

DOI: 10.19650/j.cnki.cjsi.J2108000

# 面向无人艇环境感知的改进型 SSD 目标检测方法\*

孙 备, 左 震, 吴 鹏, 童小钟, 郭润泽  
(国防科技大学智能科学学院 长沙 410072)

**摘 要:** 为了提升无人艇对典型水面小目标感知能力, 本文提出了基于多尺度卷积融合结构和空间注意力加强的改进型 SSD 目标检测算法。首先, 对 SSD 浅层网络进行多尺度卷积融合, 提升浅层网络的语义信息; 其次, 设计空间注意力结构对卷积特征层逐个增强, 提升对弱纹理小目标特征保持性; 最后, 在 VOC 公开数据集和自构水面目标数据集上进行了测试, 并基于无人艇开展了真实海域目标检测识别验证。实验结果表明, 该算法在无人艇 Nvidia 平台的运行效率可达 15 fps, 能准确检测识别浮标、桥墩、渔船、快艇和货船等目标, 在典型海面场景虚警率为 5% 时的小目标检测率相对原生 SSD 算法提升近 20.2%, 平均有效检测率达到 79.3%。

**关键词:** 目标检测; 深度学习; 视觉注意力; 多尺度融合; 无人艇

中图分类号: TP391.4 TH701 文献标识码: A 国家标准学科分类代码: 510.4050

## Object detection for environment perception of unmanned surface vehicles based on the improved SSD

Sun Bei, Zuo Zhen, Wu Peng, Tong Xiaozhong, Guo Runze

(College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410072, China)

**Abstract:** To improve the perception ability of typical small water targets for unmanned surface vehicle (USV), this paper proposes an improved SSD object detection algorithm based on multi-scale convolution layer fusion and spatial attention enhancement architecture. Firstly, a multi-scale fusion method is utilized to improve the semantic representation of SSD shallow layer for small targets. Secondly, the spatial attention architecture is designed for each convolutional feature extraction layer to improve feature retention of small targets with weak texture. Finally, the proposed algorithm is evaluated on VOC and self-constructed surface target dataset. The real sea target detection and identification verification based on USV are carried out. Experimental results show that the proposed method can reach high operating efficiency with 15 fps on the USV Nvidia platform. The targets, such as buoys, bridge piers, fishing boats, speed boats and cargo ships, can be identified accurately. Compared with the original SSD algorithm, the proposed method could achieve a higher detection rate for small targets in the typical sea scene, which is increased by nearly 20.2% when the false alarm rate is 5%. The average effective detection rate can reach 79.3%.

**Keywords:** object detection; deep learning; visual attention; multi-scale fusion; unmanned surface vehicles

## 0 引 言

作为海洋智能机器人技术的发展产物,无人艇在水文采集、巡逻搜救、排雷反潜等领域有着巨大应用价值。随着海洋环境日益复杂、作业任务日渐多样、活动范围日趋扩大,水面无人艇对环境感知与自主决策的需求越来越迫切<sup>[1]</sup>。

越迫切<sup>[1]</sup>。

无人艇自主环境感知是实现其自主性、智能性的关键。典型方法大多通过携带可见光、激光雷达、航海雷达等传感器,并通过雷达和光电联动方式,获得目标的类别、位置和运动状态。由于激光雷达作用距离有限(通常在 100 m 以内),航海雷达回波只能获取目标的运动状态,因此可见光图像成为目标感知重要环节。然而,受复

收稿日期:2021-05-26 Received Date: 2021-05-26

\* 基金项目:湖南省自然科学基金(2020JJ5672)项目资助

杂水面浪涌、水雾以及阳光反射等因素影响,海面目标大多呈现距离远、尺度小、纹理弱等特点,而现有深度学习方法主要针对典型的平视、近距离拍摄场景,难以满足海面目标精细化检测识别需求<sup>[2]</sup>。

基于此,本文设计了一种多尺度卷积融合结构和空间注意力结构,通过多尺度融合结构提升浅层网络的语义表征能力,通过空间注意力模块提升各个卷积层对小目标的特征保持性。最后基于 SSD (single shot multibox detector) 原生网络进行改进和实验验证。此外,还构建了一个典型水面场景数据集,包含水鼓、浮标、桥墩、渔船、快艇和货船等典型水面目标共 5 279 个标注目标。最后,在 VOC (visual object classes challenge) 公开数据集和自构数据集上对本文算法进行测试。测试结果表明,该算法具有较高的检测精度,在典型水面场景下,在虚警率为 5% 条件下对小目标的有效检测率相比较原有 SSD 方法提升了 20.2%。

### 1 水面无人艇环境感知相关工作

近年来,世界各国相继开展了大量水面无人艇研制工作。如图 1 所示为国内外典型无人艇,其中,以美国和以色列一直处于领先地位。



图 1 国内外典型无人艇

Fig. 1 Typical unmanned vehicles at home and abroad

#### 1.1 无人艇平台方面

##### 1) 国外方面

美国于 2002 年首次成功研制半自主控制军用无人舰艇对船舶目标进行检测与跟踪;在 2011 年~2016 年期间,通过在无人艇上搭载先进光学探测系统、雷达和测距传感器,并结合先进人工智能技术,美军对护航、自主规划、持续跟踪等任务进行了演练,并相继研制了“Spartan Scout”、“Sea Hunter”等先进无人艇<sup>[4]</sup>。以色列立足于实现白天、夜晚以及各种不利天气条件下执行侦察、识别和

拦截敌舰、反恐、水雷战、电子战和精确打击等任务,已成功研制“海星”、“黄貂语”、“保护者”等多种型号无人艇。日本于 2014 年对外宣布研制成功“Aquarius”号无人艇,用于执行港口巡逻警戒、海洋探测、舰船数据采集等任务。

##### 2) 国内方面

尽管相较于国外有所落后,但国内大学和科研机构亦取得了扎实的进展。自 2008 年国产第一艘海上无人探测船“天象一号”问世,国内先后涌现出了以上海大学、哈尔滨工程大学、上海海事大学、云洲等一大批优秀的科研团队<sup>[4]</sup>。2008 年我国生产了第一艘海上无人探测船“天象一号”,主要用于气象探测领域。2014 年,上海海事大学成功研制了“海腾 01 号”水面艇,配备了 360°全方位摄像头、声呐、雷达和立体视觉等侦测设备,能够执行全天候水上甚至水下侦测任务。2014 年,上海大学在中国珠海航展参展的“精海”号无人艇配备北斗导航系统,具有自主定位和规避障碍能力,并在南海开展了测量和巡逻等简单任务。2016 年,哈尔滨工程大学与保利集团联合推出一款高速无人艇,最高航速 80 km/h,配备高清相机和数据链系统,用于执行水面监视和港口巡逻等任务。

#### 1.2 无人艇环境感知方面

海面目标精细化感知需要获取目标类别、位置和运动状态等信息。目前,国内外无人艇环境感知主要依靠激光雷达、航海雷达和可见光等传感器<sup>[5]</sup>,如图 2 所示。其中,激光雷达由于探测距离有限(通常在 100 m)主要用于紧急避障;航海雷达只能获取目标运动信息;而可见光方式可获得目标的类别、属性等信息。因此研究可见光环境感知方式非常关键。

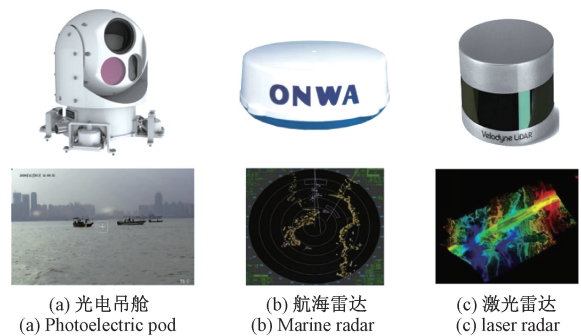


图 2 水面无人艇视觉传感器

Fig. 2 Visual sensors for surface unmanned vehicle

在可见光感知方面,现有方法主要侧重深度学习目标检测方法。例如基于区域推荐的 SPP-Net<sup>[6]</sup> (spatial pyramid pooling net)、Fast RCNN<sup>[7]</sup> (region-based convolutional network)、Faster RCNN<sup>[8]</sup>、Mask RCNN<sup>[9]</sup>等方法,通

过选择性搜索策略获取候选区域,而后基于训练器方式对候选区进行分类;基于回归的 SSD、YOLO-V3 (you only look once)、YOLO-V5 等<sup>[10]</sup>方法通过在图像像素上设置候选框,对目标类别和位置进行回归。

如图3所示为无人艇拍摄的某典型水面目标场景分析,其中目标A是渔船,目标B是水鼓,经统计渔船距离约为600 m。图中分别列举了不同分辨率下的目标成像,其中1 920×1 280是原始图像分辨率,300×300是SSD<sup>[3]</sup>(single shot multibox detector)深度学习目标检测算法的图像输入尺寸,75×75是SSD中最大特征图尺寸<sup>[2]</sup>。由图1可知,两类目标在75×75分辨率下的有用信息十分微弱,基本不可分。如何使深度学习检测算法对海面小目标有良好的特征保持性成为绝对目标感知精准性的重要因素。

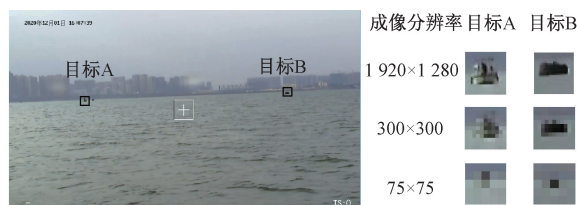


图3 某典型水面场景小目标分析

Fig. 3 Analysis of small targets in a typical water surface scene

通过对近几年深度学习目标检测算法的分析<sup>[11]</sup>,基于区域推荐的方法检测精度高,基于回归的方法检测效率高。在具体应用中,增加对两者的借鉴融合<sup>[12]</sup>,研究兼具检测精度和检测速度的目标检测算法是推向现实应用重要趋势。

### 1.3 深度学习小目标检测方面

由于深度学习检测器主要依靠浅层卷积提取小目标信息,而浅层卷积的语义特征往往不是很丰富,因此,小目标检测任务给现有方法带来了极大地挑战<sup>[13]</sup>。通过对文献的调研,我们将现有的小目标检测方法归纳为3类<sup>[14]</sup>。

1)通过提升小目标的样本比例,促进神经网络主动学习到小目标特征信息,可采取的方法通常包括人为构造小目标或修改匹配策略提升小目标匹配数量等,文献[12]采用过采样和小目标分割方法增加训练样本中的小目标比重,文献[15]通过设置小且密集的默认框提升小目标匹配的数量;还有类似的降低小目标匹配的IoU (intersection-over-union) 阈值,来控制不同尺度的目标匹配候选框数量。

2)通过高分辨特征图提升对小目标特征的保持性,以减少目标信息淹没在背景的概率。文献[16]通过训练多尺度检测器,对小尺寸特征图进行2倍插值再输入决策层;文献[17]则借鉴“卷积—反卷积”结构获得对候选框的超分辨特征表示。

3)通过改善特征提取方式提升对小目标的语义特征

提取能力。这类方法通常通过选择更深、更宽的骨干网络<sup>[18]</sup>,例如使用 ResNet (deep residual network) 及其各类变体作为骨干网络,使用特征融合方式筛选特征;此外,基于注意力机制的特征增强在自然语言处理、视频语义和图像细粒度分类等领域取得了显著效果。例如,文献[19]提出了一种挤压和激励网络 (squeeze-and-excitation network, SENet),通过压缩和扩展方法对每个通道全局信息进行建模,赋予重要通道更大的权重;CBAM<sup>[20]</sup> (convolutional block attention module) 从 Avg Pool 和 Max Pool 值中学习通道和空间注意力权重,并对两个注意模块串联执行;SCA CNN<sup>[21]</sup> (spatial and channel-wise attention) 结合多层特征图的通道注意和空间注意生成图像描述等。也可将其引入到检测网络中,使检测网络更加关注小目标信息。

通过对比上述3种方法:方法1主要是侧重对数据集的处理,没有从根本上解决问题;方法2带来了额外的计算和存储资源消耗,不具备实时性检测的优势;方法3通过改善网络对小目标的特征提取能力,是兼具检测精度和检测效率的有效途径。本文主要研究基于视觉注意力的深度学习目标检测方法,通过改进浅层特征目标保持能力,提升对小目标检测性能。

## 2 水面目标数据集

通过无人艇搭载4 K 高清相机拍摄典型海面场景的多类目标,然后经过人工去重、删除低质量图像、手工标注等步骤建立水面目标数据集共1 200张图像,5 279个标注目标。数据集场景包括晴天、雾天、雨天、逆光和大浪,数据集类别包括水鼓、浮标、桥墩、渔船、快艇和货船等,部分图像如图4所示。



图4 水面典型场景数据集

Fig. 4 Dataset of typical water surface scene

按照标注的目标像素尺寸将其分为极小目标、小目标、中目标、大目标、极大目标,并对自构数据集中各类目标数量和目标尺寸进行了统计。如图5所示为各类目标样本数量,如表1所示为目标尺寸分类标准。其中极小目

标对应 75×75 特征图 (SSD 的最大特征图) 的尺寸小于 3×3 pixel;小目标在 75×75 特征图的尺寸为 3×3~8×8 pixel。

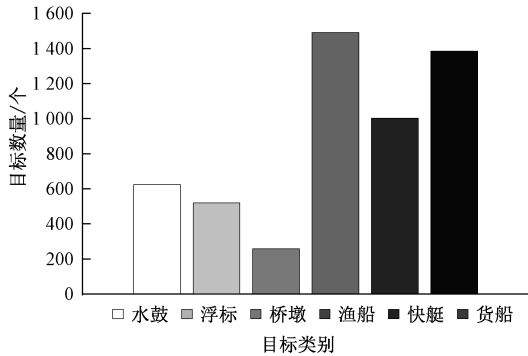


图 5 数据集各类目标数量统计

Fig. 5 Statistics of the number of various targets in the dataset

表 1 目标尺寸分类标准

Table 1 Classification criteria of target size

类别	尺寸范围 (相对 300×300)	最大特征图尺寸 (相对 75×75)
极小目标	≤10×10	≤3×3
小目标	10×10~32×32	3×3~8×8
中目标	32×32~50×50	8×8~12×12
大目标	≥ 50×50	≥ 12×12

### 3 无人艇环境感知平台搭建

如图 6 所示,为本文实验用的无人艇平台。无人艇在舱外搭载光电吊舱、航海雷达、差分 GPS、天线等模块,其中光电吊舱和航海雷达分别用于海面环境的可见光图像和雷达回波实时采集,差分 GPS 和惯导模块共同用于无人艇定位以及姿态参数获取,天线用于无人艇数据传输与信息交互。

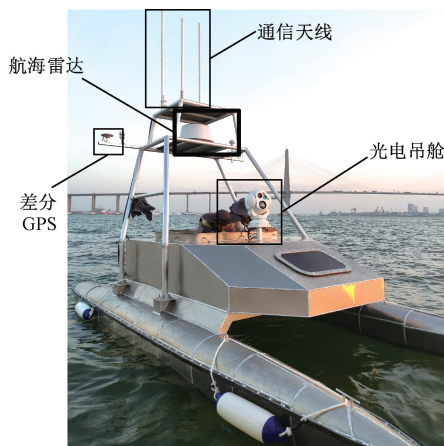


图 6 无人艇实验平台

Fig. 6 UAV experiment platform

本文对光电吊舱采集的图像进行目标检测识别,配合航海雷达回波进行目标检测与定位。所采取的雷达和光电联动方式如图 7 所示。具体包括:首先通过雷达回波发现疑似目标,然后引导光电转动到对应方位获取可见光图像,再通过图像检测算法进行目标类别的辨识,对水面目标的检测和定位。其中,航海雷达的数据处理主要采用图像平滑和闭合运算等预处理操作。光电图像的小目标检测是本文研究的重点。

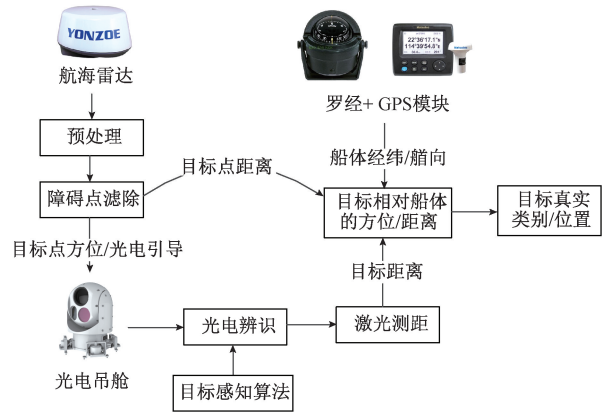


图 7 无人艇视觉感知系统基本组成

Fig. 7 The basic composition of the UAV visual perception system

图 8 所示为无人艇内部各个模块的连接实物图,分别包括 Nvidia 感知模块、决策控制模块、通信电台、电源等。其中,Nvidia Xavier 作为处理器对获得光电视频流进行实时的目标感知,Nvidia 通过串口和网络和光电吊舱连接,通过网口和航海雷达连接。Nvidia 平台通过网口获取航海雷达的 UDP 数据包,并通过解析包获得反射雷达波实现对目标的初探测;若发现目标则通过串口控制光电吊舱转动到指定指向角,然后读取实时视频流,通过光电感知算法对目标进行精确确认。因此光电感知是无人艇环境感知的关键。

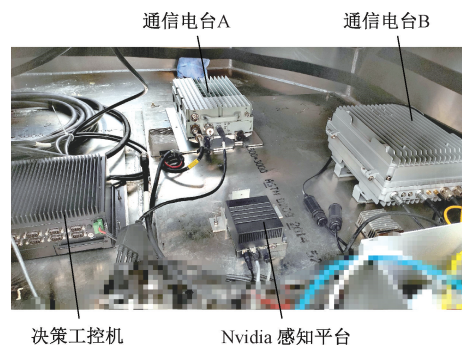


图 8 无人艇目标感知平台结构

Fig. 8 Target perception platform of UAV

## 4 基于空间注意力和级联融合的深度学习目标检测方法

### 4.1 空间注意力结构

为了提升小目标特征提取能力,避免叠加注意力掩码后的信息丢失。本文采用双分支结构提升注意力权值的有效性,其中,每个分支由连续的通道压缩卷积层、批量归一化层和非线性激活函数层组成,最终实现通道维数变换。

空间注意力结构通过残差网络学习掩膜,抑制原有特征图的背景信息。通常,注意力结构可由式(1)表示,其中 $f_1(x)$ 和 $f_2(x)$ 分别表示不同支路的注意力权值, $x$ 是输入特征图, $y$ 是注意力加强的特征图。

$$y = (f_1(x) \oplus f_2(x)) \otimes x \quad (1)$$

如图9所示是设计的空间注意力模块。两个注意力分支输出的卷积维数均为 $W \times H \times 1$ ,所示 $\oplus$ 运算是逐像素相加,即两个注意力分支计算的各自注意力图对应位置的权值相加;所示 $\otimes$ 运算是逐个像素相乘,即原特征图每个通道的值都乘以注意力图对应位置的权值系数。

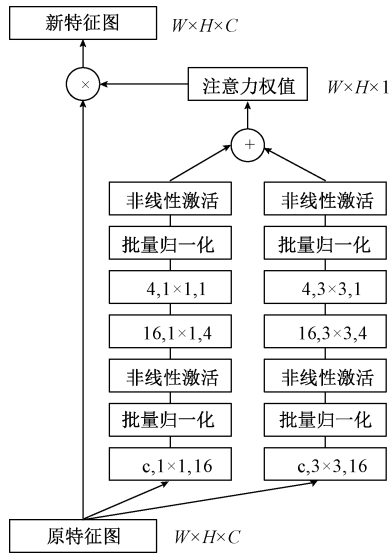


图9 双分支空间注意力结构

Fig. 9 Dual-branch spatial attention architecture

其中,原始特征图尺寸为 $W \times H \times C$ ( $C$ 为卷积通道),批量归一化层(batch normalization, BN)采用如式(2)所示的运算。其中, $\mu$ 为均值, $\sigma^2$ 为方差。

$$y_i = \gamma \frac{x_i - \mu}{\sqrt{\sigma^2 + \varepsilon}} + \beta \quad (2)$$

非线性激活层使用如式(3)所示的Leaky ReLU激活函数,并引入膨胀卷积(所述 $3 \times 3$ 结构)。其中 $a_i$ 为系数。

$$y_i = \begin{cases} x_i, & x_i \geq 0 \\ a_i x_i, & x_i < 0 \end{cases} \quad (3)$$

如图10所示是空间注意力增强后的特征图对比情况。由图10可明显看到,经过注意力增强的特征图对小目标的特征保持性更加显著。由于结构完全可微,参数可以在反向传播中得到有效更新,因此可学习到一张响应小目标区域的语义掩码,最终用于与原始特征逐通道逐元素相乘,以抑制背景响应、突出弱纹理目标区域,从而达到增强小目标区域响应的目的。

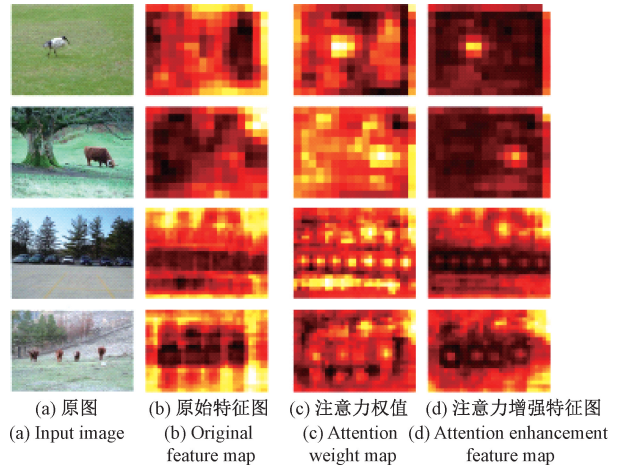


图10 空间注意力增强特征图

Fig. 10 Enhancement feature map based on spatial attention

### 4.2 级联融合结构

特征融合模块是通过融合不同层级的卷积层数据,缓解特征不变性与等变性矛盾,以提高浅层特征图的语义特征<sup>[16]</sup>。

如图11所示为所设计的特征融合模块。为了优化计算,特征融合模块首先利用两个 $1 \times 1$ 卷积层将输入特征图的通道数缩减为256,然后基于反卷积(转置卷积)将低分辨率特征图上采样到与高分辨率特征图一致的尺寸,其中卷积核大小分别设置为 $3 \times 3$ 或 $2 \times 2$ ,步幅为2。其中, $1 \times 1$ 卷积用于来调整通道数,每次卷积和反卷积后接ReLU激活函数层和批量归一化层(BN层),用于提升模型的非线性表达能力,并保证不同尺度的特征图具有相似分布和取值范围,以防止特征差异过大造成的特征发散。特征图融合的计算规则如式(4)所示。

$$y = F(f_a(x_1) + f_b(x_2)) \quad (4)$$

其中, $x_1$ 和 $x_2$ 分别对应特征图1和特征图2; $f_a(x_1)$ 和 $f_b(x_2)$ 分别表示融合前对特征图1和特征图2的处理; $f_a(*)$ 对应图11所述的 $\odot$ 操作,为两个矩阵对应位置元素相加。

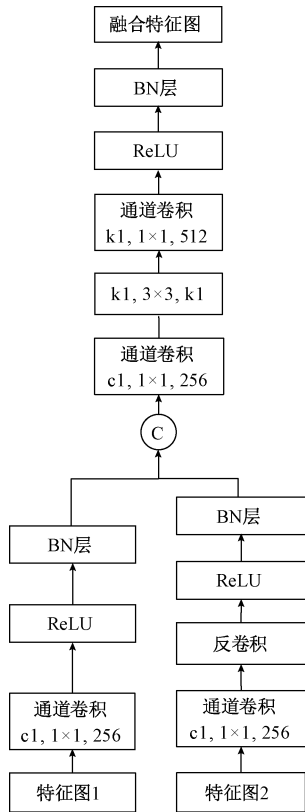


图 11 多尺度特征融合结构

Fig. 11 Multi-scale feature fusion architecture

### 4.3 网络结构

#### 1) 网络结构

如图 12 所示,为改进后的目标检测网络。在原生 SSD 网络的基础上引入上述空间注意力模块和特征融合模块。其中,级联融合模块和空间注意力模块主要作用在小目标信息集中的浅层网络,以提高网络对小目标的检测能力。本文共引入两个级联融合模块,分别对原生 SSD 网络的 Conv3\_3 层和 Conv4\_3 层,Conv4\_3 层和 Conv7 层进行融合。引入的注意力模块分别对级联融合输出的特征图、Conv8\_2、Conv9\_2 共 4 个特征图进行增强,最后将空间注意力增强的特征图输入决策层。

#### 2) Anchor 设置

Anchor 是一组预设边框,目标检测网络先根据获得大致可能的目标位置框,然后在这些预设框基础上进行调整,获得目标的准确位置。因此,设置合理的 Anchor 可有效提高目标位置的回归速度。为了提高对小目标的枚举效果,本文采取如式(5)所示计算 Anchor 尺寸。

$$s_k = s_{\min} + \frac{s_{\max} - s_{\min}}{m - 1} (k - 1) \quad (5)$$

其中,  $m$  为所特征图数量(本文中  $m$  取 6);  $k$  为特征图序号;  $s_{\min}$  和  $s_{\max}$  按照经验分别设置为 0.2 和 0.9。根据 Anchor 尺寸计算每个特征图上的预设框。

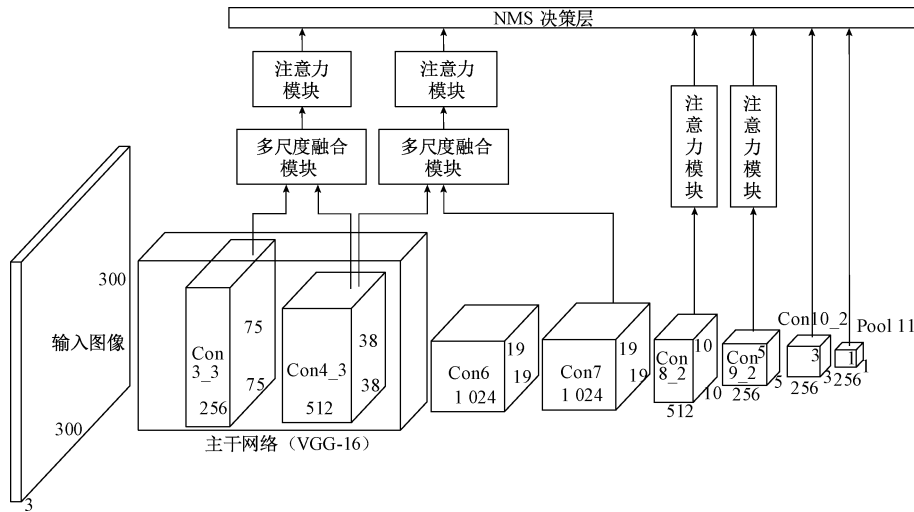


图 12 改进后的 SSD 检测网络结构

Fig. 12 The improved SSD detection network architecture

#### 3) 损失函数

本文采用和原生 SSD 一致的损失函数。如式(6)所示,损失函数由位置损失和类别损失组成。

$$L(x, c, l, g) = \frac{1}{N} (L_{conf}(x, c) + \alpha L_{loc}(x, l, g)) \quad (6)$$

其中,位置损失  $L_{loc}(x, l, g)$  采用 Smooth L1 损失函

数,如式(7)所示。

$$L_{loc}(x, l, g) = \sum_{i \in Pos} \sum_{m \in \{cx, cy, w, h\}} x_{ij}^k smooth_{L1}(l_i^m - \hat{g}_j^m) \quad (7)$$

其中,  $l^m$  表示预测框相对默认框的偏移量;  $\hat{g}^m$  表示真实框相对默认框的偏移量;  $cx, cy$  表示边框中心坐标;  $w, h$  表示边框宽高尺寸。

类别损失  $L_{conf}(x, c)$  采用 Softmax 损失函数, 如式(8)所示。

$$L_{conf}(x, c) = - \sum_{i \in Pos} x_{ij}^p \log(\hat{c}_i^p) - \sum_{i \in Neg} x_{ij}^p \log(\hat{c}_i^p) \quad (8)$$

$$\text{其中, } \hat{c}_i^p = \frac{\exp(c_i^p)}{\sum \exp(c_i^p)}.$$

#### 4) 正负样本选取

损失函数计算取决于正负样本选取。本文基于预测输出候选框与目标框真实位置 (ground truth, GT) 计算交并比 (IoU) 如式(9)所示, 并选取 IoU 最大的候选框作为正样本, IoU 小于阈值的候选框作为负样本。

$$IoU = \frac{A \cap B}{A \cup B} \quad (9)$$

为了控制正负样本集的比例, 按照类别置信度对候选框进行排序, 从低到高选取一定数量的负样本, 并控制正负样本比为 1:3。引入  $x_{ij}^p = \{1, 0\}$  指示第  $i$  候选框与第  $j$  个目标框关于类别是否匹配, 其中 1 表示匹配上, 0 表示未匹配上。

## 5 结果与分析

实验的硬件环境为 Intel I9-10900X @ 3.70 GHz, TITAN RTX GPU×2; 系统环境为 Ubuntu 16.04 LTS; 软件环境为 CUDA10.1 和 cuDNN7.6, 深度学习框架为 Pytorch 1.1.0。算法的衡量指标主要包括平均检测精度 (mean average precision, mAP) 和平均检测时间 (单张图片循环检测 100 次的平均时间), 其中, mAP 由 PASCAL VOC 提供的 MATLAB 工具包计算获得。

为了更加合理的评价算法性能, 本文定义了虚警率指标和检测率指标, 其中虚警率 (false alarm rate, FAR) 为预测错误边框数量占所有预测边框的比例, 检测率 (detection rate, DR) 为检测出的目标占实际目标数量的比例。是否预测正确的条件是按照预测框与真实框的 IoU 是否大于 0.5 评判。IoU 的计算方式如图 13 所示, 为真实区域和检测区域的重叠区域与交集区域之比。

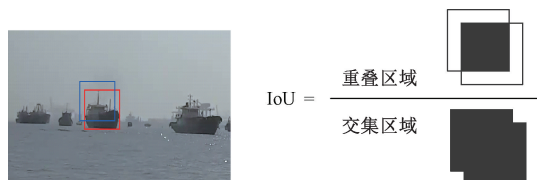


图 13 IoU 评价指标

Fig. 13 IoU evaluation index

### 5.1 VOC 数据集结果与分析

为了进行公正的比较, 首先在 VOC2007 公开数据集上对本文算法进行对比实验。训练集和测试集保留

VOC2007 原始设置, 并基于 mAP 和平均检测时间作为评价指标。

表 2 所示为引入注意力模块和特征融合模块的消融实验测试结果。由表 2 可知, 相比较原生 SSD, 引入多尺度融合结构, 模型 mAP 提升了 0.4%; 引入注意力结构, 模型 mAP 提升了 0.3%; 同时引入注意力结构和多尺度融合结构, 模型 mAP 提升了 0.5%。

表 2 基于 VOC 数据集的消融实验结果

Table 2 Results of ablation experiments based on VOC dataset

方法	注意力结构	特征融合结构	mAP/%	FPS / (帧/s)
SSD300	无	无	77.7	69.7
改进 1	无	有	78.2	54.2
改进 2	有	无	78.9	52.4
改进 3	有	有	79.3	45.3

为直观表示本文算法对小目标的检测效果, 图 14 所示为部分 VOC2007 数据集检测结果。其中第 1 行为原生 SSD 检测结果, 第 2 行为本文算法检测结果。在第 1 组中原生 SSD 算法漏掉了棕色小马; 第 2 组中原生 SSD 未检测出空中飞机; 第 3 组中原生 SSD 漏掉了被遮挡的人。

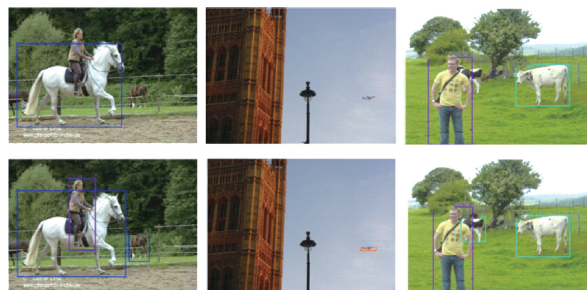


图 14 VOC 通用数据集检测结果

Fig. 14 Detection results based on VOC dataset

上述结果表明, 改进后算法相较于原有 SSD 目标检测算法, 对小尺寸目标的检测能力更加有效。这是由于原有 SSD 网络以 VGG-16 作为主干网络, 其浅层网络的语义信息较为缺乏。经过引入注意力增强和级联融合结构, 将深层网络和浅层网络进行级联特征增强, 提高了小目标显著性, 使其在网络逐层卷积过程中不被淹没在背景中。

### 5.2 典型水面目标数据集实验与分析

进一步, 为了验证真实海试场景下检测效果, 在自构的水面目标数据集上进行了对比测试。本文从图 4 所示的水面目标数据集中随机挑选了 3 000 张作为训练集,

1 000 张作为测试集,并控制随机挑选的不同类别和不同尺寸目标比例基本相同。为了保证训练过程的公平性,以原生 SSD 预训练模型作为初始化参数,Batch Size 设置为 32,采用默认的 SGD 优化器,初始学习率  $10^{-3}$ ,训练 1 000 步后调整为  $10^{-4}$ ,训练 6 000 步时调整为  $10^{-6}$  直至结束。最终选取迭代训练 7 000 步的模型用于测试并基于虚警率和检测率指标进行评价。

虚警率测试结果的性能曲线如图 15 所示。其中菱形标注曲线和圆圈标注曲线分别表示新增级联融合结构和注意力结构模块后的 SSD 算法结果。由图 15 可知,在相同的虚警率下(同一纵轴值),新增级联融合结构和注意力结构后的算法检测准确率要好于原生 SSD 算法(横轴值越大,检测率越好)。

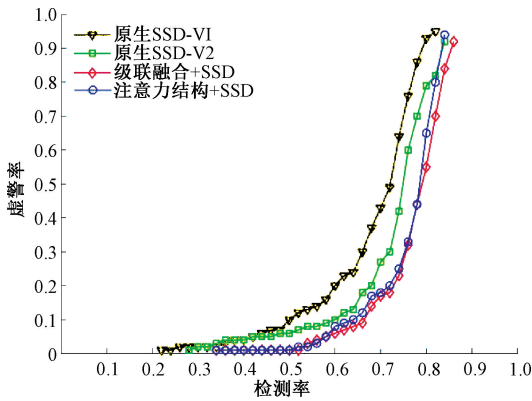


图 15 虚警率和检测率对比结果

Fig. 15 Comparison of false alarm rate and detection rate

表 3 按照不同目标尺寸,列举了虚警率分别为 0.05, 0.10, 0.20 时的检测率对比结果。其中表中所述的目标尺寸按照表 1 方式进行区分。

表 3 虚警率为 0.05, 0.10, 0.20 时检测率对比

Table 3 Comparison of detection rates with the false alarm rate 0.05, 0.10, and 0.20

虚警率	方法	目标尺寸		
		极小	小	中
0.05	SSD300	0.178	0.910	0.979
	本文方法	<b>0.380</b>	<b>0.978</b>	0.979
0.10	SSD300	0.289	0.955	0.979
	本文方法	<b>0.486</b>	<b>0.978</b>	0.979
0.20	SSD300	0.418	0.970	0.979
	本文方法	<b>0.592</b>	<b>0.978</b>	0.979

由表 3 可知,本文算法对小目标检测结果要明显优于 SSD 算法,对极小尺寸目标、小尺寸目标以及中等尺寸目标的有效检测率分别提升 20.2%, 19.7% 和 10.2%。

所述极小目标在  $300 \times 300$  图像上的像素区间小于  $10 \times 10$ , 对应在  $75 \times 75$  分辨率(SSD 最大特征图)不足  $3 \times 3$  像素,充分体现了本文算法对小目标的检测性能。

### 5.3 实时性结果与分析

表 4 所示为本算法与 SSD 原生算法综合性能对比情况。由表 4 可知,相较于原生 SSD 算法,本文算法在 mAP 指标上具有显著优势,至少提升了 0.5%。由于引入的新结构带来了计算力的增加,本文算法的检测效率有所下降,但经过测试,最终在 Titan X 平台上约为 20 fps,在 Nvidia Xavier 平台约为 15 fps,实时性基本满足应用要求。

表 4 基于 VOC 数据集的 SSD 改进算法性能对比

Table 4 Comparison of the performance of improved SSD algorithm based on VOC dataset

方法	GPU	mAP/%	FPS/(帧/s)
SSD300	TITAN RTX	77.7	69.7
CSSD300	Titan X	78.1	40.8
FSSD300	1080Ti	78.8	65.8
MDSSD300	1080Ti	78.6	38.5
RSSD300	Titan X	78.5	35.0
本文方法	TITAN RTX	79.3	45.3

### 5.4 真实海域目标检测应用

本算法最终搭载在无人艇实验平台,用于水面目标检测识别,算法在 Nvidia Xavier 平台的平均运行效率达到 15 fps,运行效率基本满足实时检测需求。部分实际检测结果如图 16 所示。从图 16 中可看出,本文算法对海面小目标检测结果较好,能准确检测并识别海面出现的水鼓、浮标、桥墩、渔船、快艇和货船等目标,基本满足无人艇环境感知需求。



图 16 某海域实艇目标检测识别结果

Fig. 16 Real object detection and recognition results in a certain sea area

如图 17 所示,本文研制的算法最终支撑了无人艇集群的自主环境感知与控制试验。试验过程中,将某艘己方无人艇指定为护航目标,护航目标通过艇间 UDP 协议



向其他无人艇发送自身位置;将某类渔船指定为威胁目标,基于本文研制的无人艇水面目标检测识别算法对海面环境进行感知,发现指定渔船则进行威胁目标预警,派出距离最近的无人艇遂行前出侦察任务。



(a) 俯视视角下无人艇集群任务

(a) USV cluster mission under vertical view



(b) 无人艇视角下目标由远至近运动

(b) The target moving from far to near under USV view

图 17 支持典型多无人艇集群任务场景

Fig. 17 Supporting typical multi-UAV cluster mission scenario

## 6 结 论

本文提出了基于空间注意力结构和级联融合结构改进的 SSD 水面小目标检测算法,空间注意力能够提升网络对弱纹理目标的特征利用,而级联融合结构可有效利用网络深层和浅层的语义信息,两者共同作用可有效提升网络对小目标的检测精度,并构建了基于无人艇的典型水面目标数据集。本文在 VOC 数据集和自构水面目标数据集上进行了训练和测试,并和 SSD 算法进行了对比。在 VOC 数据集上, mAP 提升了 1.6%;在自构数据集上,在虚警率为 5% 条件下,本文算法相比 SSD 对小目标有效检测率提升了 20.2%。

无人艇作为水面无人系统的代表性装备,其目标感知能力是无人艇自主航行和行为决策的前提。光电目标检测可和雷达图像有效联合,共同支撑无人艇水面目标感知。未来的研究应当考虑如何在精简网络的基础上提升检测效率;在样本不平衡条件下实现少样本目标稳定感知,并提升模型网络模型的泛化能力等。

### 参考文献

[1] 张卫东,刘笑成,韩鹏. 水上无人系统研究进展及其面临的挑战[J]. 自动化学报, 2020, 46(5):847-857.  
ZHANG W D, LIU X CH, HAN P. Progress and challenges of overwater unmanned systems [J]. ACTA Automatica Sinica, 2020, 46(5):847-857.

[2] 郑浦,白宏阳,李政茂,等. 抖动干扰下运动目标精准检测与跟踪算法设计[J]. 仪器仪表学报, 2019, 40(11):90-98.  
ZHENG P, BAI H Y, LI ZH M, et al. Design of accurate detection and tracking algorithm for moving target under jitter interference [J]. Chinese Journal of Scientific Instrument, 2019, 40(11):90-98.

[3] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[J]. IEEE European Conference on Computer Vision, 2016, 10:21-37.

[4] 王雅琳,刘都群,杨依然. 2019 年水下无人系统发展综述[J]. 无人系统技术, 2020, 3(1):55-59.  
WANG Y J, LIU D Q, YANG Y R. Summary of the development of unmanned undersea systems in 2019 [J]. Unmanned Systems Technology, 2020, 3(1):55-59.

[5] 朱健楠,虞梦苓,杨益新. 无人水面艇感知技术发展综述[J]. 哈尔滨工程大学学报, 2020, 41(10):1486-1492.  
ZHU J N, YU M L, YANG Y X. Overview of development of unmanned-surface-vehicle sensing technology [J]. Journal of Harbin Engineering University, 2020, 41(10):1486-1492.

[6] 徐晓光,李海. 多尺度特征在 YOLO 算法中的应用研究[J]. 电子测量与仪器学报, 2021, 35(6):96-101.  
XU X G, LI H. Application research of multi-scale features in YOLO algorithm [J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(6):96-101.

[7] GIRSHICK R. Fast R-CNN [J]. IEEE 2015 Computer Vision and Pattern Recognition, 2015, 12(13):1440-1448.

[8] 张晓丽,董昱. 面向铁路货车车号定位的 Faster R-CNN 卷积神经网络[J]. 电子测量与仪器学报, 2020, 34(10):65-73.  
ZHANG X L, DONG Y. Faster R-CNN convolutional neural network for the location of freight train number [J]. Journal of Electronic Measurement and Instrumentation, 2020, 34(10):65-73.

[9] 蔡彪,沈宽,付金磊,等. 基于 Mask R-CNN 的铸件 X 射线 DR 图像缺陷检测研究[J]. 仪器仪表学报, 2020, 41(3):61-69.  
CAI B, SHEN K, FU J L, et al. Research on defect detection of X-ray DR images of casting based on Mask R-CNN [J]. Chinese Journal of Scientific Instrument, 2020, 41(3):61-69.

[10] 李晖晖,周康鹏,韩太初. 基于 CReLU 和 FPN 改进的 SSD 舰船目标检测[J]. 仪器仪表学报, 2020, 41(4):183-190.

- LI H H, ZHOU K P, HAN T CH. Ship object detection based on SSD improved with CReLU and FPN [J]. Chinese Journal of Scientific Instrument, 2020, 41(4): 183-190.
- [11] 李成美, 白宏阳, 郭宏伟, 等. 一种改进光流法的运动目标检测及跟踪算法[J]. 仪器仪表学报, 2018, 39(5): 249-256.
- LI CH M, BAI H Y, GUO H W, et al. Moving object detection and tracking based on improved optical flow method [J]. Chinese Journal of Scientific Instrument, 2018, 39(5): 249-256.
- [12] 张业宝, 徐晓龙. 基于改进 SSD 的安全帽佩戴检测方法[J]. 电子测量技术, 2020, 43(19): 80-84.
- ZHANG Y B, XU X L. Safety helmet wearing detection method based on improved SSD [J]. Electronic Measurement Technology, 2020, 43(19): 80-84.
- [13] 张培培, 王昭, 王菲. 基于深度学习的图像目标检测算法研究[J]. 国外电子测量技术, 2020, 39(8): 34-39.
- ZHANG P P, WANG ZH, WANG F. Review on image target detection algorithm based on depth learning [J]. Foreign Electronic Measurement Technology, 2020, 39(8): 34-39.
- [14] KRISHNA H, JAWAHAR C V. Improving small object detection [J]. The 4th IAPR Asian Conference on Pattern Recognition, 2017, 11: 340-345.
- [15] CUI L SH, MA R, LYU P, et al. MDSSD: Multi-scale deconvolutional single shot detector for small objects [J]. IEEE Computer Vision and Pattern Recognition, 2020, 2(25): 98-100.
- [16] JIANG D, SUN B, SU SH J, et al. FASSD: A feature fusion and spatial attention-based single shot detector for small object detection [J]. Electronics, 2020, 9(9): 1536.
- [17] ZHAO Q J, SHENG T, WANG Y T, et al. CFENET: An accurate and efficient single-shot object detector for autonomous driving [J]. IEEE Computer Vision and Pattern Recognition, 2018, 10(10): 1-6.
- [18] 刘寒迪, 赵德群, 陈星辉, 等. 基于改进 SSD 的航拍施工车辆检测识别系统设计[J]. 国外电子测量技术, 2020, 39(7): 127-132.
- LIU H D, ZHAO D Q, CHEN X H, et al. Design of detection and recognition system of aerial photography construction vehicle based on improved SSD [J]. Foreign Electronic Measurement Technology, 2020, 39(7): 127-132.
- [19] JIE H, LI SH, GANG S, et al. Squeeze-and-excitation networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 9(42): 2011-2023.
- [20] 王生霄, 侯兴松, 黑夏萌. 嵌入 CBAM 结构的改进 YOLOV3 超宽带雷达生命信号检测算法 [J]. 国外电子测量技术, 2020, 39(3): 1-6.
- WANG SH X, HOU X S, HEI X M. Improved YOLOV3 UWB radar life signal detection algorithm embedded in CBAM structure [J]. Foreign Electronic Measurement Technology, 2020, 39(3): 1-6.
- [21] CHEN L, ZHANG H W, XIAO J, et al. SCA-CNN: Spatial and channel-wise attention in convolutional networks for image captioning [J]. IEEE Computer Vision and Pattern Recognition, 2017, 7(6): 5659-5667.

### 作者简介



孙备, 分别于 2012 年、2014 年和 2018 年于国防科技大学获得学士、硕士和博士学位, 现为国防科技大学讲师, 主要研究方向为智能侦测和目标感知。

E-mail: sunbei08@nudt.edu.cn

**Sun Bei** received his B. Sc., M. Sc. and Ph. D. degrees all from National University of Defense Technology in 2012, 2014 and 2018. He is currently a lecturer at National University of Defense Technology. His main research interest includes intelligent detection and target perception.



左震(通信作者), 分别于 2002 年、2004 年和 2012 年于国防科技大学获得学士、硕士和博士学位, 现为国防科技大学副研究员, 主要研究方向为智能侦测和目标感知。

E-mail: z.zuo@nudt.edu.cn

**Zuo Zhen** (Corresponding author) received his B. Sc., M. Sc. and Ph. D. degrees all from National University of Defense Technology in 2002, 2004 and 2012, respectively. He is currently an associate professor at National University of Defense Technology. His main research interest includes intelligent detection and target perception.