

## 面向嵌入式平台的光学遥感图像舰船检测识别

何鑫宇, 陆陈鑫, 冯书谊, 欧阳尚荣, 穆文涛

### 引用本文

何鑫宇, 陆陈鑫, 冯书谊, 欧阳尚荣, 穆文涛. 面向嵌入式平台的光学遥感图像舰船检测识别[J]. 计算机科学, 2024, 51(6A): 230700117-7.

HE Xinyu, LU Chenxin, FENG Shuyi, OUYANG Shangrong, MU Wentao. [Ship Detection and Recognition of Optical Remote Sensing Images for Embedded Platform](#) [J]. Computer Science, 2024, 51(6A): 230700117-7.

---

### 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

#### Similar articles recommended (Please use Firefox or IE to view the article)

#### [基于SAMNV3的滚动轴承智能故障诊断方法](#)

Intelligent Fault Diagnosis Method for Rolling Bearing Based on SAMNV3

计算机科学, 2024, 51(6A): 230700167-6. <https://doi.org/10.11896/jsjcx.230700167>

#### [基于LSTM和注意力机制的远程会诊需求预测](#)

Forecasting Teleconsultation Demand Based on LSTM and Attention Mechanism

计算机科学, 2024, 51(6A): 230800119-7. <https://doi.org/10.11896/jsjcx.230800119>

#### [融入类别标签和主题信息的用户兴趣识别方法](#)

User Interest Recognition Method Incorporating Category Labels and Topic Information

计算机科学, 2024, 51(6A): 230500169-8. <https://doi.org/10.11896/jsjcx.230500169>

#### [基于注意力的多尺度蒸馏异常检测](#)

Attention-based Multi-scale Distillation Anomaly Detection

计算机科学, 2024, 51(6A): 230300223-11. <https://doi.org/10.11896/jsjcx.230300223>

#### [基于改进DeepLabv3+算法的滚珠丝杠驱动表面点蚀缺陷检测](#)

Detection of Pitting Defects on the Surface of Ball Screw Drive Based on Improved DeepLabv3+ Algorithm

计算机科学, 2024, 51(6A): 240200058-6. <https://doi.org/10.11896/jsjcx.240200058>

# 面向嵌入式平台的光学遥感图像舰船检测识别

何鑫宇 陆陈鑫 冯书谊 欧阳尚荣 穆文涛

上海航天电子技术研究所 上海 201109

(hexinyu5688@163.com)

**摘要** 建设海洋强国是我国当前大力发展的战略方向。针对现有基于深度学习的遥感图像舰船目标检测识别算法在嵌入式平台上存在检测分类识别率低、运行速率慢等问题,提出了一种基于寒武纪 MLU220 嵌入式平台改进的 Mix-YOLO 网络模型。该模型以 YOLOv7-tiny 网络为基本框架,首先,引入 MobileNet 系列网络模块对特征提取网络进行部分替换,减少网络参数量;然后,引入 ULSAM 注意力机制,以便增强网络学习分类能力,减少虚警率;最后,为了显著提升嵌入式平台检测速率,采用将网络模型大模块拆分为小模块的编写方式。实验结果表明:Mix-YOLO 算法在原 YOLOv7-tiny 网络基础上,参数量和计算量分别下降了 39.70% 和 29.70%,处理帧率由 97.27 fps 提升至 120.88 fps,精度提高了 7.7%,能够实现遥感图像中舰船目标的实时检测识别。

**关键词:** 光学遥感图像;舰船检测;轻量级网络;注意力机制;嵌入式平台

**中图分类号** TP751

## Ship Detection and Recognition of Optical Remote Sensing Images for Embedded Platform

HE Xinyu, LU Chenxin, FENG Shuyi, OUYANG Shangrong and MU Wentao

Shanghai Aerospace Electronic Technology Research Institute, Shanghai 201109, China

**Abstract** The construction of a maritime power is a current strategic direction for China's vigorous development. In response to the low detection and classification recognition rate and slow operation speed of existing deep learning-based remote sensing image ship target detection and classification algorithms on embedded platforms, this paper proposes an improved Mix-YOLO network model based on the Cambricon-MLU220 embedded platform. The model is based on the YOLOv7-tiny network as the basic framework. Firstly, the MobileNet series network module is introduced to replace the feature extraction network partially, reducing the network parameter volume. Then, the ULSAM attention mechanism is introduced to enhance the network's learning and classification ability, reducing the false alarm rate. Finally, in order to make the detection speed improvement effect more obvious on the embedded platform, the network model is programmed by splitting the large module into small modules. Experimental results show that the Mix-YOLO algorithm reduces the parameter volume and calculation by 39.70% and 29.70%, respectively, on the basis of the original YOLOv7-tiny network. The processing frame rate is increased from 97.27 fps to 120.88 fps, and the accuracy is improved by 7.7%. It can achieve real-time detection and recognition of ship targets in remote sensing images.

**Keywords** Optical remote sensing image, Ship detection, Lightweight network, Attention mechanism, Embedded platform

## 1 引言

海洋权益已成为国家重要核心利益,海上监控技术是维护我国海洋权益的重要手段,舰船是海上监控任务的重要目标。随着航天遥感技术的发展,光学遥感图像可获取信息越来越丰富,这为研究舰船检测识别技术提供了数据支撑,同时也要求检测硬件平台具备快速处理图像的能力。基于 AI 算力的嵌入式平台性能强大,虽满足了海上监控的需求,但实际应用中存在以下难点:1)图像背景复杂,容易对舰船检测造成干扰;2)算法复杂,影响硬件平台的处理速率。因此,面向嵌入式平台的光学遥感舰船检测识别已成为目前重要的应用研究方向。

舰船检测识别方法分为传统方法和深度学习方法。传统方法首先对图像进行海陆分割处理,其次利用图像分析等方法获取舰船目标区域,最后提取舰船特征。近年来,随着深度

学习技术发展,研究者提出了基于深度学习的目标检测识别方法,主要分为两阶段方法<sup>[1-3]</sup>(Two Stage)和一阶段方法<sup>[4-5]</sup>(One Stage)。两阶段算法识别精度高,但网络参数和计算量大,推理速度慢,无法满足实时处理需求。一阶段方法分为基于锚框定位算法和无锚框定位算法。YOLO<sup>[6-10]</sup>系列网络是目前流行的检测识别模型,它主要包括 v1-v7 网络,其中 YOLOv5 在硬件部署上得到广泛应用,YOLOv7 网络在 v5 基础上轻量化无锚框算法。

以上大多数算法已经得到了应用。Cui 等<sup>[11]</sup>将深度可分离卷积模块应用于 YOLO 算法框架。但未能在遥感图像目标上得到验证。Su 等<sup>[12]</sup>提出了定向可分离卷积模块,并应用于 YOLO 系列网络中。Yu 等<sup>[13]</sup>结合灰度共生矩阵,将图像分割为若干子图,以稀疏 MobileNetV2 作为骨干网络进行特征提取,算法流程复杂。Nong 等<sup>[14]</sup>通过精简 YOLOv3-tiny 网络结构和引入空间注意力模块改进了小目标检测效果;

Liu 等<sup>[15]</sup>在 YOLOv5 模型中加入 GhostNet 结构优化了提取特征网络; Wang 等<sup>[16]</sup>将 ShuffleNet v2 用于 YOLOv5s 模型的特征提取网络; Liu 等<sup>[17]</sup>在 YOLOX-s 模型中加入位置注意力模块,改善了特征融合阶段网络。但上述方法均存在以下问题:1)网络参数冗余,图像数据相同的情况下计算量大;2)硬件支持度低,处理速度慢。

基于以上分析,本文提出面向嵌入式平台的轻量级 Mix-YOLO 舰船检测分类识别算法。该算法以 YOLOv7-tiny 为基础框架,首先,针对网络深层特征提取网络结构复杂的问题,引入 MobileNet-block 模块;其次,对原激活函数进行优化,将 h\_swish 激活函数替换为 Leaky\_ReLU 函数,提升了

硬件处理速率;最后,引用 ULSAM 注意力模块加入至最后一层骨干网络,减少了通道信息特征冗余。

## 2 Mix-YOLO 算法

YOLOv7-tiny 网络是 YOLOv7 的简化版本,是部署到硬件计算平台的选择之一。尽管网络复杂度大大减小,但 YOLOv7-tiny 模型也存在问题:采用了大量 ELAN 模块,每个 ELAN 模块内部由标准卷积层密集连接构成,在对舰船等大目标特征进行提取时容易造成特征冗余和存在数据溢出风险。本文设计的轻量级 Mix-YOLO 舰船检测识别算法如图 1 所示。

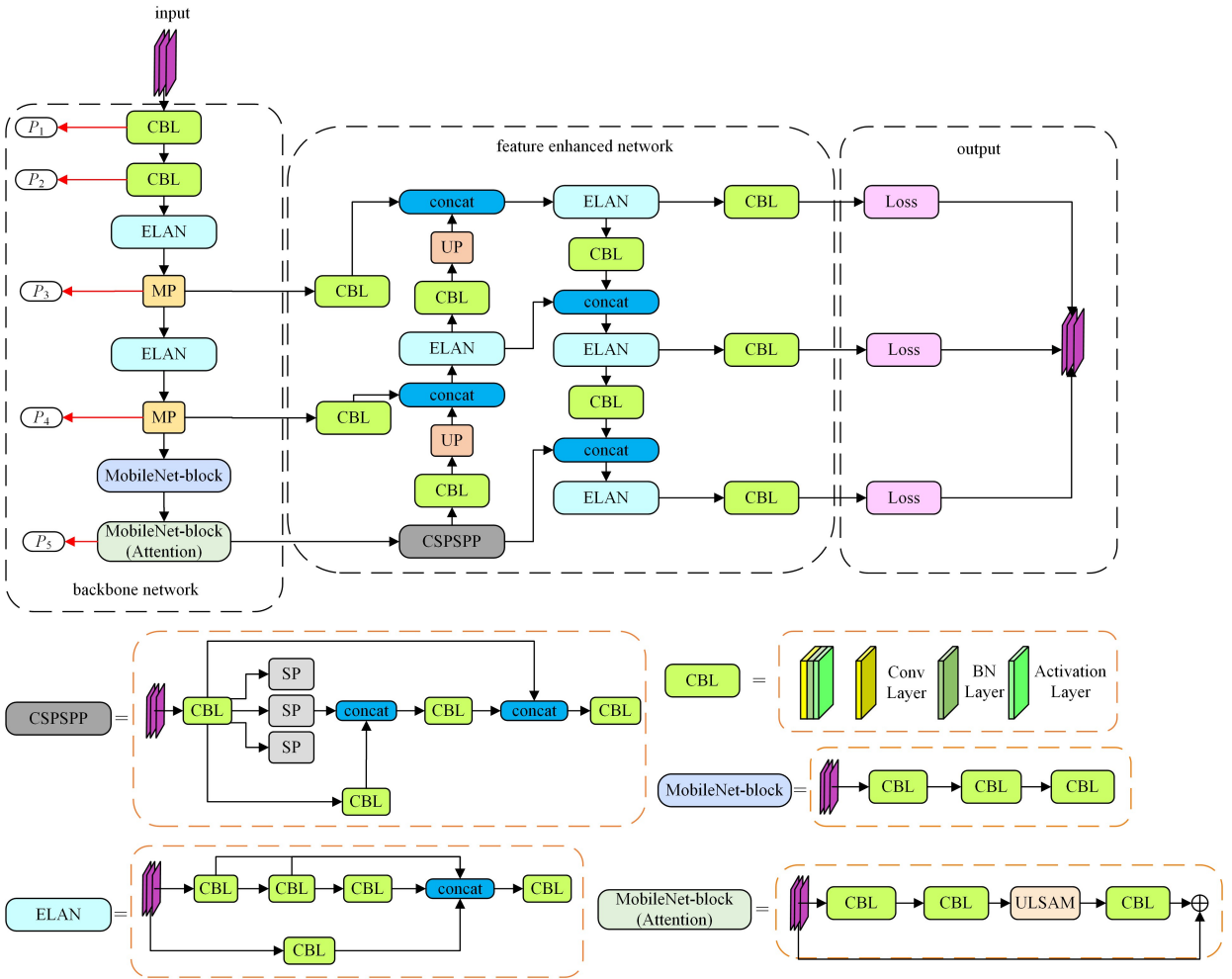


图 1 Mix-YOLO 网络结构图

Fig. 1 Structure diagram of Mix-YOLO

网络在输入端采用了 YOLOv7-tiny 的数据增强方法,除了对输入的训练图像进行剪裁、旋转、翻转和色调调整之外,还进行了基于马赛克(Mosaic)的图像数据扩充,将 4 幅经过不同数据增强方式后的图像进行拼接,扩大了训练图像数量的同时也丰富了舰船的图像背景,增强了网络检测识别的鲁棒性。

经过数据增强后,对原图像进行 Resize 操作统一,以  $1536 \times 1536 \times 3$  的大小输入至特征提取网络。特征提取网络总共由 6 个提取模块组成,第一、第二模块为卷积操作模块(CBL),CBL 由步长为 2 的卷积网络层、批归一化(BN)层和激活函数层组成,用于提取舰船的低层特征  $P_1$  和  $P_2$ ;第三、第四模块保留了 YOLOv7-tiny 的 ELAN 模块,用于提取舰船

的中层特征  $P_3$  以及中高层  $P_4$  特征,ELAN 模块分为两个并行网络,网络一和网络二都对特征图进行了通道调整,之后网络一对图像进行特征提取,将提取结果与网络二的输出结果进行连接操作(concatenate),最后对连接结果执行卷积操作和池化操作(MP)得到  $P_3$  和  $P_4$  特征;第五模块为 MobileNet-block 模块,第六模块为加入 ULSAM 注意力的 MobileNet-block 模块,第五模块进一步提取  $P_4$  特征,第六模块中的 ULSAM 注意力模块对通道注意力重新进行了分配,使得高层特征  $P_5$  对舰船的种类信息具有较强的表征能力。

各层级特征会在特征融合增强网络中对特征进行学习 and 整合。特征融合增强网络沿用了 YOLOv7-tiny 的

网络结构,包括 CBL、CSPSP、上采样操作模块(UP)。按照传统的 PAFPN 网络设计思想,网络对  $P_3$  和  $P_4$  以及  $P_5$  中层次不同、尺度不同的特征进行提取,与融合增强网络中学习得到的特征进行不同尺度的充分融合,不同层次的特征经过通道连接(concatenate)融合后会输入对应的检测头进行预测。

检测头是 Mix-YOLO 网络输出端的主要部分。它结合了 anchor 机制,根据特征融合增强网络得到的  $P_3$ 、 $P_4$  和  $P_5$  特征信息,输出对应层级的舰船目标特征图。同时利用常规卷积对通道数进行调整,再通过 CIOU 损失函数和非线性极大值抑制(NMS)处理,得到最终预测效果。

## 2.1 骨干网络模块 MobileNet-block

深度可分离卷积(Depthwise Separable Convolution, DSC)可以大大降低网络计算量和参数量,被广泛应用于轻量级网络设计中。典型的 DSC 网络就是 MobileNetV1-V3 和 ShuffleNetV1-V2,其中 MobileNetV3<sup>[18]</sup>和 ShuffleNetV2 性能较为突出。ShuffleNetV2 使用 Channel Split 操作,实现了更好的检测精度和速度,但是该操作在 AI 板卡上耗时长,对于大部分 AI 设备不够友好。MobileNetV3 是 MobileNet 系列最新的版本,也是性能最优的版本。MobileNetV3 模块结构如图 2 所示。

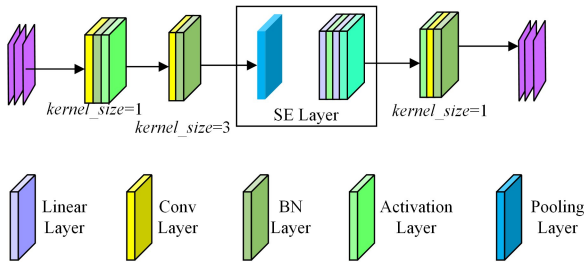


图 2 MobileNetV3 模块

Fig. 2 MobileNetV3 block

DSC 将标准卷积分解为深度卷积和逐点卷积。深度卷积里的卷积核只有一维,负责对不同通道的图像单独进行卷积处理,假设深度卷积的卷积核大小为  $k \times k \times 1$ ,其中有  $M$  个卷积核。逐点卷积的卷积核大小为  $1 \times 1 \times M$ ( $M$  为图像的

通道数),其中有  $N$  个卷积核,将之前的特征图在通道方向进行升维和降维操作。每个深度可分离卷积由一层深度卷积和一层逐点卷积构成。深度可分离卷积核的参数数量计算如下:

$$A_1 = M \times (k^2 + N) \quad (1)$$

假设标准卷积的输入特征图为  $H \times W \times M$  大小,经过  $N$  个  $k \times k \times M$  的标准卷积后得到输出特征图的尺寸为  $H \times W \times M$  大小。则标准卷积参数数量为:

$$A_2 = k^2 \times M \times N \quad (2)$$

深度可分离卷积和标准卷积参数量相比的计算式如下:

$$\frac{A_1}{A_2} = \frac{M \times (k^2 + N)}{k^2 \times M \times N} = \frac{1}{N} + \frac{1}{k^2} \quad (3)$$

MobileNetV3 模块首先对特征图通过  $1 \times 1$  的逐点卷积进行升维操作和  $3 \times 3$  的深度卷积处理,再利用  $1 \times 1$  的逐点卷积进行降维操作。这样能够保证信息流通过 h\_swish 非线性激活函数时减少信息损失。在网络通路中还添加了轻量级的注意力模块 SE<sup>[19]</sup>(Squeeze and Excitation)。

MobileNetV3 模块的 SE 注意力模块结构过于简单,无法对复杂场景下的舰船识别实现特征通道筛选作用,因此在已加入 SE 模块的 MobileNetV3 模块中去掉该模块,最终得到 MobileNet-block 模块。

虽然 MobileNet-block 模块能使网络达到减少参数量的效果,但对于骨干网络而言,它并不能保证一些特征信息能够保留。这是因为 MobileNet-block 模块采用了大量 DSC,网络参数数量不足,对于舰船这种尺度差异较大的目标提取特征有限。在特征提取阶段,低层特征  $P_1$  和  $P_2$  主要包括舰船的几何、纹理以及位置信息,中高层的语义特征  $P_3$ 、 $P_4$ 、 $P_5$  则是从低层特征中继续提取得到的。如果低层特征无法得到充分提取,对于中高层语义特征的获取也是不利的,会直接影响后续特征增强融合效果。同时,考虑到 YOLOv7-tiny 网络的 ELAN 结构在高级特征提取中存在的特征通道冗余问题,应该对特征增强融合前的骨干网络进行 MobileNet-block 模块替换。如图 3 所示,本文将特征提取网络进行改进,只将提取高层特征的骨干网络替换成 MobileNet-block,保留了原提取低层特征的骨干网络,也保证了网络检测识别精度。

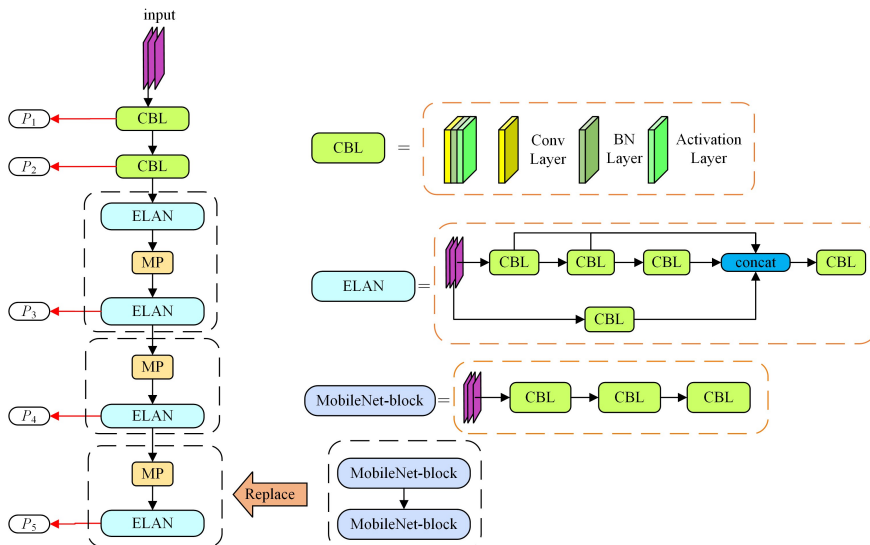


图 3 Mix-YOLO 骨干网络结构图

Fig. 3 Structure diagram of Mix-YOLO backbone network

## 2.2 ULSAM 注意力机制模块

SE 注意力模块在图像处理方面只是简单考虑了图像的通道上的权重分配方式,未考虑到图像空间上目标像素的权重注意力分配,导致了该注意力机制在不同类型的船舶检测识别上存在差异,该模块对于高层语义通道特征的提取能力不足。因此对于中高级特征通道筛选,还需要更高效的注意力模块对原有 MobileNetV3 的 SE 模块进行替换,以便对舰船种类特征信息进行更好的学习。

在网络训练中,不同尺度的特征有助于提升网络提取目标特征的能力。除了空间尺度能提升网络学习的效果之外,特征通道也能增强网络的目标分类能力。ULSAM(Ultra-Lightweight Subspace Attention Module)模块<sup>[20]</sup>是于 2020 年提出的一种轻量注意力机制模块。ULSAM 首先将输入的特征图划分为互斥的特征图,之后再对每个特征子空间进行不同的注意力映射,继而在通道上得到相应的多尺度特征表示。在 ULSAM 中,特征图  $F(H \times W \times M)$ ,  $H$  和  $W$  为图像大小,  $M$  为通道数。首先,将  $F$  分为  $n$  个互斥特征图组,其中  $[F_1, F_2, \dots, F_i, \dots, F_n]$  为一组中间特征图。

$$C_i = \text{softmax}(PW(\text{maxpool}(DW(F_i)))) \quad (4)$$

先对  $F_i$  进行深度卷积操作,之后进行核大小为  $3 \times 3$  的最大池化操作,再进行逐点卷积和 softmax 归一化后得到该组注意力图。利用式(6)得到的注意力特征图  $C_i$  与  $F_i$  进行元素相乘得到空间特征注意力权重分配情况,并与原  $F_i$  特征图元素相加得到优化后的特征图。最终将所有的  $F_i$  进行通道连接。

$$F_i = (C_i \otimes F_i) \oplus F_i \quad (5)$$

$$F = \text{concat}([F_1, F_2, \dots, F_i, \dots, F_n]) \quad (6)$$

在 ULSAM 的注意力特征整个获取过程中,与 SE 模块不同的是,该模块在通道压缩过程前采用了深度卷积操作,因此保留了图像像素通道间依赖信息。然后在激励阶段采用逐点卷积和 softmax 归一化操作替代了 SE 部件的全连接层+激活函数模块形式,从而可以利用不同特征子空间的线性关系对舰船特征的跨通道信息完成整合。因此该模块获取的特征能为高层特征即舰船种类识别阶段提供更具针对性的舰船目标种类信息,避免了在后续的特征融合增强阶段产生冗余信息的情况。引入 ULSAM 注意力模块的 MobileNet-block 示意图如图 4 所示。

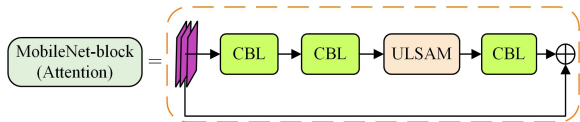


图 4 引入 ULSAM 的 MobileNet-block 模块

Fig. 4 Structure diagram of MobileNet-block attached with ULSAM

## 3 实验

### 3.1 舰船数据集及评价指标

由于舰船目标受到国防安全和技术手段等因素限制,现有针对舰船目标的公开光学遥感图像数据集较少,本文实验需要对不同类型舰船目标进行识别,以验证改进网络对不同类型舰船目标的识别效果,因此本文自建了港口舰船目标数

据集。该数据集图像分辨率在  $0.4 \text{ m} \sim 0.6 \text{ m}$  之间,图像大小为  $6144 \times 6144$  像素,原始图像数据来自于 Google Earth 和 Bing Aerial 开源瓦片图像源,图像覆盖了全球不同时相的港口基地。本文对不同气候条件(云团、浓雾)、不同光照条件以及多种背景干扰(建筑物、船坞、植被、岛屿、海滩)的港口场景进行了整理,按照人工视觉解译,对占比  $30 \times 30$  像素以上的舰船目标利用 LabelImg 等数据标注工具进行标定分类,粗略分为驱护巡(DFC)、两栖攻击舰(Amphibious Ship)、航空母舰(Aircraft Carrier)、民船(Civil Ship)、其他类(Other)。该数据集总共有 326 张  $6144 \times 6144$  大小的光学遥感图,其中训练集 252 张,测试集 74 张,将图像大小设置为  $1536 \times 1536$  大小送入网络进行训练。

本实验采用精度 P(Precision)、召回率 R(Recall)、平均精度 mAP(mean Average Precision)、平衡参数  $F_1$ 、网络参数量(Params, Parameters),以及反映模型复杂度的计算量(FLOPs)作为模型评价指标。mAP@0.5 和 mAP@0.5:0.95 为评估模型整体性能的指标。帧率(fps, Frame Per Second)为硬件平台 1s 内处理图像的帧数,图像处理速率为帧率的倒数,即硬件平台处理一张图像所需要的时间。

### 3.2 实验平台和网络参数设置

本实验网络训练平台为台式计算机,操作系统是 Linux-4.15.0,中央处理计算器为 i7-8700K CPU @ 3.70GHz,两张 GeForce GTX1080Ti 显卡,深度学习框架为 Pytorch-1.7.1。本文采用的嵌入式平台为寒武纪 MLU220-SOM 智能计算模组,该硬件平台具有时延低、功耗低、集成水平高、应用范围广等优点。中央处理器(CPU)为 ARM A55,支持处理图像最大尺寸为  $8192 \times 8192$  像素。嵌入式平台算力为 16TOPS (Tera Operations Per Second),功耗大小为 15W,芯片尺寸大小为  $60 \text{ mm} \times 55 \text{ mm}$ ,可以满足低功耗和实时检测处理的要求。网络初始学习率设置为 0.01,惯性动量为 0.937,权重衰减率为 0.0005,训练次数为 400,图批量大小为 8。

### 3.3 实验结果分析

#### 3.3.1 MobileNet-block 替换实验

为验证 MobileNet-block 模块对特征提取网络的影响。本文设计了如下实验。首先,以 MobileNet-block 大模块替换原 YOLOv7-tiny 网络所有提取特征的骨干网络,并将该网络命名为 YOLOv7-tiny-Mob 网络。实验结果如表 1 所列。

表 1 MobileNet-block 替换对比实验结果

Table 1 Comparison of capabilities of MobileNet-block replacement

Model	mAP@0.5	Params/MB	FLOPs/G
YOLOv7-tiny	0.612	5.939	75.2
YOLOv7-tiny-Mob	0.535	3.353	35.4

由表 1 可知,与 YOLOv7-tiny 网络相比,YOLOv7-tiny-Mob 网络的参数量和计算量下降了约 32.77% 和 47.3%,但 mAP@0.5 下降了 7.7%,说明 MobileNet-block 模块虽然能够大幅减少原网络计算量和参数量,但网络检测精度也大幅降低。因此不能对所有的特征提取层网络模块进行模块替换。

#### 3.3.2 不同特征提取层替换对比实验

为探究在 YOLOv7-tiny 骨干网络中不同提取特征层模块替换为 MobileNet-block 模块对网络检测识别能力的影响,本文在原骨干网络的不同位置进行 MobileNet-block 的对比

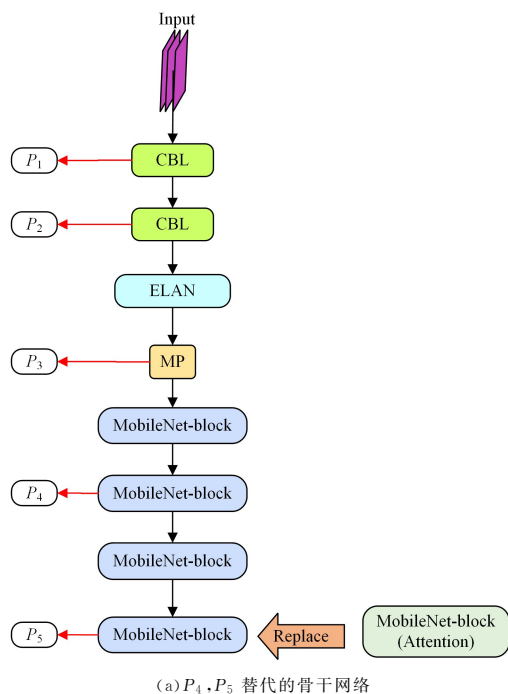
实验,实验结果如表 2 所列。

表 2 MobileNet-block 模块在不同位置的网络检测情况

Table 2 Comparison of detection capabilities of MobileNet-block applied at different positions in backbone network

Model	$mAP@0.5$	Params/MB	FLOPs/G
$P_2, P_3, P_4, P_5$	0.535	3.353	35.4
$P_3, P_4, P_5$	0.546	3.358	36.6
$P_4, P_5$	0.587	3.387	44.7
$P_5$	0.618	3.500	52.9
—	0.612	5.939	75.2

由表 2 得知,对产生低、中、高级特征( $P_2, P_3, P_4, P_5$ )的骨干网络替换后的网络,其参数量和计算量下降了 43.5% 和 53.5%,但  $mAP$  下降了 7.7%;对产生中、高级特征( $P_3, P_4, P_5$ )的骨干网络替换后的网络,参数量和计算量下降了 43.46% 和 51.91%,但  $mAP$  下降了 6.6%;对产生高级特征网络( $P_4, P_5$ )的骨干网络替换后的网络,其参数量和计算量下降了 43.0% 和 41.3%,但  $mAP$  下降了 2.5%;对只产生



高级语义特征  $P_5$  的骨干网络替换后的网络,其参数量和计算量下降了 41.1% 和 30.5%, $mAP$  上升了 0.6%。

对 YOLOv7-tiny 骨干网络替换过多的 MobileNet-block 模块( $P_2, P_3, P_4, P_5$ )后,由于网络参数量减少,因此骨干网络无法提取足够的特征,从而带来了网络检测精度的下降的问题。但同时可以看到,保留原提取低、中级特征的网络模块,将提取高级特征( $P_5$ )的网络模块替换成 MobileNet-block,能够保证原 YOLOv7-tiny 网络的检测精度。因此,Mix-YOLO 骨干网络应只对产生  $P_5$  的特征网络模块进行 MobileNet-block 的替换。

### 3.3.3 注意力机制对比实验

从 3.3.2 节中的实验可得到,产生高级特征( $P_5$ )的骨干网络模块在训练中并不会对网络检测精度能力产生显著影响。注意力添加示意图如图 5 所示。对该网络模块位置施加 ULSAM 注意力机制模块,从而对比出 ULSAM 模块对骨干特征提取网络的影响。

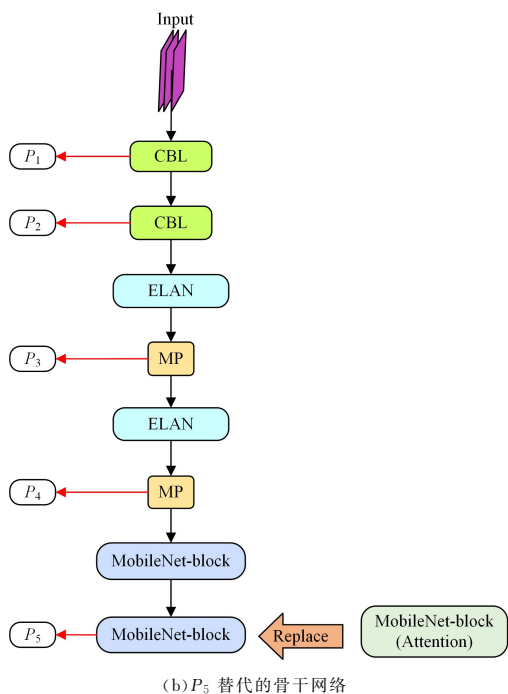


图 5 对不同骨干网络施加注意力机制示意图

Fig. 5 Diagram of ULSAM block applied in different backbone network

由表 3 可知,ULSAM 注意力模块能够大幅提高网络检测的召回率  $R$ ,同时对  $mAP$  指标有着小幅提升。在  $P_4$  和  $P_5$  网络中,加入 ULSAM 能使原网络召回率提升 8.2%, $mAP$  提升 1.6%;在  $P_5$  网络中,加入 ULSAM 能使原网络召回率提升 12.2%, $mAP$  提升 7.1%。由于网络已经在骨干网络中提取到较多特征,因此,ULSAM 模块能够对特征通道间注意力进行有效的筛选融合,网络提升效果更为明显。

表 3 施加注意力机制的网络检测对比

Table 3 Comparison of detection of networks attached with attention mechanism

Backbone	Attention	$R$	$mAP@0.5$	Params/MB	Flops/G
$P_4, P_5$	—	0.525	0.587	3.387	44.7
$P_4, P_5$	ULSAM	0.607	0.603	3.389	44.7
$P_5$	—	0.582	0.618	3.500	52.9
$P_5$	ULSAM	0.704	0.689	3.581	53.5

### 3.3.4 不同模型对比实验

为了验证改进 Mix-YOLO 网络在宽幅光学遥感图像的舰船检测识别情况,本文将当前主流的轻量化目标检测模型(YOLOv7-tiny 网络<sup>[10]</sup>、YOLOv5s 网络、YOLOv5n 网络)与本文方法进行了对比实验。由表 4 可得,YOLOv7-tiny 模型舰船检测精度最高, $F_1$  值在所有模型中较为平衡,目标检测精度和召回率比其他模型更具有竞争力,但模型参数量和计算量偏大,不适用于硬件部署和实时舰船目标检测应用,对实际场景的落地实现需要更多的说服力。YOLOv5n 的参数量和计算量最小,只有 YOLOv7-tiny 模型的 28.4% 和 32.1%,但损失了大量的目标准确率和召回率,不符合实际工程的指标要求。YOLOv5s 是 YOLOv5n 的扩张版本,其网络宽度提高了一半,参数量和计算量骤然上升,精度上涨幅度有限。

表 4 不同轻量级网络对比实验

Table 4 Comparison of capabilities of different lightweight networks

Model	P	R	mAP@0.5	mAP@0.5:0.95	F <sub>1</sub>	Params/MB	FLOPs/G
YOLOv7-tiny	0.682	0.616	0.612	0.458	0.662	5.939	75.2
YOLOv5s	0.710	0.494	0.522	0.349	0.583	3.384	49.3
YOLOv5n	0.536	0.418	0.355	0.290	0.470	1.689	24.4
Mix-YOLO	0.751	0.704	0.689	0.552	0.727	3.581	53.5

Mix-YOLO 网络的检测精度为 68.9%，相较于 YOLOv7-tiny 提升了 7.7%，相较于 YOLOv5s 提升了 16.7%，相较于 YOLOv5n 提升了 33.4%；Mix-YOLO 网络的参数量为 3.581MB，约为 YOLOv7-tiny 的 60.30%，是 YOLOv5s 的 1.05 倍，是 YOLOv5n 的 2.12 倍；Mix-YOLO 计算量比 YOLOv7-tiny 网络减少了 29.70%，比 YOLOv5s 和 YOLOv5n 分别增加了 8.52% 和 119.26%。

在回归和分类置信度设置相同的条件下，对上述训练完成的模型进行推理，图 6 选取了 3 个不同场景的舰船检测分类识别效果图。场景 1 (Scene 1) 是舰船维修停靠港口的情况，舰船分布较为密集，总计有 1 艘航空母舰和 8 艘其他类舰船。Mix-YOLO 可以将所有舰船完全检测识

别出来；YOLOv5n 检测识别出 3 艘其他类舰船；YOLOv5s 检测识别出 1 艘航空母舰，4 艘其他类舰船；YOLOv7-tiny 检测识别出 6 艘其他类舰船。场景 2 (Scene 2) 和场景 3 (Scene 3) 舰船分布情况较为稀疏，分别包含 2 艘航空母舰、3 艘其他类舰船和 1 艘航空母舰、2 艘驱护巡舰船、1 艘其他类舰船。在以上两个场景中，Mix-YOLO 网络均能有效地完成对所有舰船的检测和识别；YOLOv5n 检测识别出 1 艘航空母舰、3 艘其他类舰船和 1 艘航空母舰、2 艘驱护巡舰船；YOLOv5s 检测识别出 3 艘其他类舰船和 1 艘航空母舰、2 艘驱护巡舰船；YOLOv7-tiny 检测识别出 2 艘航空母舰、3 艘其他类舰船和 1 艘航空母舰、2 艘驱护巡舰船。

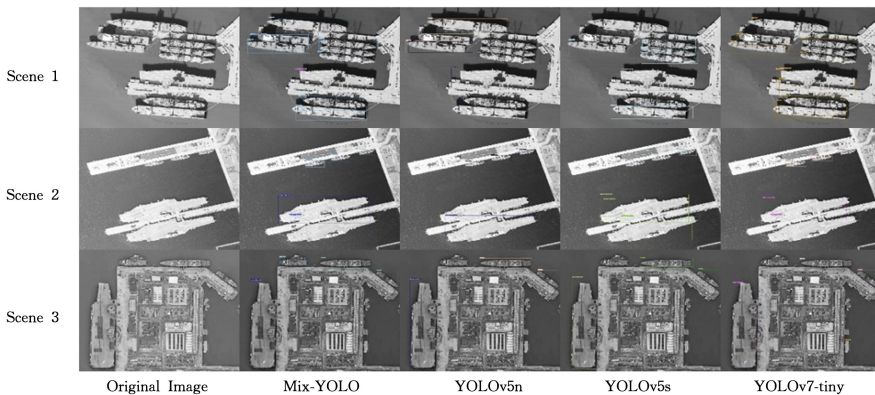


图 6 不同轻量级网络检测识别效果图

Fig. 6 Comparison of detection and recognition effects of different lightweight networks

表 5 为不同轻量级网络对不同类舰船的检测识别情况，评价指标为 mAP@0.5。由表 5 可知，Mix-YOLO 模型的舰船检测识别效果相比 YOLOv7-tiny, YOLOv5s, YOLOv5n 均有提升。Mix-YOLO 模型在驱护巡、两栖攻击舰、航空母舰的识别上达到了一定效果，但在民船和其他类方面不尽人意，除了网络层数少，骨干特征提取少等因素影响外，还与该数据集的设计有关。在构造数据集的时候，除了考虑船舶的用途分类外，还要考虑到舰船几何形状和目标大小。除此之外，在某些类别船只数量偏少的情况下需要对该类型船只进行适当的数据扩充。如何在某舰船类型数量较少和正负样本不均衡的情况下，提升舰船种类识别准确率，是后续网络改进的重点方向。

表 5 不同轻量级网络舰船检测识别情况对比

Table 5 Comparison of results of different lightweight network for ship detection and recognition

Class	YOLOv7-tiny	YOLOv5s	YOLOv5n	Mix-YOLO
DFC	0.642	0.610	0.300	0.726
Amphibious ship	0.807	0.619	0.502	0.852
Aircraft carrier	0.864	0.838	0.550	0.955
Civil ship	0.467	0.276	0.206	0.511
Other	0.278	0.267	0.217	0.401

### 3.3.5 嵌入式平台对比实验

在嵌入式平台中，若网络配置文件采用大模块构成，训练编译生成的离线模型在 MLU220 设备上运行速率会有一定程度的下降。通过实验分析，发现 MLU220 设备生成离线模型时，并没有把大模块自动拆分，而是把整个模块当作一个节点运行，这样在硬件上进行实际推理时速率会大幅降低，因此在网络训练时需要将大模块拆分成小模块（例如卷积、池化以及合并等基础模块），这样可以极大地发挥设备性能。按照上述方式对训练好的 Mix-YOLO 网络和 YOLOv7-tiny 网络分别进行嵌入式平台部署，Mix-YOLO 模型处理帧率为 120.88 fps，YOLOv7-tiny 模型处理帧率为 97.27 fps。

**结束语** 对于可部署的网络模型，需要同时兼顾模型检测性能和硬件平台支持友好度。从模型性能上来说，YOLOv7-tiny 的特征提取网络对于复杂场景的舰船目标检测具有优势，但该模型部署在小型嵌入式平台上存在计算量高、参数量大、帧率低的问题。以 MobileNet 系列网络为特征提取骨干网络的模型在计算量、参数以及帧率方面对硬件平台较为友好，但不符合实际目标检测需求。因此本文结合两者优势，设计出 Mix-YOLO 模型，对中低层特征提取的网络层采用 YOLOv7-tiny 网络的 ELAN 模型，对高层特征采用 MobileNet-block 作为提取网络，并加入 ULSAM 注意力模块，提

高了特征通道信息的利用效率。Mix-YOLO 模型精度上提高了 7.7%, 参数量和计算量比原 YOLOv7-Tiny 模型降低了 39.7% 和 29.7%, 帧率提升了 23.61, 达到了 120.88fps。基本上达到实际项目中所要求的模型计算量小, 处理速率快, 检测性能可靠的要求。下一步, 将在保证检测精度的情况下, 研究如何让网络实现小样本学习。

### 参 考 文 献

- [1] GIRSHICK R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile. IEEE, 2016:1440-1448.
- [2] JIANG H Z, LEARNED-MILLER E. Face Detection with the Faster R-CNN[C]//2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017). Washington, DC, USA. IEEE, 2017:650-657.
- [3] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy, IEEE, 2017:2980-2988.
- [4] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[J]. Computer Vision, 2016, 1(14):21-37.
- [5] WANG Y Y, WANG C, ZHANG H, et al. Automatic ship detection based on RetinaNet using multi-resolution Gaofen-3 imagery [J]. Remote Sensing, 2019, 11(5):531.
- [6] REDMON J, DIVVALA S, GIRSHICK R, et al. You Only Look Once: Unified, Real-Time Object Detection [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA, 2016:779-788.
- [7] REDMON J, FARHADI A. YOLO9000: Better, Faster, Stronger [C]//IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2017:6517-6525.
- [8] REDMON J, FARHADI A. Yolov3: An incremental improvement[J]. arXiv:1804.02767, 2018.
- [9] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection [J]. arXiv:2004.10934, 2020.
- [10] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//CVPR. 2023:7464-7475.
- [11] CUI J H, ZHANG Y Z, WANG Z, et al. Light-weight object detection networks for embedded platform[J]. Acta Optica Sinica, 2019, 39(4):0415006.
- [12] SU N, HUANG Z, YAN Y, et al. Detect Larger at Once: Large-

- Area Remote-Sensing Image Arbitrary-Oriented Ship Detection [J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19:1-5.
- [13] YU J, PENG X, LI S, et al. A Lightweight Ship Detection Method in Optical Remote Sensing Image under Cloud Interference [C]//2021 IEEE International Instrumentation and Measurement Technology Conference (I2MTC). IEEE, 2021.
- [14] NONG Y J, WANG J J. Remote sensing image caption method based on attention and reinforcement learning[J]. Acta Optica Sinica, 2021, 41(22):2228001.
- [15] LIU T, ZHOU B J, ZHAO Y S, et al. Ship Detection Algorithm based on Improved YOLO V5[C]//2021 6th International Conference on Automation, Control and Robotics Engineering (CACRE). Dalian, China, 2021:483-487.
- [16] WANG H Y, WANG C P, FU Q, et al. Lightweight Ship Detection Based on Optical Remote Sensing Images for Embedded Platform[J]. Acta Optica Sinica, 2023, 43(12):1212001.
- [17] LIU X W, PIAO Y J, ZHENG L L, et al. Ship detection for complex scene images of space optical remote sensing[J]. Optics and Precision Engineering, 2023, 31(6):892-904.
- [18] HOWARD A, SANDLER M, CHU G, et al. Searching for MobileNetV3 [C] // Proceeding of the IEEE/CVF International Conference on Computer Vision. 2019:1314-1324.
- [19] JIE H, LI S, GANG S. Squeeze-and-Excitation Networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018.
- [20] SAINI R, JHA N K, DAS B, et al. ULSAM: Ultra-Lightweight Subspace Attention Module for Compact Convolutional Neural Networks[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2020:1627-1636.



**HE Xinyu**, born in 1999, postgraduate. His main research interest is image processing applied to satellite.



**FENG Shuyi**, born in 1984, professor. His main research interests include image processing and artificial intelligence.