基于深度学习的高精度船舶检测方法

作者一[1, 2] 作者二[2] 作者三[1]

摘要当前主流的目标检测器在公有数据集上都有优异的性能表现，但是在特定数据集上的性能还有待提升。我们设计了一种基于深度学习的高精度船舶检测方法。为了提升方法准确率，我们以 Scaled-YOLOv4 为基础，引入迁移学习，并在设计、训练和测试过程中分析目标域数据的特点，对模型做针对性调整，同时使用马赛克数据增强、CIOU Loss 损失计算、测试时增强等技术进一步提升我们的模型准确率，最终我们设计的方法在海洋目标智能感知国际挑战赛中获得了第二名。

关键词 目标检测，迁移学习，船舶检测

中图分类号：TPxxx. x

Deep Learning Based High Accuracy Ship Detection Method

FIRST Author-Aa[1,2] SECOND Author-Bb[2]
THIRD Author-Cc[1]

Abstract The current mainstream object detectors have excellent performance on public datasets, but the performance on the special dataset needs to be improved. We propose a deep learning based high accuracy ship detection method. To improve accuracy, we choose Scaled-YOLOv4 as our foundation. We use transfer learning, and analyze the characteristics of the target domain data to adjust our model. During designing, training and testing, we use lots of tricks like Mosaic Data Enhancement, CIOU Loss, Test Time Augmentation, etc, to improve our model's performance. By the method we designed, we won the second prize of the Ocean Target Intelligent Perception Competition.

Key words Object Detection, Transfer Learning, Ship Detection

1 引言

智能感知现已被应用于各行业的智能管理中，水上交通也不例外。智能感知技术在航道安全监控、构建智能船舶等方面都发挥着至关重要的作用，其中基于视觉的船舶检测技术更是重中之重，河岸、港口装载摄像头对目标进行检测，可以对航行中的船舶进行精准定位，为航速监控、船舶偏航预警、吃水预警等提供决策；船舶装载摄像头检测目标，可以为船舶航行提供附近船只、障碍物、河岸距离等关键信息，保证船舶的安全行驶。[1]

目标检测是人工智能领域里一个经典但永不过时的课题，近年来，基于深度学习的目标检测算

法不断发展进步，在检测准确率和效率方面都有着本质的提升，各行业也开始将基于深度学习的目标检测算法投入到实际的应用场景中，如自动驾驶、人脸识别、字符识别等等。也陆续有研究人员将深度学习的方法应用到船舶检测中去，但是形如船舶检测这种特殊应用场景或特殊行业在使用基于深度学习的方法时，总会面临如下几个问题：

（1）数据量少；
（2）数据采集成本高；
（3）数据标注成本高。

作为数据驱动的深度学习方法，拥有的数据的数量和质量直接决定了我们设计的方法的质量，为在船舶检测中有效解决上述问题，我们以通用目标检测领域的SOTA方法为基础，以通用目标检测为源域，船舶检测为目标域进行迁移学习，设计并训练了一种高精度船舶检测方法，并在海洋目标智能感知国际挑战赛中获得了第二名。

2 相关工作

2.1 基于深度学习的目标检测方法

目前，基于深度学习的目标检测方法按照监督方法可分为监督学习目标检测和弱监督学习目标检测两大类；其中监督学习目标检测可分为基于锚框的目标检测方法和无锚框检测方法，基于锚框的目标检测方法又可分为一阶段检测方法和二阶段检测方法，无锚框检测方法又可分为基于关键点检测的方法和基于目标中心检测的方法。

2.1.1 基于锚框的目标检测方法

二阶段检测方法：Faster R-CNN[2]、Mask R-CNN[3]及其衍生方法是二阶段目标检测方法的代表，Faster R-CNN 检测目标时先由 RPN（Region Proposal Network）对整张图片进行粗检测，提取可能存在物体的区域，再由 R-CNN 对提取的区域进行精检测得到最总检测框和类别信息；Mask R-CNN 在 Faster

R-CNN 基础上改进了 RPN 与 R-CNN 的连接方式，并在 R-CNN 最终的输出增加了 Mask 分支，使整个网络的性能获得了进一步提升。

一阶段检测方法：以 Mask R-CNN 为首的二阶段检测方法曾经在很长的一段时间内都是各项检测任务的 SOTA（State of the Art）方法，但是二阶段方法的检测速度非常慢，计算资源消耗非常大，为提升检测效率，减少资源消耗，部分研究工作者开始把注意力放在一阶段方法的设计上。SSD（Single Shot MultiBox Detector）[1]的提出成为了一阶段检测方法蓬勃发展的开端，在神经网络提取特征

后设计了多个基于锚框的输出层对不同尺度的目标进行检测。

随后陆续有研究工作者对其二者进行了不同方向的改进，其中 YOLO 系列[4][5][6][7]的更新发展让我们看到了一阶段检测方法不仅在检测速度更快，而且检测准确率也在不断提升，Scaled YOLOv4[8]更是一度成为 COCO Detection 数据集的 SOTA 方法。

2.1.2 无锚框目标检测方法

基于锚框的检测方法锚框的宽高、数量等设定较为复杂，网络结构的复杂程度也受锚框设定的影响，而无锚框目标检测的方法拜托了锚框设定的限制，使网络结构变得更加简洁，提升了模型训练和使用时的效率，Focal Loss[11]的发明更是进一步提升了无锚框检测方法的准确率。

基于关键点的检测方法通过学习能够检测到目标一个或多个关键点，在根据检测到的关键点生成预测的边界框。CornerNet[9]首先检测图像中所有目标的左上角和右下角关键点，再对每个关键点进行配对组成同一个目标的一组关键点，再围绕这一组关键点生成边界框；CenterNet[10]只检测目标中心点，并围绕中心点预测目标边界框宽高。

基于目标中心点或中心部分的检测方法把目标中心点或中心部分视作图像前景并定义为正样本，然后预测正样本点到目标边界框四条边的距离，以此确定最终的预测框。FCOS[12]将目标标签框内的所有像素点全部视为正样本，然后计算所有点到边界框四条边的距离和所有点靠近中心的程度。随后许多研究人员从各方面尝试改进提升 FCOS 的性能，其中，ATSS[13]提出了一种全新的自适应正样本选择方法，改进了训练方法，将 FCOS 的性能进一步提升；Xiang Li 等设计了 Generalized Focal Loss[14]，改进损失计算，使模型检测性能在 ATSS 的基础上进一步提升。

2.2 迁移学习

随着人工智能的蓬勃发展，越来越多的行业都开始尝试用深度学习的方法在解决问题，其中表现较好的监督学习需要对大量的数据进行标注，数据标注是一项极其枯燥无味且耗费大量人力物力财力的工作，所以迁移学习受到越来越多的关注。

常用的迁移学习方法是，把我们需要解决的问题视为目标域，选取一个数据量大、特征相比目标域更加泛化的数据集作为源域进行预训练，然后将预训练的模型作为目标域训练的起点再进行训练。

2.3 模型缩放

人工智能蓬勃发展已有数年，对于人工智能的各个方向都有优秀的研究成果，现阶段人工智能的主要任务已经由方法研究转移到的工业应用部署，而在实际的部署使用过程中我们真正需要的是一套完整的解决方案，即为了适应实际应用中不同的硬件条件，我们需要一套深度学习的模型而不是单单一个。

近两年 YOLOv4[7]、YOLOv5、EfficientDet[15]等在设计网络模型时都对模型准确率、推理速度、参数量等进行综合考量，进行模型缩放（Model Scaling）设计了一整套适用于不同硬件条件的网络模型。而 Scaled YOLOv4[8]更是通过理论论证和实验证明对模型缩放做出了更加详细的定义，根据新的定义所设计的一整套模型在与同等级的模型中检测准确率和效率都有更优秀的表现。

2.4 基于深度学习船舶检测方法

目前，多数基于深度学习的船舶检测方法都是在通用目标检测方法的基础上进行定制化设计，但是其中应用的网络模型多为多年前的方法，使用技术不够新。
为此我们基于通用目标检测领域最新的研究成果设计了一种高精度船舶检测方法，并在海洋目标智能感知国际挑战赛中获得了第二名。

3 模型和方法描述

我们选择以目前 COCO Detection 数据集目前的 SOTA 模型 Scaled YOLOv4 为基础，并以通用目标检测为源域，船舶检测为目标域进行迁移学习。

3.1 模型和方法原理

3.1.1 模型结构

Scaled YOLOv4 的核心在于严格的模型缩放（Model Scaling），原文中作者通过对特征图宽高、通道数、网络参数量等对计算精度和速度的影响进行严格的理论和实验论证，设计出了一整套适用于不同硬件条件的网络模型。

Scaled YOLOv4 P6 网络结构如图 1，形如上一代 YOLOv4 和 YOLOv5，总体上采用 FPN-PAN[16][21]结构，FPN-PAN 网络结构的优点在于在特征金字塔的基础上更加细致的进行多尺度检测，并在前者的基础上加深了网络结构，将输出分为了四个层次，对不同尺度的目标进行更加细致的检测。

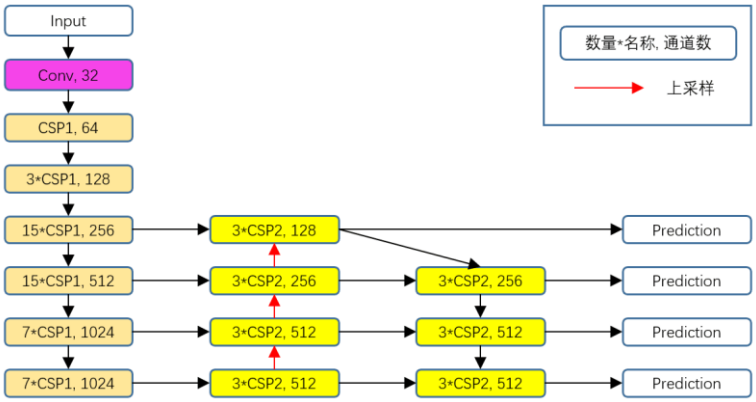

图 1 Scaled YOLOv4 P6 模型结构
Fig.1　Model structure of Scaled YOLOv4 P6

3.1.2 基础模块

CBM（Convolution+Batchnorm+Mish）：卷积+批归一化+Mish 激活函数[17]，结构图如图 2，其中 Mish 激活函数为：

$$f(x) = xtanh(ln(1+e^x)) \quad (1)$$

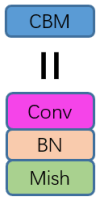

图 2 CBM 模块结构
Fig.2　Module structure of CBM

其函数图像如图 3。Mish 是神经网络结构中一种非单调的自动正则化的激活函数，原作用实验证明 Mish 的输出要比 ReLU 和 Swish[18]更加平滑，而输出平滑的激活函数会提高深层神经网络特征图的质量，使整个模型准确性更高，泛化能力更强。

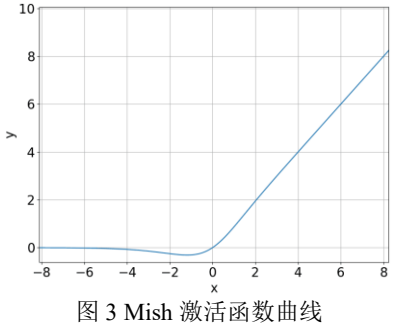

图 3 Mish 激活函数曲线
Fig.3　Curve of Mish activation function

残差结构：

$$y = x + f(x) \quad (2)$$

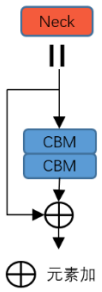

图 4 残差结构

Fig.4 Structure of residual block

其中 *f()*为两个 CBM 模块堆叠，结构图如图 4。残差结构首次出现是在 ResNet 中，为了解决当时深度神经网络训练难度大的问题，现已成为多数神经网络在设计 Backbone 时的常用结构。

CSP1 模块：CSP1 是 Scaled YOLOv4 的 backbone 中的主要组成结构，如图 5 左侧。参考 CSPNet（Cross Stage Partial Network）[19]设计时重点从网络结构搭建的角度出发来减少模型的计算量这一点，将基础层的特征映射划分为两部分，然后通过类残差结构进行跨阶段合并，减少了网络优化时重复的梯度信息，在削减计算量的同时保证了准确率。

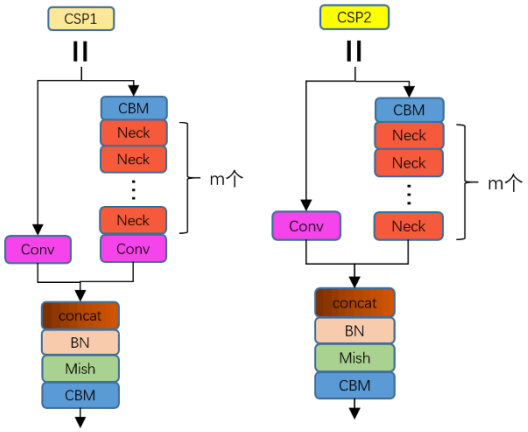

图 5 CSP1 和 CSP2 模块结构

Fig.5 Module structure of CSP1 and CSP2

CSP2 模块：CSP2 模块是 Scaled YOLOv4 中 FPN-PAN 结构的主要组成部分，如图 5 右侧，与 CSP1 模块的主要区别在于残差块堆叠的分支最后没有再加卷积层（Convolution）。

3.1.3 损失计算

Scaled YOLOv4 的损失分为三部分：CIOU Loss、Object Loss、Class Loss，其中 CIOU Loss 是 DIOU Loss[20]的改进版本。DIOU Loss 如式（3），能够在优化过程中最小化预测框与标签框之间的归一化距离，如图 6（a）和（b），当标签框包裹预测框的时候，直接考察两个框的距离，相对于先前的 GIOU Loss 收敛的更快。

$$L_{DIOU} = 1 - (IOU - \frac{D_2}{D_C}) \quad (3)$$

其中 IOU 为预测框、标签框交集面积和预测框、标签框并集面积的比值，D_2 为标签框中心点与预测框中心点之间的欧氏距离，D_C 为标签框与预测框最小外接矩形对角线距离。

但是 DIOU Loss 并不是完美的，如图 6(c)、(d)两图中预测框中心点位置相同，且面积相同，但是两个框存在着肉眼可见的差别，这时候 DIOU Loss 并不能区分二者差别，CIOU Loss 应运而生。

CIOU Loss 在 DIOU Loss 的基础上加入了对预测框宽高和预测框宽高的考量，如式（4）：

$$L_{CIOU} = 1 - \left(IOU - \frac{D_2}{D_C}\right) - \frac{v^2}{(1-IOU)+v} \quad (4)$$

其中，v 是衡量标签框与预测框宽高比一致性的参数，如式（5），w^{gt} 和 h^{gt} 表示标签框的宽高，w^p 和 h^p 表示预测框宽高：

$$v = \frac{4}{\pi^2}(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w^p}{h^p})^2 \quad (5)$$

而 Object Loss、Class Loss 都是标准的交叉熵损失。

并且在测试时，Scaled YOLOv4 采用 DIOU 非极大抑制，即在对目标进行边界框回归时考虑边界框中心点位置，这种方法尤其在多目标重叠时的表现要优于普通非极大抑制。

3.2 改进策略

3.2.1 迁移学习

海洋杯大赛船舶检测训练集共有 9800 张图片，为了方便在训练过程中进行调优，我们随机抽取了 980 张图片作为验证集使用，这样用于训练的数据就减少了，为应对数据

减少的问题同时增强模型的泛化能力，我们

图 6 DIOU Loss 在目标检测中计算示例。(a)和(b)是 DIOU Loss 参数示例，(c)和(d)是 DIOU Loss 缺陷示例
Fig.6　Examples of DIOU Loss computation in the object detection. (a) and (b) show examples of DIOU Loss parameters, (c) and (d) show examples of DIOU Loss defects.

选择 COCO Detection 通用目标检测数据集作为源域来进行迁移学习，Scaled YOLOv4 的开源代码中已经附带了 COCO 数据集上训练调优后的模型参数。

3.2.2 Focal Loss

Focal Loss 主要是为了解决一阶段目标检测方法中正负样本比例严重失衡的问题，大大降低了简单负样本在训练过程中所占的比重，如式（6）：

$$L_{fl}=\begin{cases}-(1-y')^{\gamma}logy', & y=1\\ -y'^{\gamma}log(1-y'), & y=0\end{cases} \quad (6)$$

我们在实验过程中尝试将 Object Loss、Class Loss 由交叉熵损失更换为 Focal Loss，但是最终测试的准确率比先前要低。

3.2.3 测试时增强

测试时增强（Test-time Augmentation）是指在测试过程中，对同一张图片进行旋转、翻转、缩放等增强后进行多次测试，并将所有的测试结果取平均值，作为该图像的最终输出结果。经过实验证明，使用 TTA 后输出的测试结果更加优秀。

4 实验结果

4.1 数据集处理

本文使用的数据集是由大赛官方提供的共 10800 张图片，其中包括用于训练的 9800 张带标签的图片和用于盲测的 1000 张不带标签的图片。数据集包括 6 类检测目标，分别为：liner , container ship , bulk carrier , island reef, sailboat, other ship。为了更好地训练和优化模型，我们利用数据集做了以下两点工作：（1）我们对 9800 张带标签的图片做了详细的属性分析，包含分类别目标总数分析、目标相对位置分析、目标相对大小分析；（2）为了更好地了解模型训练情况，我们按比例划分了训练集和验证集（以下将数据集统称为训练集、验证集、A 测试集）。

对数据集的属性分析是为了方便针对数据集的特点使用更佳的优化策略。对 9800 张图片的标签我们进行了 bbox 可视化，便于直接观察各类别目标的特点及 bbox 的标记情况。各类别目标总数分析是按 6 个类别

对各类目标作了数量统计，统计结果可以显示出数据集中各类别目标的不同数量，如图7(a)所示，从图中可以看出，6 类目标的数目总数呈现明显的类别不平衡现象；目标相对位置分析是统计目标中心点相对于所在图片左上角点的水平及垂直距离，并按图片宽高作归一化处理，获得比例值，绘制分布图，如图 7(b)所示，可以看出目标相对图片的分布位置主要集中在水平方向居中部分，并且分布于图片中心点处目标较多；目标相对大小分析，是统计目标宽高相对于所在图片的宽高的比例值，并绘制分布图，如图 7(c)所示，从分布图可以看出目标整体呈现小目标多，大目标少的特点，在大目标部分中宽度大、高度小的目标又相对较多，即存在部分体型较大的船舶或其它类别目标。

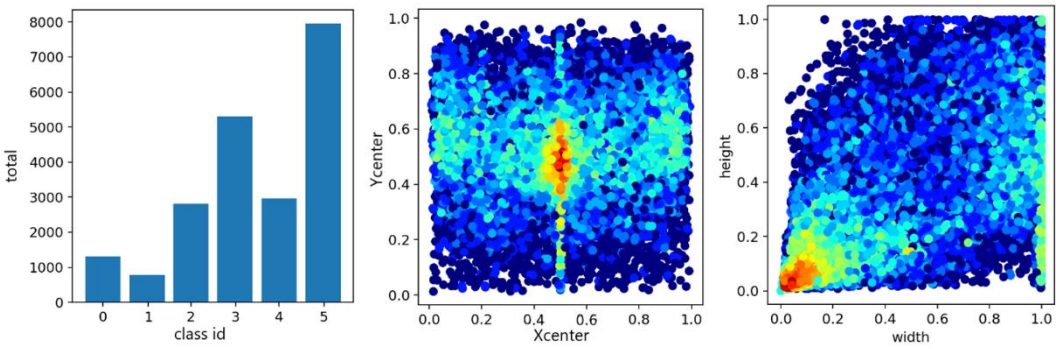

图 7 数据集目标属性分析。(a) 显示了数据集中 6 个类别目标的总数；(b)是目标中心点的相对位置分布图；(c)是目标尺寸相对大小分布图

Fig.7 Attribute analysis of objects in dataset. (a) shows the total of the six categories' objects in the dataset; (b) shows the relative position distribution map of the objects' center point; (c) shows the relative size distribution map of the objects' size

4.2 模型 baseline 选取

基于前一节中数据集的特点分析，结合当前目标检测研究领域的现状，我们从一系列最先进(state-of-the-art)的方法中选取了实验所用的 baseline。我们从 one-stage 和 two-stage 两类目标检测算法中选取了部分性能优异的算法，按检测方法又可以分为基于 anchor 的方法和 anchor-free 的方法。

早期的检测算法中，two-stage 方法的检测准确率普遍优于 one-stage 方法，虽然其速度不占优势，但完全适用于更侧重准确率的任务。考虑这一点，我们从 two-stage 方法中选取了 Faster R-CNN 模型。此外，由于近几年来，研究者们越来越重视检测模型的实际使用效果和研究项目的工程化落地，对于模型检测速度更加关注。即时使用高性能设备，two-stage 缓慢的检测速度和极低的检测帧率仍然无法满足实时检测的要求。相较于 two-stage 缓慢的检测速度，one-stage 方法自诞生以来一直以较高的检测帧率为其主要优势，虽然在早期发展的模型中其精度较 two-stage 方法低，但是 one-stage 方法在实时视频目标检测中完全能胜任工作。随着近年的发展，one-stage 方法已经在检测精度上有了很大进步，部分检测器在精度上已经超过了部分 two-stage 方法，同时在速度上仍然保持巨大的优势。据此我们选取了 yolo 系列中最新的 yolov4、yolov5、scaled-yolov4 三种模型，以及 EfficientDet、CenterNet、FCOS 模型，其中 CenterNet 是我们基于论文中关键点检测方法，使用 Encoder-Decoder 替换主干网络改进而来的，以下简称为 E-D CenterNet。

对于以上共七种模型实验，我们均使用迁移学习方法，加载公有数据集上的预训练参数，在同一设备上做了模型训练、验证和结果测试，如表 1 所示。EfficientDet-D6 是我们根据设备性能从 D1-D7 中选取的认为性能较优的模型，但是初期我们对数据集的误操作导致实验结果无效，无法明确其在本数据集上的性能。同时我们在 YOLOv5 模型

表 1 部分最先进的模型在实验数据集上的精度比较
Table 1 Comparison of the most advanced models' accuracy on the experimental dataset

Type	Model	Val-mAP	ATest- mAP
one-stage	YOLOv4	39.20	39.64
	YOLOv5-X	60.75	60.95
	Scaled-YOLOv4-P6(+TTA)	64.70	65.00
	EfficientDet-D6	-	-
	FCOS	-	52.80
	E-D CenterNet	49.00	52.27
two-stage	Faster R-CNN	49.60	50.01

上取得了领先的成绩，因此暂停了此模型的实验。考虑到它在当前 SOTA 榜上仍然名列前茅，在 COCO 数据集上性能指标大幅领先于很多最新模型，我们认为它在本次数据集上仍然可能取得更好的结果，因此我们在本文中保留了这一模型的实验记录，作为后续继续改进策略的参考。FCOS 模型没有采用划分验证集的方法，直接将 9800 张图片作为训练集，只获取了 A 测试集的结果，这种扩充训练集的方法在后面实验中也被我们采纳为一种改进策略。其它 5 种模型我们分别列出了验证集 mAP 和 A 测试集 mAP 结果，其中 Scaled-YOLOv4-P6 的 A 测试集结果直接使用了 TTA 策略，这是我们在之后实验中广泛采用一种改善检测性能的 trick，其对于 mAP 的提升效果大约在 1%-2%左右，基于模型的优劣上下浮动。可以看到，我们划分的验证集的结果与 A 测试集的结果相差比较小，这充分证明了划分验证集的策略非常恰当，在训练过程中加入验证集的验证，有利于观察模型的收敛情况，也可以快速评估模型的优劣，最重要的是可以根据验证集的 mAP 从众多训练代数中筛选出最佳的模型参数。参考上述模型的实验结果之后，我们选取了 Scaled-YOLOv4-P6 作为方案的 baseline，以下改进策略都是围绕该 baseline 模型开展的。

4.3 Test-Time Augmentation

Test-Time Augmentation (TTA) 是对测试数据集进行数据扩充的方法，它的使用可以改进测试性能。使用 TTA 时会将图片左右翻转，并在三个不同的分辨率下进行操作，将这些增强的图片送入模型进行检测，检测结果合并后再进行非极大值抑制 (NMS) 处理，增加了 2-3 倍的推理时间，但能有效提高大约 1%-2%的 mAP。

实验中，我们首先在 YOLOv5 上增加附带 TTA 的测试实验，如表 2 所示，A 测试集的 mAP 有显著提高，从 baseline 实验中的 60.95 上升到 63.13，提高了 2.18。优异的性能提升效果使得我们确定了 TTA 作为我们的一种常规改进策略，并直接在 Scaled-YOLOv4 中使用了该策略，实现了 65.00 的 mAP 分数。

表 2 模型在实验数据集上的精度比较
Table 2 Comparison of the most advanced models' accuracy on the experimental dataset

Model	WithoutTTA	WithTTA
YOLOv5	60.95	63.13
Scaled YOLOv4 P6	-	65.00

4.4 IoU 阈值

IoU 阈值是非极大值抑制 (NMS) 中使用的关键参数。基于 anchor 的检测方法一般都需要进行后处理操作，使用非极大值抑制对重复冗余的目标框进行剔除，最后得到适当数量的目标框作为检测结果。这也是这类检测器的关键步骤，其中 IoU 阈值的选取非常关键，它决定着目标框的筛选，如图 8 所示，IoU 阈值设置较大时，如 iou-thre=0.95 时，表示当目标框与最大置信分数目标框的交并比大于 0.95 时，才会认为这些目标框是冗余的，会被剔除；当 iou-thre=0.50 时，表示当目标框与最大置信分数交并比大于 0.50 时，就会认为这些目标框是冗余的，此时剔除的冗余框会更多，保留的目标框更少。

因此不同的IoU阈值选取直接影响着检测结果。

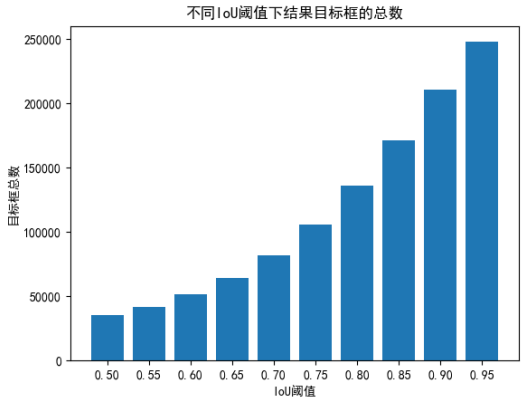

图 8 不同 IoU 阈值下结果目标框的总数

Fig.8 The total of result's object boxes under different IoU thresholds

在 4.1 节中我们对数据集进行了目标属性分析，发现此数据集存在目标类别不平衡、目标大小不均匀等特点，尤其是小目标数量较多，而小目标检测本身就是当前目标检测领域的难题。通过在一组不同 IoU 阈值设定下进行测试实验，分析其各类别的 AP 值，结果数据如表 3 所示。我们观察到 IoU 阈值的选取会影响到不同尺寸目标的检测精度，也会影响到同条件下 APval50 和 APval75 的值。而 COCO 格式的 AP 计算方法是统计 IoU 阈值从 0.50 到 0.95 之间十个参数下的各个目标类别的平均 AP，对十个值取平均值后得到的 mAP，它可以衡量模型整体的精度。数据集中小目标的数量相对较多，其在 APval 中占据的比重也较大，当 Aps 值提升时，APval 也会获得一定的提升

表 3 不同 IoU 阈值下的各类别精度比较

Table 3 Comparison of each category's accuracy under different IoU thresholds

IoU thresholds	APs	APm	APl	APval	APval50	APval75
0.50	21.2	45.1	73.2	64.7	84.9	68.8
0.55	21.3	45.2	73.4	64.9	84.8	69.2
0.60	21.4	45.3	73.6	65.1	84.6	70.0
0.65	21.7	45.4	73.7	65.3	84.4	70.6
0.70	22.3	45.3	73.8	65.4	84.0	71.2
0.75	22.5	45.2	73.9	65.5	83.4	71.5
0.80	22.7	44.9	73.8	65.3	82.5	71.3
0.85	21.6	44.2	73.3	64.6	80.8	70.6
0.90	19.0	41.6	72.0	62.9	77.4	68.7
0.95	14.5	34.9	67.3	57.4	68.5	62.5

从表 3 中可以看到，当 IoU 阈值 0.80 时，小目标检测精度 Aps 相比 IoU 阈值取 0.50 时提高了 1.5，从 21.2 上升到 22.7，APm 出现 0.2 的小幅下降，APl 从 73.2 上升到 73.8，综合来看，整体精度 APval 从 64.7 提高到 65.3，提高了 0.8。另外由于 APval 是各尺度目标检测精度的综合结果，同时受其它尺寸的目标检测精度的影响，而从表中可以看到当取不同的 IoU 阈值时，APm 和 APl 都会出现不同幅度的波动，单纯的提高小目标检测精度 Aps，并不能保证 APval 一定提高。此外，我们发现 IoU 阈值的取值对精度的影响并非线性的，当IoU阈值超过0.75后，各个类别的 AP 都出现不同程度的下降，且下降幅度较大。我们对数据集进行了仔细检测，发现造成这一现象的主要原因是由于数据集中存在较多数量的多目标重叠度很高的图片数据，并且这些目标多是同类别目标，其中小目标也相对较多。这些包含高重叠度目标的图片通过模型检测后得到大量的目标框，在不同的 IoU 阈值下，通过 NMS 后处理会产生有一定差异的结果集，而这些不同的结果集作为测试结果进行 AP 评估时，也会获得不一样的精度结果。

我们在不同 IoU 阈值下，置信度阈值设为 0.4，对验证集的部分目标重叠度很高的图片进行了可视化实验。对于编号为 08753 的图片，其真值标签一共 17 个类型为 other

ship 的目标，并且标注存在较高的重叠度。当IoU阈值设为0.50时，经过NMS后处理，结果保留了 15 个 other ship 目标；当 IoU 阈值设为 0.75 时，结果包含 24 个 other ship 目标；当 IoU 阈值设为 0.95 时，结果包含 130 个目标。其它图片也显示出同样的效果，如下图所示。

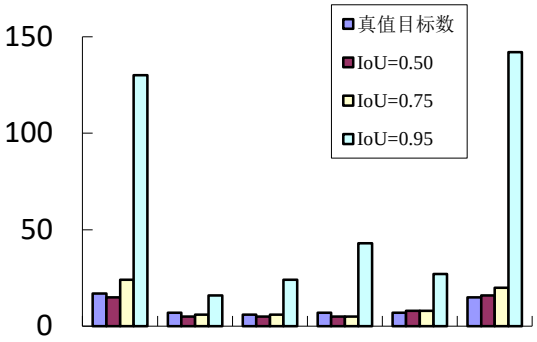

图 9 部分图片在不同 IoU 阈值下结果目标框总数对比

Fig.9 Comparison of object boxes' total for some pictures under different IoU thresholds

如图 9 所示，当 IoU=0.50 时，结果包含的目标数可能少于真值目标数，即出现漏检的情况，而当 IoU=0.95 时，保留目标过多，超过真值目标数几倍，易造成错检多捡等情况，造成精度下降。IoU=0.75 时，结果目标数基本接近真值目标框数，此时这类图片的检测准确率得到提升，整体的 AP 在此时能达到最大值 65.5。设置 IoU 阈值=0.75，使我们的模型在 A 测试集上获得 65.58 的精度。

4.5 非正向效果的策略

除了前面叙述的有效改进策略之外，我们还尝试了一些其它的改进策略，这些策略没取得正向的效果，但我们仍然想将其列举出来，以供参考。

4.5.1 加深网络宽度

加大网络宽度通常可以获得更丰富的特征信息，我们将主干网络每层宽度提高20%，在相同实验条件下进行训练和验证，APval 为 58.6，性能退化。我们分析主要原因可能是修改网络模型后，由于时间关系，没有做公有数据集上的训练，没有利用迁移学习，数据量太小，模型训练没有拟合好；另一方面，可能更多的卷积核导致特征信息提取冗余，不利于目标检测。

4.5.2 Adam 优化器

考虑到 SGD 优化器中学习率选取不当容易使训练陷入局部最优，在加深网络宽度的基础上，我们将其改为 Adam 优化器，同时学习率从 0.01 降至 0.001，试图进一步推进模型的收敛，但是性能出现大幅下降，APval 仅为 51.9，我们分析认为选取最优学习率 0.01 的 SGD 优化器的效果比 Adam 优化器要好。

4.5.3 FocalLoss 损失函数

结合 4.1 节进行的数据集目标属性分析，数据集存在明显的类别不平衡特点，并通过一系列模型的验证实验，我们发现数据集还存在识别难易目标数量不对等的现象，考虑可以将类别损失中使用的 BCE 损失函数改为 FocalLoss 损失函数，以解决类别不平衡现象，调整难易样本在损失函数中的比例，使得训练更侧重于难识别样本，以此提高模

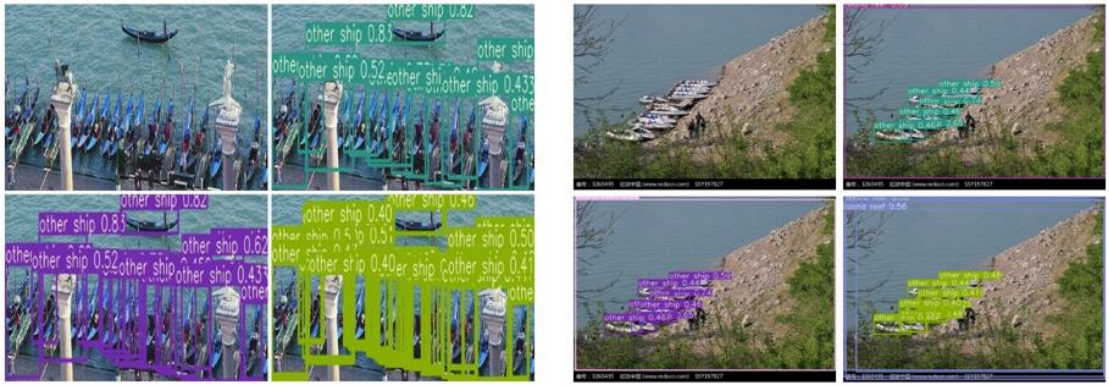

图 10 部分图片在不同 IoU 阈值下结果对比

Fig.10 Comparison of results for some pictures under different IoU thresholds

型精度。但是实验效果不佳，精度下降 0.7；分析原因可能是参数 γ 的取值没有设置最优，由于时间限制，没有再做进一步的取值对比实验。

4.5.4 扩充训练集

增加这一策略的原因主要是我们在实验后期，考虑到划分训练集和验证集的操作使得模型训练牺牲了 10%的数据，在当前训练数据下，模型训练可能已经达到最优化了。我们打算加入 10%的验证集，使用全部 9800 张带标签的图片训练模型，以获得更加丰富的数据信息，使模型学到更多特征信息，提高精度，并加强模型泛化能力。继续前面最佳实验权重进行训练，设置不同的 epoch，但是精度都出现了下降，主要原因是训练过程中没有验证数据的参考，无法判断在哪一轮训练后模型达到最优，需要反复修改 epoch 总数进行实验，获得最佳结果。

5 结论

我们设计了一种基于深度学习的高精度海洋目标检测方法，在现有的 SOTA 方法中，选取了最优的 baseline 模型，结合数据集特点，对模型检测器进行了大量优化改进，探讨了 TTA 引入对结果精度的影响，改进 IoU 阈值提高小目标检测精度，所有改进都进行了详细的实验验证分析，最终采用了经过验证的最优实验方法和参数，在首届海洋目标智能感知国际挑战赛中获得了总检测精度第二的分数，结果也显示我们的方法模型泛化能力更强，这可能为其它海洋目标检测领域提供参考和帮助。

References

1. Liu, Wei, et al. "Ssd: Single shot multibox detector." European conference on computer vision. Springer, Cham, 2016.
2. Ren, Shaoqing , et al. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." IEEE Transactions on Pattern Analysis and Machine Intelligence 39.6(2015).
3. Kaiming, He , et al. "Mask R-CNN." IEEE Transactions on Pattern Analysis & Machine Intelligence PP(2017):1-1.
4. Redmon, Joseph, and Ali Farhadi. "YOLO9000: better, faster, stronger." Proceedings of he IEEE conference on computer vision and pattern recognition. 2017.
5. Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
6. Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." arXiv preprint arXiv:1804.02767 (2018).
7. Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. "YOLOv4: Optimal Speed and Accuracy of Object Detection." arXiv preprint arXiv:2004.10934 (2020).
8. Wang, Chien-Yao, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. "Scaled-YOLOv4: Scaling Cross Stage Partial Network." arXiv preprint arXiv:2011.08036 (2020).
9. Law, Hei, and Jia Deng. "Cornernet: Detecting objects as paired keypoints." Proceedings of the European Conference on Computer Vision (ECCV). 2018.
10. Zhou, Xingyi, Dequan Wang, and Philipp Krähenbühl. "Objects as points." arXiv preprint arXiv:1904.07850 (2019).
11. Lin, Tsung-Yi, et al. "Focal loss for dense object detection." Proceedings of the IEEE international conference on computer vision. 2017.
12. Tian, Zhi, et al. "Fcos: Fully convolutional one-stage object detection." Proceedings of the IEEE international conference on computer vision. 2019.
13. Zhang, Shifeng, et al. "Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.
14. Li, Xiang, et al. "Generalized Focal Loss: Learning Qualified and Distributed Bounding Boxes for Dense Object Detection." arXiv preprint arXiv:2006.04388 (2020).
15. Tan, Mingxing, Ruoming Pang, and Quoc V. Le. "Efficientdet: Scalable and efficient object detection." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.
16. Liu, Shu, et al. "Path aggregation network for instance segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
17. Misra, Diganta. "Mish: A self regularized non-monotonic neural activation function." arXiv preprint arXiv:1908.08681 (2019).
18. Ramachandran, Prajit, Barret Zoph, and Quoc V. Le. "Searching for activation functions." arXiv preprint arXiv:1710.05941 (2017).
19. Wang, Chien-Yao, et al. "CSPNet: A new backbone that can enhance learning capability of cnn." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020.
20. Zheng, Zhaohui, et al. "Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression." AAAI. 2020.
21. Lin, Tsung-Yi, et al. "Feature pyramid networks for object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

作者一　文章末页需附作者中英文简介，简介不超过 100 字.

.

(FIRST Author-Aa　Authors of accepted Papers are requested to supply their biographies (100 words or less). For style, see biographies in the latest issue of Acta Automatica Sinica.)