度感知的轻量化水下目标检测

赵瑞金,李海涛,陆光豪
(青岛科技大学 信息科学技术学院,山东 青岛 266061)

摘要:提出了一种通道空间深度感知的轻量化水下目标检测网络 CSDP-L-YOLO;该网络基于 YOLOv5 网络进行改进,由特征感知模块和双注意门控策略组成;特征感知模块旨在将解码器中的多级特征自适应抑制或增强,优化类内学习的一致性,解决水下场景复杂导致的误检和漏检问题;通过线性操作和混洗结构生成特征映射,减少冗余特征的融合和计算,以减少模型的参数量和计算量;双注意门控策略是在编码器中同时引入并发通道空间挤压-激励机制模块和卷积注意力模块,进一步关注强相关性特征,增强模型对特征的敏感度;实验结果表明,与基线模型 YOLOv5-s 相比,*mAP* 提高了 2.4%,节省了 20%参数量和 15.8%计算量,检测速度提升了 8.2 ms;此外,与目前较为先进的 YOLOv8 模型相比,*mAP* 提高了 1.9%。

关键词:水下目标检测;通道空间深度感知;注意力机制;模型轻量化;特征融合;YOLO

Lightweight Underwater Target Detection for Channel Spatial Depth Perception

ZHAO Ruijin, LI Haitao, LU Guanghao
(College of Information Science and Technology, Qindao University of Science and Technology, Qingdao 266061, China)

Abstract: A lightweight underwater target detection network CSDP-L-YOLO for channel spatial depth perception is proposed. The network is improved based on the YOLOv5 network and consists of the feature awareness module and two-attention gating strategy. The feature sensing module aims to adaptively suppress or enhance multi-level features in the decoder, optimize the consistency of in-class learning, and solve the false detection and missing detection caused by the complexity of underwater scenes. The feature mapping is generated by the linear operation and mixing structure to reduce the fusion and calculation of redundant features, so as to reduce the parameters and computational complexity of the model. The dual attention gating strategy introduces both the concurrent channel space squeezing excitation mechanism module and convolutional attention module into the encoder simultaneously, so as to further focus on the strong correlation features and enhance the sensitivity of the model on the features. Experimental results show that compared with the baseline model, the proposed model improves the mean average precision (*mAP*) by 2.4%, saves the parameters by 20% and the computation by 15.8%, and improves the detection speed by 8.2 ms. In addition, compared to currently more advanced YOLOv8 model, the *mAP* of the proposed model improves by 1.9%.

Keywords: underwater target detection; channel spatial depth perception; attention mechanism; model lightweight; feature fusion; YOLO

0 引言

海洋是各国战略利益竞争的制高点,物质和能源资源丰富,同时也是人类赖以生存和发展的重要空间。对水下生物进行检测识别在海洋研究领域占据着不可动摇的地位。一方面,水下生物目标检测是海洋生物研究的重要手段之一,在海洋生物统计、水下养殖和环保等方面发挥着重要作用,并有着广泛应用。另一方面,由于水下光照条件差,待检目标重叠、遮挡的问题十分普遍[1],加上水下环境的复杂性和可变性[2],导致水下生物对比度、清晰度、能见度较低更增加了目标检测的挑战性。许多水下目标检测的研究都是基于声学检测方法[3],但由于小型水下生物的声源级较低,容易被背景噪声淹没,因此这些方法并不适用于探测小型水下生物。此外,声学探测方法的特征多样性可能无法满足区分水下生物之间微小差异的需求。因此,基于光学图像实现水下目标的准确定位和识别是开展科学研究和水下智能作业的难点和关键[3]。

传统目标检测流程分为两个步骤。先使用滑动窗口遍历待检图像并生成目标候选项区域。对目标区域进行特征提取后与相应的分类器结合后分类。这种方法的关键难点有两点,一是如何决定以滑动窗口区域选择为基础的策略,

收稿日期:2024-03-03; **修回日期**:2024-04-23。
基金项目:山东省重点研发计划(科技示范工程)(2021SFGC0701);青岛市海洋科技创新专项(22-3-3-hygg-3-hy)。
作者简介:赵瑞金(1998-),男,硕士。
通讯作者:李海涛(1978-),男,博士,副教授,硕士生导师。
引用格式:赵瑞金,李海涛,陆光豪.通道空间深度感知的轻量化水下目标检测[J].计算机测量与控制,2024,32(9):86-93.

且该方法时间复杂度高、针对性低，二是传统算子对水下图像的适应性不强，造成算法特征鲁棒性不佳。

深度学习等新技术为提高检测水平提供了重要动力[4]。基于深度学习的目标检测算法大致可以分为两大类[5-9]。Girshick 等人提出了一种简单、可扩展的检测算法，支持将大容量卷积神经网络 Convolutional Neural Networks（CNNs）应用于从底部到顶部提取区域线索，实现目标的精确定位与分割。同时提出了 R-CNN。两阶段算法涉及大量的重复计算操作[10]，导致推理速度较慢。

近年来，水下目标检测话题在国内逐渐变热，对该领域的研究和应用逐渐增加。2021 年 Xiangning Chen 等人提出的基于对抗学习的鲁棒和精确目标检测[11]，目标检测模型通常基于微调预训练分类器，进一步探索了对象探测器的微调空间，从分类和定位分支选择对抗样本，用于进一步提高模型准确率和鲁棒性。薛永杰[12]等人在 AlexNet 卷积神经网络的基础上，通过去除一些多余的卷积层来加速模型的学习，从而提高了模型的学习效率。在此基础上还提出了基于迁移学习的鱼群辨识模型。闫党康[13]提出了基于 Mask R _ CNN 的水下鱼类识别算法。

OverFeat[14]是最早开发的单级探测器之一。随后，YOLO 系列[15-19]在实际工程中表现出了较强的性能。YOLOv5 作为通用性和性能较强的一款模型，在 YOLO 系列算法中有很高的研究价值。同时，因其检测速度较快、准确度的优势，能在保持较高检测准确率值的同时满足及时性要求。YOLOv5 也被广泛运用于及时性较强的场景和工业检测当中。但是，在一般环境下有着较好表现的 YOLOv5 在模糊目标和特殊场景下的表现不尽人意。因此，通过改进 YOLOv5 提高水下目标的检测效果是很有研究意义的。

本文对上述问题进行研究，提出一种通道空间深度感知的轻量化底栖生物目标检测模型。首先，提出 CSDP 模块代替了原有模型 Neck 结构的 C3 模块。在 CSDP 模块内部引入卷积注意力机制 Convolutional Block Attention Module（CBAM），在空间和通道维度自适应地加强特征相关性，优化类内学习的一致性。其次，提出双注意门控策略，实现通道和空间感兴趣特征的多尺度深度感知，加强信息区分度和上下文交互能力。最后，使用组群洗牌策略卷积 Group Shuffle Convolution（GSConv）减少计算量，实现轻量化。经实验验证，与基线模型和其他经典算法相比，本问题出的模型在 URPC2020 数据集上有着更好的表现。

1　YOLOv5 目标检测算法简介

YOLO 系列代码自发布以来一直在进行更新迭代升级版本。本文基于 YOLOv5，比对原生的不同网络深度和训练复杂度的 s、m、l、x 四个版本，选择 YOLOv5-s 进行改进。YOLOv5-s 的网络结构包含 4 个部分：input（输入端）、Backbone（主干网络）、Neck（颈部）和 output（输出端）。主干网络采用 Focus 结构和瓶颈结构 Bottleneck Cross Stage Paritial（BCSP）模块构建。Focus 结构用于切片操作，以便于提取足够充分的图像特征，用减少计算量和原始图像信息的信息损失的方式提高模型训练速度。颈部结构采用 BCSP 和空间金字塔池化 Spatial Pyramid Pooling（SPP）构建特征金字塔 Feature Pyramid Network（FPN）＋路径聚合网络 Path Aggregation Network（PANet）的结构，BCSP 模块可以帮助网络有效地利用不同尺度的特征信息，增强模型对于输入图像的感知能力，同时减少了由于深度网络引入的梯度消失和过拟合等问题，FPN＋PAN 可以充分融合不同尺度的特征，确保特征图能包含图像的语义和位置信息。Prediction 部分采用 3 个特征层预测不同尺度的目标，输出一个包含目标对象的类别概率、预测权重和该目标对象边界框位置的向量。

2　CSDP-L-YOLO 模型

传统的卷积神经网络存在特征信息冗余，特征提取、融合能力不强，计算量大，模型参数量大，训练时间较长等缺点。YOLOv5-s 模型是 YOLOv5 系列模型中参数量和计算量最小的一个，这意味着其在检测精度方面表现不如

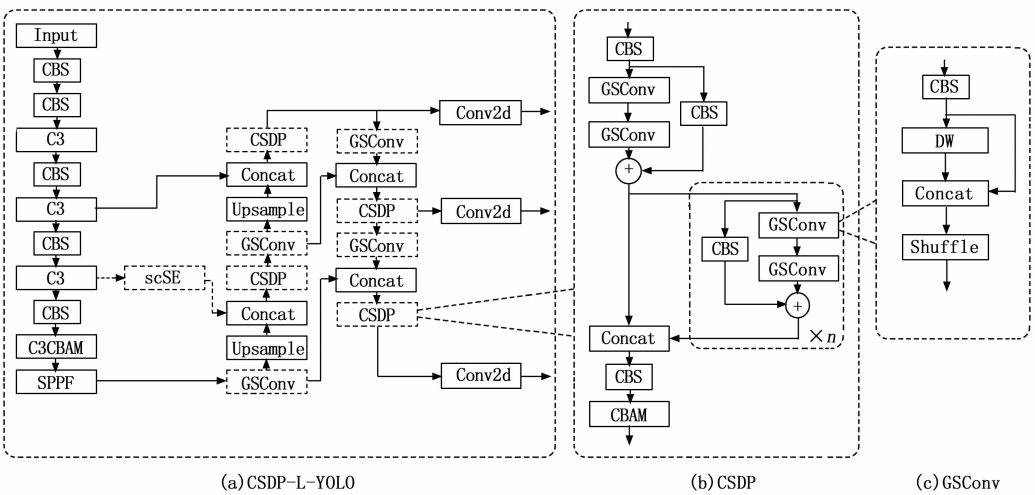


图 1　CSDP-L-YOLO 模型结构图

同系列的 n、m、l 算法。对于有挑战性的 URPC2020 水下数据集来讲，提升该算法的精度和实现算法轻量化设计是非常必要的。图 1（a）展示了改进后的通道空间深度感知的轻量化水下目标检测模型 Channel Spatial Depth Perception-Light-YOLO（CSDP-L-YOLO）模型结构图。改进内容如下所示：

1）提出 CSDP 模块代替了原有模型 Neck 结构的 C3 模块。在 CSDP 模块内部引入 CBAM 注意力机制，在空间和通道维度自适应地加强特征相关性，优化类内学习的一致性。

2）提出双注意门控策略，该策略包括通道空间挤压、激励模块和卷积块注意力模块，对主干网络深层特征提取的输出进行进一步处理，实现通道和空间感兴趣特征的多尺度深度感知，加强信息区分度和上下文交互能力。

3）使用 GSConv 卷积减少计算量，实现轻量化。

2.1 特征感知模块 CSDP 模块设计

YOLOv5 的颈部结构包含 PANet 结构，其作用是通过将来自不同特征层的信息融合在一起，实现提高检测精度的作用，是 Neck 部分中最重要的部分。不同特征层信息融合需要大量计算，不仅增加训练和推理的时间成本，增加模型的计算复杂度，还可能导致过拟合的情况，导致模型效率较低且检测结果与真实结果偏差较大。针对这些问题，提出了 CSDP 模块。

CSDP 模块参考 C3 模块设计，模块包含普通卷积模块、Bottleneck 结构（瓶颈结构）和注意力模块三部分。图 1（b）给出了 CSDP 模块的结构。CSDP 模块是一种类似 C3 的瓶颈结构，采用单次聚合方法[20] One-shot Aggregation（OSA）和模块融合注意力机制思想设计。只聚合每个模块的最后一层特征，也就是在每个模块的最后一层，对该模块的前面所有层的特征进行拼接，只进行一次聚合。在 CSDP 模块中使用标准卷积和深度可分离卷积 Depthwise Separable Convolution（DSC）加信道混洗机制，一方面保留 DSC 的优势，另一方面最小化 DSC 的缺陷对模型的影响。在减少计算量的基础上保留大部分有效特征，使得模型能够融合更多有效特征。在 CSDP 模块内添加 CBAM 注意力机制有助于强化空间和信道上的特征表达，增加模型对模糊背景、目标重叠场景下的检测能力。

DSC 可以极大程度上减少参数和浮点操作（FLOPs，浮点运算次数），实现轻量化的目的，DSC 结构如图 2 所示。DSC 主要分为两部分操作：逐深度卷积 Depthwise Convolution（DW）和逐点卷积 Pointwise Convlution（PW）。DW 卷积为了得到与输入特征矩阵的深度相同的特征矩阵，首先将卷积核拆分成逐个信道，再对每个信道进行单独卷积计算。PW 为卷积核大小固定为 1 的标准卷积。

注意力机制的思想来源于人类视觉模式并模仿人眼从复杂图像或复杂背景中提起典型特征的能力。注意力机制倾向作用于特征映射，生成二维注意权重和通道－空间两个独立维度的注意映射，并将权重作用到输入特征映射上，

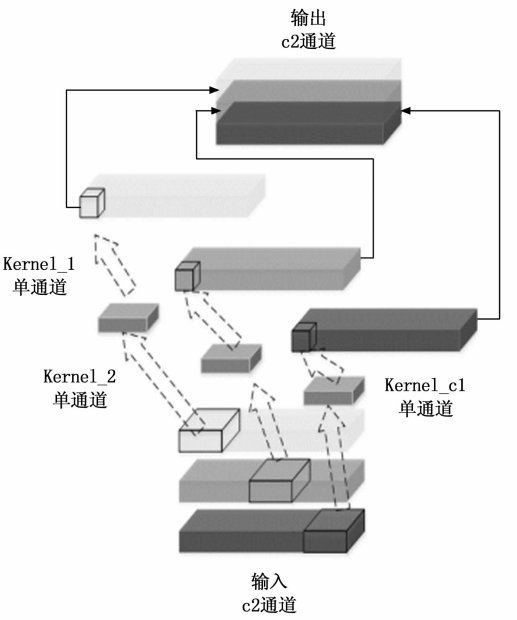


图 2 DSC 结构图

最后细化自适应特征。在模型恰当位置上添加注意力机制可以降低背景噪声带来的干扰，使得神经网络更关注于目标物体的显著特征区域，生成质量更优的特征图。在 CSDP 模块中添加 CBAM 注意力机制来提高目标检测准确率，对抗复杂环境和目标重叠带来的漏检和误检问题。CBAM 注意力机制结构如图 3 所示，包含两个独立的子模块，通道注意力模块和空间注意力模块，分别沿着两个独立的维度（通道和空间）依次推断注意力图，然后将注意力图与输入特征图相乘以进行自适应特征优化。

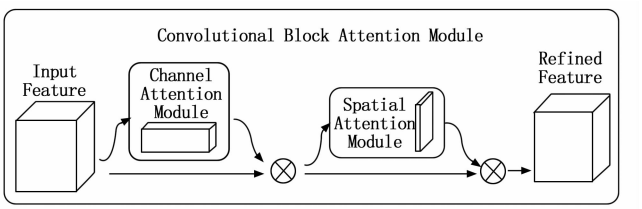


图 3 CBAM 注意力机制结构

先通过通道注意力机制，在空间维度上依次进行最大值池化与平均值池化，得到两个只有通道维度的向量，然后将这两个向量分别通过一个共享全连接层，两个特征相加后经过 sigmoid 函数，得到通道注意力向量。输入特征映射分别进行基于 w 的全局最大池化和全局平均池化。对共享的全连接层进行求和和 Sigmoid 激活操作，得到通道关注特征图，得到输出，计算公式如式（1）表示：

$$M_C(F)=\sigma\{MLP[AP(F)+MLP(MP(F)]\} \quad (1)$$

其中：$M_C(F)$ 表示通道注意机制的输出特征映射。空间注意机制以通道注意模块的输出作为输入，执行基于通道的全局最大池化和全局平均池化，将两个结果进行连接，通过卷积运算降维到一个通道，然后通过 sigmoid 生成空间

注意特征。空间注意机制的输出特征图公式如式（2）所示：

$$M_S(F)=\sigma(f^{7\times7}([AP(F);(MP(F)])) \tag{2}$$

其中：$M_S(F)$ 为二维空间注意图。式（1）和（2）中以中间特征图 $F\in R^{C\times H\times W}$ 为输入，CBAM 模块依次推导出一维通道注意图中间特征图 $M_C\in R^{C\times1\times1}$ 和二维空间注意图 $M_S\in R^{1\times H\times W}$，$AP$ 代表平均池化（AvgPooling），MP 代表最大池化（MaxPooling）。

提出的 CSDP 模块可以实现在减少计算量的前提下提高检测精度的效果。此外还可以根据 MobileNet 的思想，使用更多的卷积模块和瓶颈结构继续拼凑更多层数。参考深度可分离卷积的特点，结合实验验证，在残差结构的前提下，使用一个卷积层和一个 GSBottleneck 结构的 CSDP 模块可以很大程度上减少深度卷积的负面影响。

2.2　双注意门控策略设计

在主干网络尾部添加注意力机制的做法非常常见，这类改动在一定程度上避免了浅层网络重的底层语义信息对模型学习的干扰，强化了重要特征的表达。在模型进行特征融合时，将这主干网络最后一层特征提取的输出进行进一步处理，再进行特征融合，使得这些重要特征融合更加充分，加强信息区分度和上下文交互能力。

并发通道空间挤压和激励机制模块 Concurrent Spatial and Channel Squeeze and Excitation（scSE）是一个兼顾通道信息和模型上下文的注意力模块，它将通道与空间两个尺度的刺激元素叠加在一起，并在此基础上增加了对信道与信道两种重要的特征信息的权重。让模型学习更有意义的特征信息。与 CBAM 不同的是，scSE 由串行处理通道信息变更为并行处理通道信息。scSE 模块结构如图 4 所示。

左侧操作表示沿着空间维度来进行特征压缩，该部分采用两个全连接层的瓶颈结构，先降维，再恢复至原始维度，逐特征通道生成权重。整合整个空间特征当作一个全局特征，也就是将通道中的全局池化处理，可以通过式来表示：

$$Z\in R^{1\times1\times C}=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}U(i,j) \tag{3}$$

其中：包含全局信息的特征图规格为 $H\times W\times C$，输入特征映射 $U=[u_1,u_2,\cdots,u_C]$，将其直接压缩成一个 $1\times1\times C$ 的一维特征向量 $\boldsymbol{Z}$。其中 C 个特征图的通道特征都被压缩成为数值。在保证输出信道数目与输入信道数目相同的情况下，获得全局感受野。

之后，将输出的特征通过一个类循环神经网络中门的模块来抓取特征通道之间的关系。公式为：

$$s=\sigma[W2\mathrm{FReLU}(W1Z)] \tag{4}$$

其中：$W1\in R^{\frac{C}{r}\times C}$ 和 $W2\in R^{C\times\frac{C}{r}}$ 分别代表中间两个全连通层的权值。使用 FReLU 激活函数。S 为所得向量，用于重新校准并赋予特征映射 U 一个压缩后的权重。在通道维度上，对初始特征进行权值校正，对信息量较大的通道进行更多的处理，剔除信息量较小的通道。

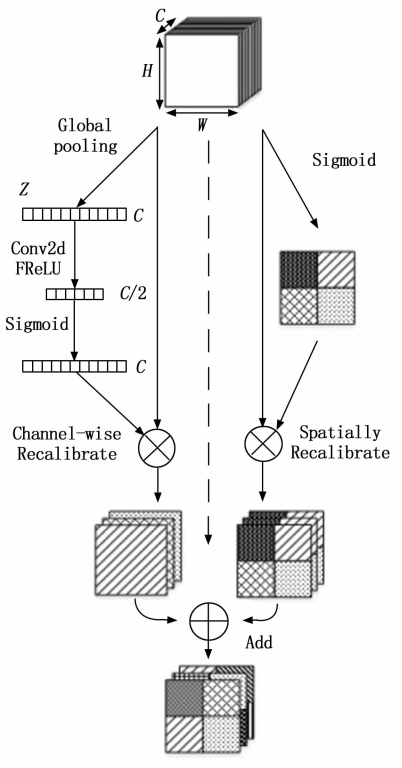


图 4　scSE 模块结构

模型图右侧是一个通道挤压和空间激励模块，沿着通道挤压特征图并进行空间激励。基于空间压缩的方法，先对投影张量进行卷积，然后将其各通道进行线性拼接，最后利用激活函数对其进行激活缩放。

此外，CSDP-L-YOLO 还在主干网络末端的引入 C3CBAM 模块，模块结构如图 5 所示。经过主干网络特征提取之后，通过注意力模块，再次强化重要特征表达，抑制次要特征对模型的影响。该模块与 scSE 模块构成双注意门控结构。

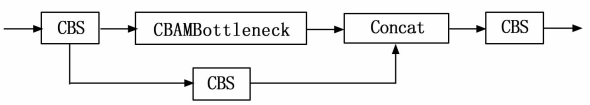


图 5　C3CBAM 模块结构

2.3　GSConv 卷积模块

轻量化设计可以有效地降低当前阶段的高计算成本。早些年推出的轻量级设计如 Xception[21]、MobileNets[22]、ShuffleNets[23]和 GhostNet[24]通过 DSC 卷积操作大大提高了检测器的检测速度，但或多或少影响了检测的准确率。实际上这些轻量级设计提出了一些方法来缓解 DSC 的缺点，其中 MobileNet 使用大量 1×1 密集卷积来融合通道信息，但大量的密集卷积反而占用了更多的计算资源；ShuffleNets 使用“通道洗牌”策略来实现通道信息的交互；GhostNet 使用特征图减半策略来保留通道信息，但依旧回归 SC 的计算方式。还有许多轻量级设计使用以上类似思路

来设计基本结构，即从头到尾只使用 DSC 深度神经网络，同时这也导致了 DSC 的缺陷都在模型的骨干网络中被直接放大。

通过通道洗牌策略生成的特征映射仍然可以进行深度分离操作。本文引入了 DSC 和混洗融合的模块——GSConv 卷积模块[25]，来达到经过 DSC 操作的输出尽可能接近 SC 的输出。GSConv 模块通过对多个通道的稠密卷积运算所产生的信息进行叠加，利用多个通道间的局部特征信息进行融合。GSConv 由三层组成：卷积层，批归一化层和激活层。蓝色标记的“DWConv”模块表示深度可分离卷积 DSC 操作。GSConv 模块结构如图 6 所示。

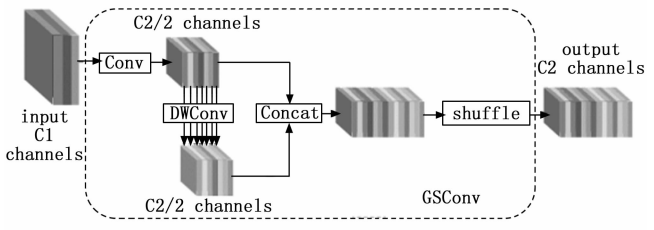


图 6 GSConv 模块结构

为了最终加快预测的计算速度，CNN 中的馈入图像几乎总是要在主干中经历类似的变换过程：将空间信息逐级向信道传递。而每次特征图的空间（宽度和高度）压缩和通道扩展都会造成部分语义信息的丢失。通道密集卷积计算最大限度地保留了每个通道之间的隐藏连接，而通道稀疏卷积完全切断了这些连接。一般情况下卷积计算的时间复杂度是由 FLOPs 来定义的。因此，SC（信道密集卷积）、DSC（信道稀疏卷积）和 GSConv 的无偏置时间复杂度如式（5）～（7）表示：

$$TimeSC \sim O(W \cdot H \cdot K_1 \cdot K_2 \cdot C_1 \cdot C_2) \tag{5}$$

$$TimeDSC \sim O(W \cdot H \cdot K_1 \cdot K_2 \cdot 1 \cdot C_2) \tag{6}$$

$$TimeGSConv \sim O\left[W \cdot H \cdot K_1 \cdot K_2 \cdot \frac{C_2}{2}(C_1+1)\right] \tag{7}$$

式中，W 为输出特征图的宽度，H 为输出特征图的高度，K_1、K_2 为卷积核的大小，C_1 为每个卷积核的通道数，也是输入特征图的通道数，C_2 为输出特征图的通道数。同等条件下，GSConv 的计算成本约为普通卷积的 50%（0.5+0.5 C_1，C_1 值越大，比值越接近 50%），但 GSConv 对模型学习能力的贡献与普通卷积相当。使用轻量化的卷积模块 GSConv 代替原有的卷积模块，使得既可以保证网络可以最大限度的保留每个通道之间的隐藏层链接，而不会把网络变得过于冗长。

3 实验结果及分析

3.1 数据集选用

本文所有实验都在 Underwater Robot Professional Contest 2020（URPC2020）数据集上实现。该数据集包含 4 个类别：海参、海胆、扇贝和海星，共有分辨率为 1 920×1 080的真实水下图像 5 543 张。将 VOC 格式转换为 YOLO 适配格式，按照训练集、验证集、测试集 8：1：1 的比例划分数据集。

3.2 实验环境与参数设置

实验所需的硬件及环境配置如表 1 所示。

表 1 硬件及环境配置

配置	版本
操作系统	linux
CPU	Intel(R) Xeon(R) Gold 6326 CPU @ 2.90 GHz
GPU	NVIDIA GeForce GTX 3090
Python	3.8.16
PyTorch	1.10.1
CUDA	11.6

实验的部分关键参数如表 2 所示。

表 2 关键超参数配置

超参数	值	超参数	值
初始学习率	0.01	预训练权重	YOLOv5－s
优化器	SGD	训练图像大小	640×640
IOU 阈值	0.2	训练轮次	200
权重衰减	0.000 5	batch_size	16
学习率动量	0.937	预热学习动量	0.8
Anchor 长宽比	4.0	Mosaic 增强	1

3.3 实验衡量指标

本文实验采用的衡量标准包括准确率 P（Precision）、召回率 R（Recall）、平均精确率均值 mean Average Precision（mAP）、计算量（GFLOPs）、参数量（Params）和推理速度作为评估指标。P 是是预测出的真正正例在预测出的所有正例中所占比例；召回率 R 是指预测出的真正正例在所有真实正例中所占比例；mAP 是对精确率－召回率曲线（PR 曲线）和坐标轴之间所包围的面积上的 Precision 值求平均值，一般使用积分法计算。计算公式如式（8）、（9）所示：

$$xAP = \int_0^1 P(R) \tag{8}$$

$$mAP = \frac{1}{t}\sum_{i=1}^{t} xAP_i \tag{9}$$

其中：xAP 代表平均精确度，xAP_i 代表第 i 类目标检测的平均精确度，t 代表标记的类别。

$mAP@0.5$ 的公式如式（10）所示，其定义为在交并比阈值 IOU 为 0.5 的情况下的平均精确度均值：

$$mAP@0.5 = \frac{1}{t}\sum_{i=1}^{t} xAP@0.5i \tag{10}$$

其中：$xAP@0.5i$ 代表第 i 类目标在交并比阈值为 0.5 时的平均精确度，t 代表标记的类别。

实验中对精确度的计算如式（11）所示，召回率的计算公式如式（12）所示：

$$P = \frac{TP}{TP+FP} \tag{11}$$

$$R = \frac{TP}{TP + FN} \tag{12}$$

其中：TP（True Positive）为真正例，代表真实情况和预测情况同时为正例。FP（Flase Positive）为假正例，代表真实为反例而预测为正例。FN（Flase Negative）为假反例，代表真实情况为正例而预测为反例。

计算量和参数量计算公式如式（13）、（14）所示：

$$GFLOPs = W \cdot H \cdot K \cdot K \cdot C_{in} \cdot C_{out} \tag{13}$$

$$Params = C_{in} \cdot C_{out} \cdot K \cdot K \tag{14}$$

其中：W 和 H 分别为输入特征图的宽度和高度，K 为卷积核大小，C_{in}和 C_{out}分别为输入和输出的特征信道数。

4　实验结果与分析

实验结果分为两部分。第一部分为不同模型对比实验，主要分析改进后的网络与经典目标检测网络和目前常用目标检测网络的训练效果。第二部分为消融实验，主要体现不同改进模块在原始网络上的涨幅差异。同时设计了大量模块组合实验以证明改进后的网络在性能和结果上具有优势。

4.1　数据预处理

拼接数据增强是一种将数据集中任意四张图片数据进行混合、拼接和裁剪，从而得到一幅新图片数据的方法。结果包含了更丰富的目标信息，在一定程度上扩展了训练数据，使得网络可以在数据集上进行更充分的训练。图 7 显示了马赛克增强后的图像。

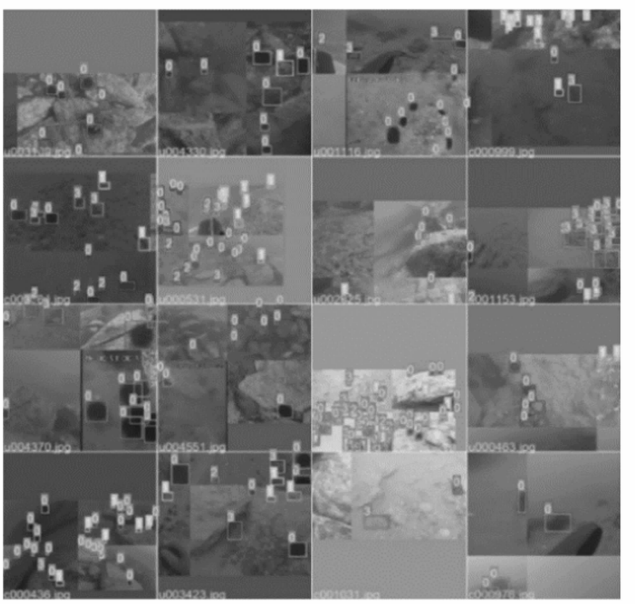


图 7　在训练中使用马赛克增强图像

4.2　不同模型对比分析

本部分实验选取经典的 Faster RCNN、SSD、RetinaNet、YOLOv3、YOLOv4、YOLOv5 系列模型和较新的 YOLOv8 模型进行比较。经实验验证模型的 mAP 较 Faster R-CNN、SSD、RetinaNet 有较大提升，mAP 均提升 30% 左右。与 YOLOv5-s 模型相比，在海胆类上上涨了 0.5%，在扇贝类上上涨了 1%，在海星类上上涨了 9.1%，在所有类别上的 mAP 上涨了 2.4%。虽然模型的整体性能得到提上，但改进后模型在海参上表现略有下降。经典算法对照实验结果如表 3 所示。

表 3　经典算法对照实验结果对比

模型名称	AP/%				mAP/%
	海参	海胆	扇贝	海星	
Faster R-CNN	22.6	73.8	66.6	44.3	51.8
SSD	35.2	75.0	34.3	67.8	53.1
RetinaNet	37.2	76.8	52.7	69.4	59.0
YOLOv3	57.3	87.8	66.8	79.1	72.8
YOLOv4	71.9	89.3	85.9	77.4	81.1
baseline(YOLOv5-s)	75.4	91.4	87.0	81.3	83.8
YOLOv5-n	64.4	90.8	83.5	88.0	81.7
YOLOv5-m	72.4	90.7	86.9	88.8	84.7
YOLOv5-l	72.0	91.1	87.0	89.0	84.8
YOLOv8	77.3	91.8	79.6	90.3	84.8
CSDP-L-YOLO(本文)	74.5	**91.9**	**88.0**	**90.4**	**86.2**

由于本文利用了 scSE 注意力，为确定 scSE 是该基线模型改进下最优模块，本小节将展示 scSE 注意力模块与一些其他常见注意力模块加入到网络同一位置的效果差距。实验效果对比如表 4 所示。

表 4　常见注意力模块对照实验结果对比

模型名称	AP/%				mAP/%
	海参	海胆	扇贝	海星	
YOLOv5＋simAm[26]	73.9	**91.9**	86.4	86.6	84.7
YOLOv5＋SGAMAttention	72.1	89.2	84.3	85.6	82.8
YOLOv5＋S2-FPN[27]	73.7	91.3	83.5	88.3	84.2
YOLOv5＋CA[28]	72.6	89.3	87.2	86.9	84.0
YOLOv5＋SE[29]	71.9	91.4	86.8	89.1	84.8
scSE(本文)	**74.0**	**90.8**	**88.0**	**89.3**	**85.5**

4.3　消融实验结果分析

本小节对所提方法的单个模块进行消融实验，以 YOLOv5 为基线模型，来评估各个组件是否有效。首先，将每个模块依次插入到基线模型中，结果比较如表 5 所示。

表 5　单个模块消融实验结果对比

模型名称	P/%	R/%	mAP/%	参数量/M	速度/ms	FLOPs(G)
YOLOv5	**94.3**	92	84.0	7.07	52.1	16.5
＋CBAM	92.1	93	85.5	6.71	49.5	15.7
＋GSConv	88.2	**94**	84.7	6.58	46.9	15.4
＋scSE	88.6	93	85.2	6.48	46.5	15.8
＋CSDP	91.7	93	84.9	6.87	50.1	15.9
CSDP-L-YOLO(本文)	93.1	**94**	**86.2**	**5.64**	**43.9**	**13.9**

其次，根据模块的计算方式，选择将同一模块嵌入到

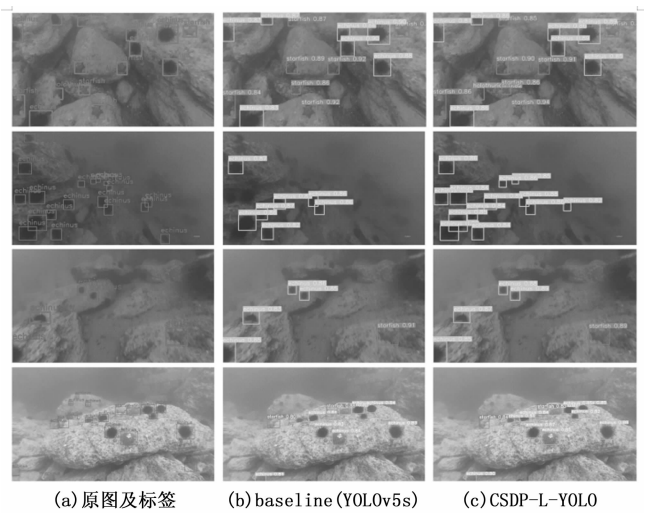
网络的不同位置来确定模块的最优位置。将不同数量的改进模块嵌入网络，来测试改进后网络的最优表现。不同结构模型结果对比实验如表 6 所示。

表 6 不同结构模型结果对比

模块名称	Backbone	Head	数量	参数量/M	*mAP*/%
GSConv	√	×	1	6.89	84.0
	√	×	4	6.62	84.5
	×	√	4	6.59	84.7
CSDP-Neck	√	×	4	5.69	84.5
	×	√	4	5.82	85.3
CSDP-Bottleneck	×	×	1	5.87	85.5
	×	×	2	6.06	85.3
	×	×	3	6.16	84.9
scSE	第二层	—	1	6.54	84.2
	第四层	—	1	6.71	84.6
	第六层	—	1	5.64	84.8

4.4 推理结果可视化比较

为了证明所提 CSDP-L-YOLO 模型的有效性和鲁棒性，本小节旨在对比分析基线模型和 CSDP-LYOLO 模型在 URPC2020 数据集上的可视化结果，如图 8 所示。

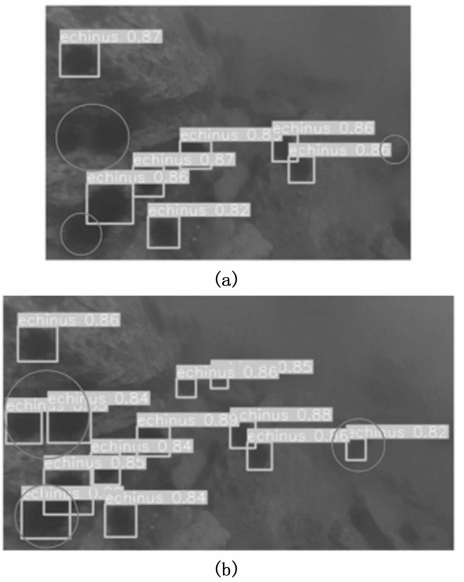


(a)原图及标签 (b)baseline(YOLOv5s) (c)CSDP-L-YOLO

图 8 模型推理结果对比

在图 8 中，(a) 列为验证集中的原图和标定准确的目标标签；(b) 列展示了基线模型的检测结果，在 (b) 列第二行中可以看出，基线模型未检测出图片左侧受背景影响较严重的待检目标和图片右侧重叠的待检目标；(c) 列是 CSDP-L-YOLO 模型训练得到的结果。与 (b) 列相比，CSDP-L-YOLO 模型可以适应海底光线暗、能见度低等复杂环境的干扰，较好的避免了密集目标和小目标的漏检问题，减小了由目标重叠和光线条件差对小目标的影响。详细情况如图 9 所示。

其中，(a) 为基线模型检测结果，(b) 为 CSDP-L-YOLO 模型推理结果。通过标记处对比可以看出，CSDP-L-YOLO 模型解决了大部分漏检问题。

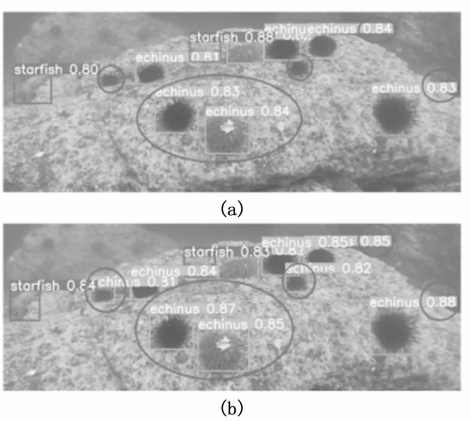


(a)

(b)

图 9 漏检问题优化

此外，CSDP-L-YOLO 模型大幅度提高了海星、海胆和扇贝类别的检测置信度，CSDP-L-YOLO 模型对于误检问题有较高的解决效果。详细情况如图 10 所示。



(a)

(b)

图 10 检测置信度提升

其中 (a) 为基线模型检测结果，(b) 为 CSDP-L-YOLO 模型推理结果，展示了在各类待检目标上的置信度的提升。

5 结束语

针对水下光照条件较差、图像模糊、目标重叠覆盖等问题，本文基于 YOLOv5-s 算法，提出了一种轻量化水下目标检测模型 CSDP-L-YOLO。通过特征感知模块和双注意门控策略，增强了模型对通道和空间感兴趣特征的深度表示和感知能力。此外，所提模型在提升平均精确度和推理过程中的置信度的同时，加快了检测速度，减少了参数量和计算量。经实验验证，所提模型对于水下的模糊目标和细小目标减少了漏检的情况，在水下目标检测领域具有应用和借鉴价值。接下来将针对不同目标检测置信度差距较

大问题进一步改善模型，研究探索不同损失函数来提升水下少样本目标的检测精度，进一步提高模型在面对其他水下目标检测任务的泛化性，使得模型能够可靠的完成更多水下复杂环境中的目标检测工作。

参考文献：

[1] SIMONYAN K，ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. ArXiv Preprint ArXiv：1409.1556，2014.

[2] JIAN M，LIU X，LUO H，et al. Underwater image processing and analysis：A review [J]. Signal Processing：Image Communication，2021，91：116088.

[3] SARKAR P，DE S，GURUNG S. A survey on underwater object detection [M]. Singapore：Springer Singapore，2022：91－104.

[4] 张志强，牛智有，赵思明. 基于机器视觉技术的淡水鱼品种识别 [J]. 农业工程学报，2011，27 (11)：388－392.

[5] BARAT C，PHLYPO R. A fully automated method to detect and segment a manufactured object in an underwater color image [J]. EURASIP Journal on Advances in Signal Processing，2010：1－10.

[6] UIJLINGS J R R，VAN DE SANDE K E A，GEVERS T，et al. Selective search for object recognition [J]. International Journal of Computer Vision，2013，104：154－171.

[7] GIRSHICK R. FAST R-CNN [C] //Proceedings of the IEEE International Conference on Computer Vision，2015：1440－1448.

[8] HE K，GKIOXARI G，DOLLáR P，et al. Mask r-cnn [C] //Proceedings of the IEEE International Conference on Computer Vision，2017：2961－2969.

[9] CAI Z，VASCONCELOS N. Cascade r-cnn：Delving into high quality object detection [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition，2018：6154－6162.

[10] DENG J，XUAN X，WANG W，et al. A review of research on object detection based on deep learning [C] //Journal of Physics：Conference Series. IOP Publishing，2020，1684 (1)：012028.

[11] CHEN X，XIE C，TAN M，et al. Robust and accurate object detection via adversarial learning [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition，2021：16622－16631.

[12] 薛永杰，巨志勇. 基于改进 AlexNet 的鱼类识别算法 [J]. Electronic Science & Technology，2021，34 (4).

[13] 闫党康. 基于深度学习的水下鱼类识别 [硕士论文]. 北方工业大学，2021.

[14] SERMANET P，EIGEN D，ZHANG X，et al. Overfeat：Integrated recognition，localization and detection using convolutional networks [J]. ArXiv Preprint ArXiv：1312.6229，2013.

[15] REDMON J，DIVVALA S，GIRSHICK R，et al. You only look once：Unified，real-time object detection [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition，2016：779－788.

[16] REDMON J，FARHADI A. YOLO9000：better，faster，stronger [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition，2017：7263－7271.

[17] REDMON J，ARHADI A. Yolov3：An incremental improvement [J]. ArXiv Preprint ArXiv：1804.02767，2018.

[18] BOCHKOVSKIY A，WANG C Y，LIAO H Y M. Yolov4：Optimal speed and accuracy of object detection [J]. ArXiv Preprint ArXiv：2004.10934，2020.

[19] 李庆忠，李宜兵，牛　炯. 基于改进 YOLO 和迁移学习的水下鱼类目标实时检测 [J]. 模式识别与人工智能，2019，32 (3)：193－203.

[20] LEE Y，HWANG J，LEE S，et al. An Energy and GPU-Computation Efficient Backbone Network for Real-Time Object Detection [C] //2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE，2019：752－760.

[21] CHOLLET F. Xception：Deep learning with depthwise separable convolutions [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition，2017：1251－1258.

[22] HPWARD A G，ZHU M，CHEN B，et al. Mobilenets：Efficient convolutional neural networks for mobile vision applications [J]. ArXiv Preprint a ArXiv：1704.04861，2017.

[23] ZHANG X，ZHOU X，LIN M，et al. Shufflenet：An extremely efficient convolutional neural network for mobile devices [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition，2018：6848－6856.

[24] HAN K，WANG Y，TIAN Q，et al. Ghostnet：More features from cheap operations [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition，2020：1580－1589.

[25] LI H，LI J，WEI H，et al. Slim-neck by GSConv：A better design paradigm of detector architectures for autonomous vehicles [J]. ArXiv Preprint ArXiv：2206.02424，2022.

[26] YANG L，ZHANG R Y，LI L，et al. Simam：A simple，parameter-free attention module for convolutional neural networks [C] //International Conference on Machine Learning. PMLR，2021：11863－11874.

[27] ELHASSAN M A M，YANG C，HUANG C，et al. S^2-FPN：Scale-ware strip attention guided feature pyramid network for real-time semantic segmentation [J]. ArXiv E-Prints，2022：ArXiv：2206.07298.

[28] HOU Q，ZHOU D，FENG J. Coordinate attention for efficient mobile network design [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition，2021：13713－13722.

[29] HU J，SHEN L，SUN G. Squeeze-and-excitation networks [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition，2018：7132－7141.