【引用格式】罗逸豪，孙创，邵成，等. 基于深度学习的水面无人艇目标检测算法综述[J]. 数字海洋与水下攻防，2022，5（6）：524-538.

基于深度学习的水面无人艇目标检测算法综述

罗逸豪[1, 2]，孙　创[1, 2]，邵　成[1, 2]，张钧陶[3]

（1. 中国船舶集团有限公司第七一〇研究所，湖北 宜昌 443003；
2. 清江创新中心，湖北 武汉 430076；
3. 军事科学院系统工程研究院，北京 100141）

摘　要　随着人工智能的发展，水面无人艇可代替人工进行危险任务作业，目标检测是其完成自主探测的核心技术。深度学习技术克服了人工特征提取精度低、通用性差等局限性，已成为图像处理的主流方法。首先，对当前基于深度学习的目标检测算法的发展现状进行了全面总结，对算法分类进行了详细的定义，并指出了不同类型算法的优缺点及适用场景；然后，分析了无人艇水面目标检测技术的研究现状，指出了各类深度学习工作的贡献、优势和局限性；最后，总结了面向水面无人艇的深度学习目标检测算法中亟需解决的关键科学问题，并对可行的方案以及该应用研究领域的未来发展做了进一步的展望。

关键词　水面无人艇；图像处理；目标检测；深度学习

中图分类号　TP391.4　　**文献标识码**　A　　**文章编号**　2096-5753(2022)06-0524-15

DOI　10.19838/j.issn.2096-5753.2022.06.007

Review on Object Detection Algorithm for Unmanned Surface Vehicle Based on Deep Learning

LUO Yihao[1, 2]，SUN Chuang[1, 2]，SHAO Cheng[1, 2]，ZHANG Juntao[3]

（1. NO.710 R&D Institute，CSSC，Yichang 443003，China；
2. Qingjiang Innovation Center，Wuhan 430076，China
3. Institute of System Engineering，Academy of Military Sciences，Beijing 100141，China）

Abstract　With the development of artificial intelligence，the unmanned surface vehicles can replace manual operations for dangerous tasks，and object detection is the core technology for autonomous detection. The deep learning technology could overcome the limitations of low accuracy and poor versatility of manual features，and has become the mainstream method of image processing. Firstly，this paper has comprehensively summarized the current development status of deep learning-based object detection algorithms，defined the classification of algorithms in detail，and pointed out the advantages，disadvantages and applicable scenarios of different types of algorithms. Then，the research status of unmanned surface vehicle object detection technology is analyzed，and the contributions，advantages and limitations of various types of deep learning are pointed out. Finally，the key scientific problems that need to be solved urgently in deep learning object detection algorithm for unmanned surface vehicle are summarized. Meanwhile，the feasible solutions and the future development of this application research field are further prospected.

Key words　unmanned surface vehicle；image processing；object detection；deep learning

收稿日期：2022-09-20

作者简介：罗逸豪（1995-），男，博士，主要从事深度学习、计算机视觉方向研究。

0　引言

水面无人艇（Unmanned Surface Vehicles，USV）作为一种无人操作的水面舰艇，具有体积小、航速快、机动性强、模块化等特点，可用于执行危险以及不适于有人船执行的任务[1]。USV 可实现自主规划与航行、环境感知、目标探测、自主避障等功能，在军事作战和民用领域中具备极高的应用价值[2]。其中无人艇自主目标检测算法是支撑任务完成的核心技术[3]。目前国内 USV 尚未进行大规模应用，一个重要的原因就是水面目标检测算法性能不足。如何提高目标检测的精度和速度，增强应对复杂场景的稳定性，以及扩充识别目标的种类，都是水面目标识别中需要解决的问题。

USV 的感知模块通常可采用以下传感器采集信息：导航雷达、激光雷达、声呐、红外热成像仪、可见光传感器。可见光相机作为轻量级、低功耗和信息丰富的传感器，虽然容易受到光照、天气等环境影响，但已成为 USV 水面目标检测的主流传感设备[4]。

可见光图像目标检测的研究可以追溯到 20 世纪 90 年代，早期的传统目标检测算法基于人工设计的特征，比如十分经典的 SIFT[5]、HOG[6]、Haar[7] 特征。然而，它们能够提取的特征信息往往局限于纹理、轮廓等，只适用于特定任务，并且需要大量的专业经验和知识进行手工设计[8]。而目前各式各样的应用环境充满着许多复杂因素和干扰，传统方法已经显得无能为力。2012 年，AlexNet[9]采用卷积神经网络（Convolutional Neural Network，CNN）在 ImageNet[10]大规模图像分类数据集上取得了突破性的效果，引发了深度学习（Deep Learning）的火热浪潮。深度学习利用大数据对网络模型进行训练，克服了传统特征的诸多缺点，已成为当下各个应用领域中目标检测任务的主流算法。

USV 水面目标检测任务是通用目标检测算法的一个重要应用方向。已有一些综述文献[11-14]对传统或基于深度学习的目标检测算法研究现状进行了综述，但它们仅采用经典的算法类型定义，并未囊括在此类型之外的最新相关工作。另一方面，文献[15-17]对无人水面艇感知技术发展进行了调研与展望，包含了检测、跟踪、定位、导航等多项技术，但未对水面目标检测进行全面深入的分析。

1　基于深度学习的目标检测算法

目标检测算法需要输出给定图像中所有物体的类别，还需用紧密的外接矩阵定位每一个目标，即分类+回归。通俗来讲，目标检测就是解决图像中所有物体“是什么”以及“在哪里”的问题。在 2012 年以前，传统的目标检测算法采用手工方式提取特征，其框架图如图 1 所示。

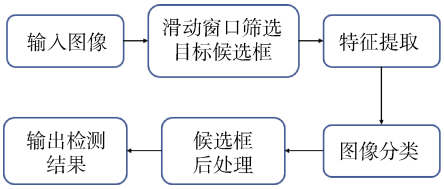

图 1　传统目标检测模型框架

Fig. 1　Framework of traditional object detection model

基于滑动窗口的筛选方法旨在枚举出输入图像中所有可能的目标外接矩形框，最终得到一系列不同大小和尺寸的初始候选框（Anchor，也称为锚框，样本参考框）。然后从输入图像中截取每一个候选框中的图像输入特征提取算法得到图像特征。得到的特征（比如 SIFT、HOG 等手工特征）被输入到分类器（比如 SVM[18]等）中以执行图像分类。最后通过后处理步骤（比如非极大值抑制[19]，Non-Maximum Suppression，NMS）根据分类得分筛选出置信度高的候选框以得到最终的检测结果。

伴随着 2012 年 AlexNet[9]兴起的深度学习研究热潮，深度神经网络（Deep Neural Network，DNN）已经成为了计算机视觉领域中提取图像特征的主流模型。在图像分类任务中 DNN 取得了杰出的精度提升，因此人们自然而然地将其引入到目标检测问题中，将传统目标检测框架中的各个组件由 DNN 进行替换，最终实现“输入→深度学习模型→结果”的端到端模型，具体框架如图 2 所示。

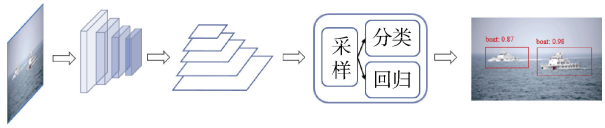

图 2 基于深度神经网络的目标检测模型框架
Fig. 2 Framework of object detection model based on DNN

不同类型的目标检测算法具有不同的采样策略。根据是否存在显式的候选框提取过程，目标检测模型可以分为两阶段（Two-stage）和一阶段（One-stage）检测方法。两阶段检测器通过候选框提取方法首先筛选出一批感兴趣区域（Region of Interest，ROI），然后再进行识别定位，整体上是一种由粗到精的检测过程；而一阶段检测器是直接使用固定的锚框进行识别定位，属于“一步到位”。这也是经典的目标检测算法分类方法。

另一方面，根据是否需要显式定义先验锚框，目标检测模型还可以分为基于锚框（Anchor-based）和无锚框（Anchor-free）检测方法。基于锚框的方法需要预先定义一定数量、尺寸、长宽比例的锚框以进行采样，而无锚框方法则不需要。大部分两阶段目标检测模型属于基于锚框的方法，而一阶段目标检测模型则两者皆有。在 2018 年左右，无锚框检测方法逐渐受到研究者的重视。

除此之外，Transformer[20]作为一种最早用于序列建模和机器翻译任务的基于注意力结构，在最近两年被广泛应用于图像目标检测领域。它提供了一个新的基于目标查询的采样方式，将目标检测作为一个集合预测问题。

接下来本章对上述类型的目标检测算法分别进行阐述。

1.1 两阶段目标检测算法

R-CNN[21]是基于深度学习的两阶段目标检测器开山之作，在传统检测框架上，它采用 CNN 来提取图像特征。R-CNN 检测器在第一个阶段中采用选择性搜索算法[22]生成约 2 000 个 ROI。不同于传统的滑动窗口，选择性搜索算法可以排除掉一部分背景区域的干扰而尽可能筛选出目标区域。在第二阶段中 R-CNN 将每一个 ROI 裁剪并缩放至同样的大小，然后使用 CNN 提取图像特征。最后将特征向量输入到训练好的 SVM 分类器和回归器中得到分类置信度得分和目标边界框的坐标参数。虽然 R-CNN 相比传统检测算法实现了更高的精度，但是它将每一个 ROI 分别输入 CNN 提取特征，这导致了大量的重复计算，致使算法实时性过低，每张图像的计算时间接近一分钟。同时 R-CNN 中的各个组件是独立的，无法以端到端的方式进行训练和推理。

针对 R-CNN 的推理速度不足，SPPNet[23]直接使用 CNN 提取整张输入图像的特征，然后将特征图输入空间金字塔池化层得到固定长度的特征，最后进行分类和回归。类似地，Fast R-CNN[24]采用 ROI 池化层处理整张特征图以提取固定大小特征，然后输入到由全连接层构造的分类器和回归器中。虽然它们在一定程度上提升了检测器的速度，但是由于候选框提取算法的限制依然无法实现端到端检测。

为了实现快速的端到端目标检测模型，Faster R-CNN[25]提出了一种新的候选框提取算法——区域推荐网络（Region Proposal Network，RPN）。RPN 由全卷积神经网络[26]构成，它在输入的特征图中每一个坐标点设置不同比例的固定锚框，输出带有前景/背景二分类结果的候选框。然后，根据所提取的候选框和映射机制可以从特征图上提取一系列 ROI 特征，输入到分类层和回归层得到检测结果。Faster R-CNN 能够以端到端的方式进行训练和推理，极大地提升了检测速度和精度，并且扩展性和泛化性强，成为了经典的两阶段目标检测器范式，被广泛地应用于学术界和工业界。

后续的两阶段目标检测研究主要是基于 Faster R-CNN 的改进工作。R-FCN[27]生成位置敏感度得分图对每个候选框进行编码来提取空间感知区域特征，同时用卷积层替换了分类层和回归层中的全卷积层，实现了更快更准确的检测。Cascade R-CNN[28]提出了一种多阶段的检测模式，通过级联的方式结合多个 R-CNN 结构对回归结果不断地优化，实现了更精准的预测框。Dynamic R-CNN[29]采用动态训练方法来调整训练过程中的 IoU 阈值，

逐步提高锚框的质量。RL-RPN[30]引入了一个顺序区域建议网络，该网络与检测器一起改进搜索策略，优化 RPN 结构。近几年越来越多的两阶段目标检测器被提出，比如 CBNet[31]、DetNet[32]等。

1.2　一阶段目标检测算法

两阶段目标检测器虽然检测精度较高，但是候选区域生成模块会带来更大的计算消耗，降低实际场景应用中的实时性。一阶段检测器没有用于候选框生成的单独阶段，将图像上所有位置都视为可能存在目标，以降低检测精度为代价来提升速度。

OverFeat[33]是第一个采用全卷积神经网络的一阶段目标检测器，它将目标检测看作是多区域分类，直接使用 CNN 来代替滑动窗口。全卷积神经网络的优势在于可以接受任意尺寸的图像输入，而全连接层的劣势正是只支持固定尺寸的输入。尽管 OverFeat 大大提升了检测速度，其粗糙的锚框生成策略和非端到端的训练策略使得它的检测精度不高。

后来 Redmon 等人提出了 YOLO[34]，把输入图像在长宽维度上划分为预设的 N×N 个网格单元。YOLO 将目标检测视为回归问题，并规定每一个网格中都存在同一个类别的一个或者多个预测框，由框的中心点来确定目标所属于的网格。最终每一个网格都会得到 C 个类别的 one-hot 编码概率，B 个预测框的坐标信息和其对应的置信度，输出的特征图尺寸（长×宽×通道）为 N×N×（5B+C）。YOLO 因为其较高的准确率和极快的速度成为了最受欢迎的目标检测模型之一。然而它也有明显的缺点：对于小目标和聚集的物体检测精度不高。这些问题在其后续的版本 v2–v4[35-37]中陆续得到了改善。直至 2022 年，YOLO 已经发展到了第七代[38]，逐渐与无锚框方法相融合。YOLO 系列模型对数据集依赖度不高，运行速度快，是工业界应用最广泛的一阶段目标检测算法。

为了在保证实时性的同时尽可能地提高检测精度，SSD[39]有效地借鉴了 RPN，YOLO 和多尺度检测的思想，仍然将输入图像划分为固定的网格单元，并设定一系列具有多个长宽比例的锚框以扩充预测框的输出空间。每一个预设的锚框都会通过回归器训练得到预测框的坐标，并且由分类器得到（C+1）个类别的概率（1 代表背景类别）。同时，SSD 在多张不同尺寸的特征图上执行目标检测，以更好地发现大、中、小尺寸的目标。SSD 的精度甚至超过了早期的 Faster R-CNN，检测速度比 YOLO 更快，因此备受推崇。基于 SSD 模型的后续研究有 DSOD[40]、RefineDet[41]、MT-DSSD[42]等，它们针对原始方法的跨域预训练、正负样本比例失衡、特征表达能力不强等问题进行优化。

考虑到一阶段探测器和两阶段探测器的精度之间的差异，普遍的观点是认为一阶段目标检测器在训练的过程中存在严重的正负样本不平衡问题，因为未经过筛选的大量锚框只有少量才包含待检测的目标。针对这一现象，RetinaNet[43]改进了交叉熵损失函数的表达式，提出了新的 Focal Loss。它减少了训练过程中简单样本（可以被轻易识别的样本）对于梯度的贡献，使得检测器更加关注容易判错的困难样本。同时，RetinaNet 引入了特征金字塔网络[44]来进行多尺度检测，大幅提高了检测精度。RetinaNet 部署简单，泛化能力强，收敛速度快且易于训练，成为了学术界一阶段目标检测器研究的基线。近几年一阶段检测算法 ATSS[45]、GFL[46]、GFLv2[47]在损失函数上进一步优化，检测精度已与两阶段方法没有差距。

1.3　无锚框目标检测算法

先前介绍的方法多是基于锚框的目标检测算法，这也是自深度学习目标检测研究以来的主流方法。然而，基于锚框的检测算法十分依赖人工预先设置的锚框，需要考虑其数量、尺度、长宽比等因素。当更换数据集之后，预先设置好的锚框参数则需要重新进行设计，这带来了巨大的工作量，使得检测器可扩展性不高。人工设置的锚框参数并不能保证最优，可能会导致训练样本失衡等问题而引起精度下降。同时，生成大量密集的锚框会使得检测器训练和推理的速度降低。因此，近几年无锚框检测算法受到了越来越多研究者的关注，成为了目标检测未来的研究方向之一。

在早期的无锚框方法研究中，UnitBox[48]率先提出了基于交并比（Intersection over Union，IoU）

的回归损失函数。交并比是指在图像中预测框与真实框的交集和并集的面积比值，这也是评价目标检测器精度的主要依据。而主流基于锚框的检测器主要是采用 L1 损失函数，以预测框与真实框的 4 个顶点坐标差的绝对值来计算误差，这与 IoU 不是等价的。极有可能存在具有相同 L1 损失值样本的 IoU 值差异大。IoU 损失函数使得检测器不需要预先设置的锚框，而以像素点为单位来进行预测，开辟了一个新的回归损失范式。

无锚框方法的另一条思路是预测目标框的关键点。CornerNet[49]采用 CNN 提取输入图像特征之后又续接了 2 个独立的分支，上分支负责预测目标框的左上角，下分支则负责预测右下角。上下两分支生成位置热图和嵌入向量，用来判定左上角和右下角是否属于同一个目标，最终使用偏移量误差来进行训练，提升了模型精度。在后续研究中，CenterNet[50]又引入了物体中心点预测来提高检测精度，ExtremeNet[51]则是采用最顶部、最左侧、最底部、最右侧 4 个极值点进行预测。

之后，FCOS[52]在结合了 Focal Loss 和 IoU Loss 的基础上，又提出了 Center-ness Loss。它将落入真实框内的坐标点视作正样本，以坐标点到真实框四条边的距离进行回归，有助于抑制低质量边界框的产生，大幅提高检测器的整体性能。Center-ness Loss 还保证了不同尺度的目标都具有足够数量的正样本，在一定程度上解决了正负样本不平衡问题，成为了代表性的无锚框检测算法配置。FSAF[53]和 Foveabox[54]同样也是采取与 FCOS 类似的思路：在 RetinaNet 检测器上添加无锚检测分支以优化预测框。最近 ObjectBox[55]不仅泛化性良好，而且超越了以往绝大多数方法的检测精度。

1.4 Transformer 目标检测算法

Transformer 模型最早出现在自然语言处理领域，最近两年许多研究者将其应用于计算机视觉，在检测、分割、跟踪等任务中均取得了优异的性能。

DETR[56]是端到端 Transformer 检测器的开山之作，它消除了手工设计的锚框和 NMS 后处理，并通过引入目标查询和集合预测直接检测所有对象，开辟了新的检测算法框架。具体地，DETR 使用编码器-解码器作为颈部模块，使用前馈网络（Feed Forward Networks，FFN）作为检测头部。输入由 CNN 主干提取，展平成一维序列，附加位置编码，然后输入到编码器。设计基于目标查询的可学习位置编码附加到输入，然后并行地传输给解码器。训练过程中，在预测框和真实框之间应用二分匹配损失匹配，以识别一对一标签分配。DETR 实现了具有竞争力的检测精度，但在小型目标上存在收敛速度慢和性能差的问题。

为了解决此问题，可变形 DETR[57]提出了可学习的稀疏注意力机制，用于加速收敛，并引入了多尺度检测结构，提升了小目标进车精度并将训练次数减少了 10 倍。ACT[58]消除编码器的冗余查询，提出了一种自适应聚类转换器，基于多轮精确欧几里德局部敏感度哈希方法，ACT 可以动态地将查询聚类到不同的原型中，然后通过将每个原型广播到相应的查询中，使用这些原型来近似查询关键注意力热图。与 DETR 相比，ACT 降低 15 GFLOPs 的运算量，仅损失 0.7%的平均精度。

DETR 还可以引入空间先验知识，与基于锚框的方法相结合。为了增强目标查询和边界框与经验空间先验的关系，SMCA [59]提出了一种基于空间交叉注意力机制的一阶段检测方法。其训练次数比 DETR 少 5 倍。Meng 等人提出了条件空间嵌入[60]方法，以空间先验明确表示目标的极端区域，从而缩小了定位不同区域的空间范围，使 DETR 收敛速度加快了 8 倍。Yao 等人观察到不同的初始化点总是倾向于类似地分布，提出了一种两阶段高校 DETR[61]，包括密集建议生成和稀疏集预测部分，将 DETR 训练次数减少 14 倍。

Transformer 结构还可以应用于目标检测模型的骨干网络和颈部模块，适用于两阶段、一阶段、无锚框等框架中。PVT[62-63]将 Transformer 构造为一个从高到低分辨率的过程，以学习多尺度特征。基于局部增强的结构将骨干网络构造为局部到全局的组合，以有效地提取短距离和长距离视觉相关性，并避免二次计算开销，如 Swin Transformer[64]、ViL[65]和 Focal Transformer[66]。与特征金字塔网络[44]类似，ZHANG 等人通过结合非局部特征和多

尺度特征，提出了 FPT[67]用于密集预测任务。在模型网络构造过程中，Swin Transformer 作为通用的视觉骨干网络，可以广泛应用于图像分类、目标检测和语义分割等任务，突破了 Transformer 检测器的应用局限性。

然而，基于 Transformer 的目标检测算法通常只能在大规模数据集上实现较大的性能提升，无法在训练数据不足的情况下进行良好的推广[68]。可以采用迁移学习[69]的方法，从足够的数据集中预先训练，然后在小型和特定的下游任务中进行微调。

2　无人艇水面目标检测技术

与传统目标检测算法类似，一些早期的研究工作利用人工设计的特征对水面目标检测进行了研究。许多方法将海上物体的检测视为显著性估计问题[70-73]。这些方法假设目标与其所处的直接背景有很好的区别。然而，此假设在很多情况下都不成立，比如在起雾和强光的环境下，以及需要检测视觉上接近于水的物体。经典的背景建模法和帧间差分法也不适合 USV，因为起伏的海面导致 USV 的持续晃动，违反了静态相机假设，导致误报率很高[74]。RAJAN 等人[75]对基于传统视觉的水面目标物体检测和跟踪做了更为全面的综述，本文不再进行赘述。

因为现实水面环境复杂多变，USV 拍摄的可见光图像的图像质量有所欠缺，包括天气起雾、运动模糊、光照变化等；另外，同一类别的水面目标物也可能在尺度、形状、纹理、大小等方面具有较大差异性。这增加了不同环境下的水面目标检测难度，在一定程度上限制了传统目标检测算法的应用范围。而深度学习目标检测算法迅速发展，已成为目前水面目标检测的主流技术。本章将从 3 个方面总结基于深度学习的水面目标检测技术进展。

2.1　两阶段与一阶段检测方法

基于深度学习的目标检测算法在 2018 年之前大多数分为两阶段或一阶段检测方法，因其技术成熟且易于实现，被广泛应用于各个领域。而在无人艇水面目标检测领域，应用深度学习技术起步较晚。

2017 年 KUMAR 等人[76]提出了一种改进的 VGG16[77] 骨干网络用于海面物体的视觉目标检测。该工作发现由于训练数据的缺乏，CNN 规模过大可能会造成过拟合现象。为了解决此问题，LEE 等人[78]采用了预训练的方式，将通用目标数据集上训练好的模型进行微调，以适用于海事目标。

在之后的研究工作中，经典的两阶段检测模型 Faster R-CNN 被频繁采用。FU 等人[79]使用了一种改进的 Faster R-CNN 方法用于海上目标检测，使用层数更深、功能更强大的 ResNet[80]骨干网络提取特征，并利用深度归一化层、在线难样本挖掘对模型进行优化。CHEN 等人[81]将多尺度策略融合到了 ResNet 的多层卷积中，并在特征图上添加了双线性插值进行上采样，以增强小目标检测的效果。Yang 等人[82]提出了一个基于 CNN 的水面目标检测和跟踪定位系统，以 Faster R-CNN 模型检测目标位置，然后使用 KFC 算法[83]在视频序列中连续跟踪该目标。在后续研究中，MA 等人[84]采用了混合骨干网络架构，通过 DenseNet[85]与 ResNet 结合的策略，再结合双向特征金字塔网络，进一步增强了两阶段检测模型的精度。

基于两阶段的检测方法倾向于算法精度，但计算复杂度相对更大；相反，一阶段的目标检测识别算法在训练和推理过程占用内存更低，模型计算更快。在不追求更高的检测精度时，一阶段检测方法更受偏爱。陈欣佳等人[86]使用 SSD 模型执行快速的无人艇目标检测任务，并借助相关滤波（Correlation Filter）方法进行快速跟踪。YANG 等人[87]使用 YOLOv3 模型实现了实时的水面无人艇检测，然后通过卡尔曼滤波器将外观特征与运动状态估计相结合，实现了一种基于数据关联的多目标跟踪方法。无独有偶，王飞等人[88]也基于 YOLOv3 开发了海雾气象条件下海上船只实时检测的深度学习算法。王孟月[89]借助 DenseNet 改进 YOLOv3 的骨干网络，以增强特征传播效率、促进有效特征重用以及提高网络性能。

2.2　基于语义分割的检测方法

语义分割需要对图像中每一个像素点进行分类，也就是解决像素级别的“是什么”问题。对分

割结果图像，可以轻松的得到目标的外接矩形框，如图 3 所示。

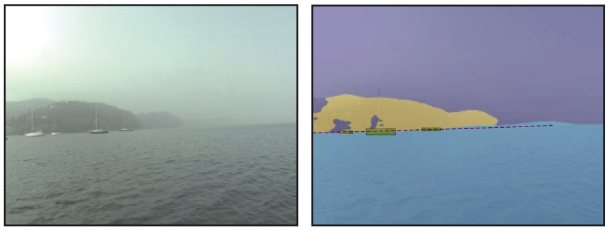

图 3 基于语义分割的检测示意图

Fig. 3 Schematic diagram of detection based on semantic segmentation

由于基于深度学习的语义分割网络模型在城市与道路场景中取得了良好的效果，一些工作[90-91]将 CNN 分割框架用于海上图像分割。为了改进早期方法在小障碍物上分割表现不佳以及镜像混淆的问题，KIM 等人[92]将跳跃连接和白化层应用于 E-Net[93]以改进小目标检测，虽然精度和效率高于同期其他的分割方法，但每秒 10 帧的计算速度依然无法达到实时的检测效果。

在后续的研究中，STECCANELLA 等人[94]提出用深度卷积替换 U-Net[95]中的传统卷积层以改进水线分割效果。在生成了水和非水区域的二进制掩码后，继续检测水中区域的障碍物。为了进一步解决小目标检测精度低和水反射误报率高的问题，BOVCON 等人[96]提出了一种新的深度非对称编码器–解码器架构，设计了注意力机制和新的损失函数，并通过视觉和惯性信息融合提高了整体分割精度。但是基于分割的方法始终难以达到实时检测的效果。

2.3 海事视觉感知数据集

早期有一些数据集用来评估海上监视和机器人导航的各种算法。FEFILATYEV 等人[97]提出了一个数据集，该数据集包含在同一天记录的 10 个序列，在同一片公海采集。然而它仅用于地平线检测评估，不包含障碍物，限制了它们的视觉多样性。BLOISI 等人[98]采集了 10 种海上目标跟踪序列。通过在一天中的不同时间进行记录，增加视觉多样性，并对船舶、船只和喷气式飞机等动态障碍物进行注释。然而，由于所有障碍物在非常明亮的水面上都是黑暗的，它们对目标检测几乎没有挑战性。MARQUES 等人[99]和 RIBEIRO 等人[100]记录了 2 个视觉上不同的海上机载探测数据集。该数据集是为无人机应用而设计的，它不具有在自主船上观察到的有利位置。

为了使数据集信息更加丰富，PATINO 等人[101]提出了一个包含 14 个多传感器序列的数据集，用于障碍物检测、跟踪和威胁识别评估。数据集包含地平线和动态障碍物的注释，但不包含小型障碍物，如浮标。KRISTAN 等人[102]构建了一个海上障碍物检测数据集，其中包含从 USV 捕获的 12 个不同序列，后来 BOVCON 等人[103]将其扩展为与惯性测量单元同步的 28 个立体摄像机序列。2 个数据集都记录在同一个场景，并包含地平线、水边和大小动态障碍物的注释，通过在不同天气条件下进行记录，保持视觉多样性。

由于深度学习模型需要数据驱动，小型数据集会使得深度学习模型出现过拟合的问题。因此，PRASAD 等人[104]提出了一个大型海上监视数据集，包含 51 个 RGB 和 30 个红外光谱序列，在一天的不同时间和不同天气条件下记录。大多数序列是从固定的岸上观测点记录的，而有些是从比机器人船更高的有利位置拍摄的。由于它主要是为监视而设计的，所以场景非常静态，几乎没有运动。为了使动态障碍物和地平线被很好地注释，最近 MOOSBAUER 等人[105]提供了通过基于颜色的半自动方法计算的粗略实例分割标签。GUNDOGDU 等人[106]提出了一个具有 400 000 补丁的数据集，用于轮船分类任务，但该数据集不能用于检测器评估，因为轮船位置没有注释。SOLOVIEV 等人[107]最近构建了具有接近 2 000 张图像的数据集，用于评估预训练的船舶探测器，因此不标注静态障碍物（如海岸）和动态障碍物（例如边界）。

大多数数据集被提出用于评估目标检测算法，只有少数数据集被设计用于训练分割方法。STECCANELLA 等人[108]提出了一个由 191 幅图像组成的逐像素注释数据集，这些图像在 7 种海域中分别单独记录，用于训练和测试分割方法。数据集包含水域和非水域 2 个语义标签，并且测

试集与训练集没有很好地分离，视觉多样性有限。BOVCON 等人[109]提出了目前用于海面图像分割的最大和最详细的数据集。数据集是在不同时间和不同天气条件下记录的，历时 2 年，包含接近 1 300 张图像，每个像素点标记水、天空或者障碍物。

由于在海洋试验现场采集数据成本高昂，许多数据集包含的图像数量较少。在 2022 年，RAZA 等人[110]使用 3D 仿真平台 AILiveSim 构建了一个舰船检测仿真数据集，包含 9471 张高分辨率（1920×1080）图像，具有船舶、岩石、浮标等动态和静态目标，并使用 YOLOv5 测试了模拟数据的可行性。最近，BOVCON 等人[111]构建了目前规模最大、最具挑战性的水面目标检测数据集 MODS，包含了超过 8 万张图像，记录了高度多样化的目标，并且设计了相应的评估方法、训练集和测试集，形成了一项新基准。这项研究工作在开源网站上进行了公开发布，系统地评估了 19 项两阶段、一阶段、基于语义分割的目标检测算法在该基准上的性能并进行排名，使得不同方法的跨论文比较更易实现。该工作使水面无人艇目标检测领域取得了关键进展。

3　水面目标检测关键问题及展望

虽然有许多研究工作将深度学习方法应用于水面目标检测任务中，但仍有一些缺陷和关键问题亟需解决。本章对关键问题进行归纳总结，并对可行的方案以及未来发展做了进一步的展望。

3.1　关键问题

1）缺乏大规模数据集和统一的评价标准。

在通用目标检测研究中，PASCAL VOC[112]数据集是 2015 年以前评价检测算法的金标准，MS COCO[113]数据集则是 2015 年以后的金标准，他们分别具有约 2 万和 16 万张图像。由于其涵盖类别多、场景复杂性高，被研究者们广泛采用，不同算法工作可以轻易地进行性能横向对比。

然而目前的许多海事数据集不能充分捕捉真实世界 USV 场景的复杂性，并且没有标准化评估方法，这使得不同方法的跨论文比较变得困难，阻碍了相关研究的进展。

2）深度学习方法陈旧。

人脸识别[114]和行人检测[115]也作为通用目标检测算法的 2 个应用子问题，分别衍生出了各自的特异性问题和新颖的算法，在现实应用场景中取得了良好的效果。而由前文内容可知，USV 水面目标检测算法的应用相较于通用目标检测算法研究滞后 2 年左右，并且所使用的方法通常为 Faster R-CNN 和 YOLOv3 等经典模型，未引入新的模型和针对于水面情况的算法，性能有待进一步提高。

3）现实场景图像质量不佳。

无人艇面临着不断变化的外部环境和突发因素的影响，例如起雾、雨水、强光、海浪等因素的干扰，复杂的背景以及快速变化的视角，或是摄像设备的突然失焦。这均会使得采集的图像质量不佳，极有可能导致算法误判，在应用场景中产生严重后果。尽管深度学习算法比传统算法的精度和鲁棒性更强，在直接处理受损图像时依然不能达到令人满意的效果。

4）可见光相机信息单一。

单一的传感器不能全面地反映复杂海况，单目可见光相机仅能获取彩色图像，无法获取距离、温度等信息。无人艇系统由各体系模块化组成，可以搭载不同的传感器进行感知探测。因此需要利用雷达、声呐、红外等多种传感器信息进行协同、融合分析，提升系统的整体性能。

5）无法应对特定目标检测任务。

目前水面无人艇目标检测数据集涵盖的目标类别通常为船舶、人、浮标、岩石等常见水面目标。然而，当某些具体的应用场景需要检测数据集中未涵盖的特定目标，现有的 USV 水面目标检测算法难以满足需求。比如，需要检测海域中的冰山，搜寻水域和岸边的濒危两栖动物，在海域作战中检测信号弹、导弹、飞机等空中目标。

3.2　展望

1）大规模数据集下的 Transformer 模型。

由于归纳偏差通常表示为关于数据分布或解空间的一组假设，在 CNN 中表现为局部性和平移不变性。局部性关注空间上紧密的元素，并将它们与远端元素隔离，变换不变性表明在输入的不同位

置重复使用相同的匹配规则。因此 CNN 在处理图像数据中更关注于局部信息，却限制了数据集规模的上限。Transformer 可以关注图像全局信息，在大规模数据集上表现出了更优越的性能。深层 Transformer 骨干网络和编码器–解码器结构可有效降低计算复杂度，避免深层特征过度平滑。

最新提出的 MODS 大规模水面目标检测数据集包含 8 万张图像和超过 6 万个目标标注，有望成为评价水面目标检测算法的金标准。因此，在大规模数据驱动下，可以引入 Transformer 进行模型设计，进一步提升水面目标检测算法的精度和泛化性。

2）新算法与模型的应用。

近几年目标检测算法在多个层面迅速发展。在骨干网络方面，ResNext[116]和 Res2Net[117]已经成为了常用的模型，可以提取表达能力更强的图像特征，并且可变形卷积[118]也被广泛使用。在颈部模块方面，AugFPN[119]和 RCNet[120]联合设计了上下文和注意力模块大幅丰富了多尺度特征信息。在检测头部方面，DOOD[121]和 TOOD[122]分别采用了解耦和联合的策略，进一步提高分类和定位的精度。除此之外还有许多训练策略[123-124]改善了正负样本不平衡问题。对于 USV 水面目标检测任务中环境复杂、小目标漏检、背景区域大等问题，需要借鉴通用目标检测算法，针对性的选择和设计解决方案。

3）基于图像重建与目标检测的多任务模型。

为解决图像质量不佳的问题，最直观的方法是引入图像重建算法对采集的图像进行预处理。CHEN 等人[125]采用偏振成像技术对强反光区域进行抑制，QIAN 等[126]结合生成对抗网络和注意力机制对雨天采集的图像进行去雨处理。然而他们仅针对单一的图像受损因素进行预处理操作，适用范围较小。设计多任务模型[127]进一步提升算法性能十分有必要。

深度学习领域中的多任务学习是指让一个神经网络同时学习多项任务，目的是让每个任务之间能够互相帮助。这有利于提高模型实时性和减少算力消耗。其主要实现方式为参数共享，多个任务之间共用网络模型的部分参数，共同进行端到端训练，产生隐式训练数据增加的效果，增强模型的能力并降低过拟合的风险。多任务模型比独立地训练单个任务能实现更好的效果。

因此，图像重建和目标检测任务可以作为子任务统一至端到端模型，在大规模数据驱动下进行多任务联合学习，提高检测器在恶劣天气条件下的性能。

4）多模态融合算法。

多模态学习即是从多个模态表达或感知事物[128]，比如通过 2 种不同成像原理的相机拍摄的图像，通过图像、音频、字母理解视频。多模态学习通常具有 2 种方式：协作和融合。

在水面目标检测任务中，基于协作的方法可以对相机、雷达、声呐、红外等多种数据的算法输出结果执行进一步分析，采用加权等方式得到最终的检测结果。基于融合的方法可以将多种传感器采集的图像进行融合，进一步探究多模态数据深层特征之间的关系，提高数据的利用率，构建鲁棒的算法系统。例如，MA 等人[129]提出了 Fusion GAN 模型，采用生成对抗网络实现红外与可见光图像融合。同时，随着 3D 目标检测[130]研究的兴起，可以将彩色图像与雷达点云数据进行配准融合[131-132]作为深度神经网络的输入。为提高 USV 感知环境的整体能力，多模态融合算法必将成为重要的发展趋势。

5）小样本、弱监督训练算法。

在特定目标检测任务中存在样本数量少、标注缺失、类别不明确、标注错误等问题。可以借助深度学习小样本学习[133]和弱监督训练[134]的方法，针对特定的水面检测任务充分利用已有的少量图像数据，解决深度学习模型欠拟合和过拟合的问题，提高目标检测算法精度。

4 结束语

水面无人艇在军事作战和民用领域中具备极高的应用价值，目标检测算法是支撑任务完成的核心技术。本文首先回顾了当前基于深度学习的目标检测算法的发展现状，从两阶段、一阶段、无锚框、Transformer 4 个类别进行了全面的总结；然后从两

阶段/一阶段方法、基于语义分割的方法、海事视觉感知数据集 3 个方面归纳无人艇水面目标检测技术的研究现状；最后阐述了水面目标检测任务面临的 4 个关键问题：缺乏大规模数据集和统一的评价标准、深度学习方法陈旧、现实场景图像质量不佳、可见光相机信息单一、无法应对特定目标检测任务，并对多任务、多模态、弱监督等新技术进行了可行性分析和展望。未来，高度智能化的水面无人艇将会成为海事任务的重要力量。

参考文献

[1] 王石，张建强，杨舒卉，等. 国内外无人艇发展现状及典型作战应用研究[J]. 火力与指挥控制，2019，44（2）：11-15.

[2] 王博. 无人艇光视觉感知研究发展综述[J]. 舰船科学技术，2019，41（23）：44-49.

[3] 熊勇，余嘉俊，张加，等. 无人艇研究进展及发展方向[J]. 船舶工程，2020，42（2）：12-19.

[4] ZHANG W，YANG C F，JIANG F，et al. A review of research on light visual perception of unmanned surface vehicles[J]. Journal of Physics：Conference Series，2020，1606（01）：012022.

[5] LOWE D G. Object recognition from local scale-invariant features[C]// IEEE International Conference on Computer Vision. Kerkyra：IEEE，1999.

[6] DALAL N，TRIGGS B. Histograms of oriented gradients for human detection[C]// IEEE Conference on Computer Vision and Pattern Recognition. San Diego：IEEE，2005.

[7] LIENHART R，MAYDT J. An extended set of Haar-like features for rapid object detection[C]// International Conference on Image Processing. Rochester：IEEE，2002.

[8] 路齐硕. 基于深度学习的目标检测方法研究[D]. 北京：北京邮电大学，2020.

[9] KRIZHEVSKY A，SUTSKEVER I，HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM，2017，60（6）：84-90.

[10] RUSSAKOVSKY O，DENG J，SU H，et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision，2015，115（3）：211-252.

[11] WU X，SAHOO D，HOI S C H. Recent advances in deep learning for object detection[J]. Neurocomputing 2020，396：39-64.

[12] ZAIDI S S，ANSARI M S，ASLAM A，et al. A survey of modern deep learning based object detection models[J]. Digital Signal Processing，2022，126：103514.

[13] ZOU Z X，SHI Z W，GUO Y H，et al. Object detection in 20 years：a survey[J/OL]. [2019-05-16]. https：//arxiv.org/pdf/1905.05055.pdf.

[14] JIAO L C，ZHANG F，LIU F，et al. A survey of deep learning-based object detection[J]. IEEE Access，2019，7：128837-128868.

[15] 张安民，周健，张豪. 水面无人艇环境感知技术及应用发展[J]. 科技导报，2021，39（5）：106-116.

[16] 侯瑞超，唐智诚，王博，等. 水面无人艇智能化技术的发展现状和趋势[J]. 中国造船，2020，61（S1）：211-220.

[17] 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.

[18] BURGES C J C. A tutorial on support vector machines for pattern recognition[J]. Data Mining and Knowledge Discovery，1998，2（2）：121-167.

[19] NEUBECK A，VAN GOOL L. Efficient non-maximum suppression[C]// 18th International Conference on Pattern Recognition. Hong Kong：IEEE，2006.

[20] VASWANI A，SHAZEER N，PARMAR N，et al. Attention is all you need[C]// Advances in Neural Information Processing Systems 30. Long Beach：Curran Associates Inc，2017.

[21] GIRSHICK R，DONAHUE J，DARRELL T，et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// IEEE Conference on Computer Vision and Pattern Recognition. Columbus：IEEE，2014.

[22] 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（2）：154-171.

[23] HE K M，ZHANG X Y，REN S Q，et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2015，37（9）：1904-1916.

[24] GIRSHICK R. Fast R-CNN[C]// IEEE International Conference on Computer Vision. Santiago：IEEE，2015.

[25] REN S Q，HE K M，GIRSHICK R，et al. Faster R-CNN：towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2017，39（6）：1137-1149.

[26] SHELHAMER E，LONG J，DARRELL T. Fully

convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2017，39（4）：640-651.

[27] DAI J F，LI Y，HE K M，et al. R-FCN：object detection via region-based fully convolutional networks[C]// Advances in Neural Information Processing Systems 29. Barcelona：Curran Associates Inc.，2016.

[28] CAI Z W，VASCONCELOS N. Cascade R-CNN：delving into high quality object detection[C]// IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City：IEEE，2018.

[29] ZHANG H K，CHANG H，MA B P，et al. Dynamic R-CNN：towards high quality object detection via dynamic training[C]// European Conference on Computer Vision. Glasgow：Springer，2020.

[30] PIRINEN A，SMINCHISESCU C. Deep reinforcement learning of region proposal networks for object detection[C]// IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City：IEEE，2018.

[31] LIU Y D，WANG Y T，WANG S W，et al. CBNet：a novel composite backbone network architecture for object detection[C]// AAAI Conference on Artificial Intelligence. New York：AAAI Press，2020.

[32] LI Z M，PENG C，YU G，et al. DetNet：design backbone for object detection[C]// European Conference on Computer Vision. Munich：Springer，2018.

[33] SERMANET P，EIGEN D，ZHANG X，et al. Overfeat：integrated recognition，localization and detection using convolutional networks[C]// The 2nd International Conference on Learning Representations. Banff：OpenReview.net，2014.

[34] REDMON J，DIVVALA S，GIRSHICK R，et al. You only look once：unified，real-time object detection[C]// IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas：IEEE，2016.

[35] REDMON J，FARHADI A. YOLO9000：better，faster，stronger[C]// IEEE Conference on Computer Vision and Pattern Recognition. Honolulu：IEEE，2017.

[36] REDMON J，FARHADI A. YOLOV3：an incremental improvement[J/OL][2018-04-08]. https：//arxiv.org/pdf/1804.02767.pdf.

[37] BOCHKOVSKIY A，WANG C Y，LIAO H Y M. YOLOv4：optimal speed and accuracy of object detection[J/OL][2020-04-23]. https：//arxiv.org/pdf/2004.10934.pdf.

[38] WANG C Y，BOCHKOVSKIY A，LIAO H Y. YOLOV7：trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[J/OL][2022-07-06]. https：//arxiv.org/pdf/2207.02696.pdf.

[39] LIU W，ANGUELOV D，ERHAN D，et al. SSD：single shot multiBox detector[C]// European Conference on Computer Vision. Amsterdam：Springer，2016.

[40] SHEN Z Q，LIU Z，LI J G，et al. DSOD：learning deeply supervised object detectors from scratch[C]// IEEE International Conference on Computer Vision. Venice：IEEE，2017.

[41] ZHANG S F，WEN L G，BIAN X，et al. Single-shot refinement neural network for object detection[C]// IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City：IEEE，2018.

[42] ARAKI R，ONISHI T，HIRAKAWA T，et al. MT-DSSD：deconvolutional single shot detector using multi task learning for object detection，segmentation，and grasping detection[C]// IEEE International Conference on Robotics and Automation. Paris：IEEE，2020.

[43] LIN T Y，GOYAL P，GIRSHICK R，et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2020，42（2）：318-327.

[44] LIN T Y，DOLLÁR P，GIRSHICK R，et al. Feature pyramid networks for object detection[C]// IEEE Conference on Computer Vision and Pattern Recognition. Honolulu：IEEE，2017.

[45] ZHANG S F，CHI C，YAO Y Q，et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle：IEEE，2020.

[46] LI X，WANG W H，WU L J，et al. Generalized focal loss：learning qualified and distributed bounding boxes for dense object detection[J]. Advances in Neural Information Processing Systems，2020，33：21002-21012.

[47] LI X，WANG W H，HU X L，et al. Generalized focal loss V2：learning reliable localization quality estimation for dense object detection[C]// IEEE Conference on Computer Vision and Pattern Recognition. Virtual Event：IEEE，2021.

[48] YU J H，JIANG Y N，WANG Z Y，et al. UnitBox：an advanced object detection network[C]// ACM Conference on Multimedia Conference. Amsterdam：ACM，2016.

[49] LAW H，DENG J. CornerNet：detecting objects as paired keypoints[J]. International Journal of Computer Vision，

2020，128（3）：642-656.

[50] DUAN K W，BAI S，XIE L X，et al. CenterNet：keypoint triplets for object detection[C]// IEEE/CVF International Conference on Computer Vision. Seoul：IEEE，2019.

[51] ZHOU X Y，ZHUO J C，KRÄHENBÜHL P. Bottom-up object detection by grouping extreme and center points[C]// IEEE Conference on Computer Vision and Pattern Recognition. Long Beach：IEEE，2019.

[52] TIAN Z，SHEN C C，CHEN H，et al. FCOS：fully convolutional one-stage object detection[C]// IEEE/CVF International Conference on Computer Vision. Seoul：IEEE，2019.

[53] ZHU C C，HE Y H，SAVVIDES M. Feature selective anchor-free module for single-shot object detection[C]// IEEE Conference on Computer Vision and Pattern Recognition. Long Beach：IEEE，2019.

[54] KONG T，SUN F C，LIU H P，et al. FoveaBox：beyound anchor-based object detection[J]. IEEE Transactions on Image Processing，2020，29：7389-7398.

[55] ZAND M，ETEMAD A，GREENSPAN M. ObjectBox：from centers to boxes for anchor-free object detection[C]// European Conference on Computer Vision. Tel Aviv：Springer，2022.

[56] CARION N，MASSA F，SYNNAEVE G，et al. End-to-end object detection with transformers[C]// European Conference on Computer Vision. Glasgow：Springer，2020.

[57] ZHU X Z，SU W J，LU L W，et al. Deformable DETR：deformable transformers for end-to-end object detection[C]// International Conference on Learning Representations. Virtual Event：OpenReview.net，2021.

[58] ZHENG M H，GAO P，ZHANG R R，et al. End-to-end object detection with adaptive clustering transformer[C]// British Machine Vision Conference. Online：Springer，2021.

[59] GAO P，ZHENG M H，WANG X G，et al. Fast convergence of DETR with spatially modulated co-attention[C]// IEEE/CVF International Conference on Computer Vision. Montreal：IEEE，2021.

[60] MENG D P，CHEN X K，FAN Z J，et al. Conditional DETR for fast training convergence[C]// IEEE/CVF International Conference on Computer Vision. Montreal：IEEE，2021.

[61] YAO Z Y，AI J B，LI B X，et al. Efficient DETR：improving end-to-end object detector with dense prior[J/OL][2021-04-03]. https：//arxiv.org/pdf/2104.01318.pdf.

[62] WANG W H，XIE E Z，LI X，et al. Pyramid vision transformer：a versatile backbone for dense prediction without convolutions[C]// IEEE/CVF International Conference on Computer Vision. Montreal：IEEE，2021.

[63] WANG W H，XIE E Z，LI X，et al. PVT V2：improved baselines with pyramid vision transformer[J/OL][2022-01-30]. https：//arxiv.org/pdf/2106.13797.pdf.

[64] LIU Z，LIN Y T，CAO Y，et al. Swin transformer：hierarchical vision transformer using shifted windows[C]// IEEE/CVF International Conference on Computer Vision. Montreal：IEEE，2021.

[65] ZHANG P C，DAI X Y，YANG J W，et al. Multi-scale vision longformer：a new vision transformer for high resolution image encoding[C]// IEEE/CVF International Conference on Computer Vision. Montreal：IEEE，2021.

[66] YANG J W，LI C Y，ZHANG P C，et al. Focal self-attention for local-global interactions in vision transformers[J/OL][2021-07-01]. https：//arxiv.org/pdf/2107.00641.pdf.

[67] ZHANG D，ZHANG H W，TANG J H，et al. Feature pyramid transformer[C]// European Conference on Computer Vision. Glasgow：Springer，2020.

[68] LIU Y，ZHANG Y，WANG Y X，et al. A survey of visual transformers[J/OL] [2022-05-02]. https：//arxiv.org/pdf/2111.06091.pdf.

[69] TAN C Q，SUN F C，KONG T，et al. A survey on deep transfer learning[C]// International Conference on Artificial Neural Networks. Rhodes：Springer，2018.

[70] ALBRECHT T，WEST G A W，TAN T，et al. Visual maritime attention using multiple low-level features and naive Bayes classification[C]// International Conference on Digital Image Computing：Techniques and Applications. Noosa：IEEE，2011.

[71] MAKANTASIS K，DOULAMIS A，DOULAMIS N. Vision-based maritime surveillance system using fused visual attention maps and online adaptable tracker[C]// International Workshop on Image Analysis for Multimedia Interactive Services. Paris：IEEE，2013.

[72] SOBRAL A，BOUWMANS T，ZAHZAH E. Double-constrained RPCA based on saliency maps for foreground detection in automated maritime surveillance[C]// IEEE International Conference on Advanced Video and Signal Based Surveillance. Karlsruhe：IEEE，2015.

[73] CANE T，FERRYMAN J. Saliency-based detection for maritime object tracking[C]// IEEE Conference on Computer Vision and Pattern Recognition Workshops.

Las Vegas：IEEE，2016.

[74] PRASAD D K，RASATH C K，RAJAN D，et al. Object detection in a maritime environment：performance evaluation of background subtraction methods[J]. IEEE Transactions on Intelligent Transportation Systems，2018，20（5）：1787-1802.

[75] PRASAD D K，RAJAN D，RACHMAWATI L，et al. Video processing from electro-optical sensors for object detection and tracking in a maritime environment：a survey[J]. IEEE Transactions on Intelligent Transportation Systems，2017，18（8）：1993-2016.

[76] KUMAR A S，SHERLY E. A convolutional neural network for visual object recognition in marine sector[C]// International Conference for Convergence in Technology. Rockville：IEEE，2017.

[77] SIMONYAN K，ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]// International Conference on Learning Representations. San Diego：OpenReview.net，2015.

[78] LEE S J，ROH M I，LEE H W，et al. Image-based ship detection and classification for unmanned surface vehicle using real-time object detection neural networks[C]// International Ocean and Polar Engineering Conference. Sapporo：International Society of Offshore and Polar Engineers，2018.

[79] FU H X，LI Y，WANG Y C，et al. Maritime target detection method based on deep learning[C]// IEEE International Conference on Mechatronics and Automation. Changchun：IEEE，2018.

[80] HE K M，ZHANG X Y，REN S Q，et al. Deep residual learning for image recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas：IEEE，2016.

[81] CHEN W，LI J L，XING J C，et al. A maritime targets detection method based on hierarchical and multi-scale deep convolutional neural network[C]// International Conference on Digital Image Processing. Shanghai：SPIE，2018.

[82] YANG J，XIAO Y，FANG Z W，et al. An object detection and tracking system for unmanned surface vehicles[C]// Target and Background Signatures III：SPIE，2017.

[83] HENRIQUES J F，CASEIRO R，MARTINS P，et al. High-speed tracking with Kernelized correlation filters[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2014，37（3）：583-596.

[84] MA L Y，XIE W，HUANG H B. Convolutional neural network based obstacle detection for unmanned surface vehicle[J]. Mathematical Biosciences and Engineering，2019，17（1）：845–861.

[85] HUANG G，LIU Z，VAN DE MAATEN L，et al. Densely connected convolutional networks[C]// IEEE Conference on Computer Vision and Pattern Recognition. Honolulu：IEEE，2017.

[86] 陈欣佳，刘艳霞，洪晓斌，等. 基于 SSD-CF 的无人艇目标检测跟踪方法[J]. 中国测试，2019，45（2）：145-150.

[87] YANG J，LI Y H，ZHANG Q N，et al. Surface vehicle detection and tracking with deep learning and appearance feature[C]// International Conference on Control，Automation and Robotics. Beijing：IEEE，2019.

[88] 王飞，刘梦婷，刘雪芹，等. 基于 YOLOv3 深度学习的海雾气象条件下海上船只实时检测[J]. 海洋科学，2020，44（8）：197-204.

[89] 王孟月. 面向无人艇的水面目标检测识别与跟踪方法研究[D]. 武汉：华中科技大学，2020.

[90] CANE T，FERRYMAN J. Evaluating deep semantic segmentation networks for object detection in maritime surveillance[C]// IEEE International Conference on Advanced Video and Signal Based Surveillance. Auckland：IEEE，2018.

[91] BOVCON B，KRISTAN M. Obstacle detection for USVs by joint stereo-view semantic segmentation[C]// IEEE/RSJ International Conference on Intelligent Robots and Systems. Madrid：IEEE，2018.

[92] KIM H，KOO J，KIM D，et al. Vision-based real-time obstacle segmentation algorithm for autonomous surface vehicle[J]. IEEE Access，2019，7：179420-179 428.

[93] PASZKE A，CHAURASIA A，KIM S，et al. ENet：a deep neural network architecture for real-time semantic segmentation[J/OL][2016-06-07]. https：//arxiv.org/pdf/1606.02147.pdf.

[94] STECCANELLA L，BLOISI D D，CASTELLINI A，et al. Waterline and obstacle detection in images from low-cost autonomous boats for environmental monitoring[J]. Robotics and Autonomous Systems，2020，124：103346.

[95] RONNEBERGER O，FISCHER P，BROX T. U-Net：convolutional networks for biomedical image segmentation[C]// International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich：Springer，2015.

[96] BOVCON B，KRISTAN M. A water-obstacle separation and refinement network for unmanned surface

vehicles[C]// IEEE International Conference on Robotics and Automation. Paris：IEEE，2020.

[97] FEFILATYEV S，SMARODZINAVA V，HALL L O，et al. Horizon detection using machine learning techniques[C]// International Conference on Machine Learning and Applications. Orlando：IEEE，2006.

[98] BLOISI D D，IOCCHI L，PENNISI A，et al. ARGOS-Venice boat classification[C]// International Conference on Advanced Video and Signal Based Surveillance. Karlsruhe：IEEE，2015.

[99] MARQUES M M，DIAS P，SANTOS N P，et al. Unmanned aircraft systems in maritime operations：challenges addressed in the scope of the Seagull project[C]// OCEANS. Genova：IEEE，2015.

[100]RIBEIRO R，CRUZ G，MATOS J，et al. A data set for airborne maritime surveillance environments[J]. IEEE Transactions on Circuits and Systems for Video Technology，2017，29（9）：2720-2732.

[101]PATINO L，NAWAZ T，CANE T，et al. PETS 2017：dataset and challenge[C]// IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu：IEEE，2017.

[102]KRISTAN M，KENK V S，KOVACI S，et al. Fast image-based obstacle detection from unmanned surface vehicles[J]. IEEE Transactions on Cybernetics，2016，46（3）：641–654.

[103]BOVCON B，PERS J，KRISTAN M，et al. Stereo obstacle detection for unmanned surface vehicles by IMU-assisted semantic segmentation[J]. Robotics and Autonomous Systems，2018，104：1-13.

[104]PRASAD D K，RAJAN D，RACHMAWATI L，et al. Video processing from electro-optical sensors for object detection and tracking in a maritime environment：a survey[J]. IEEE Transactions on Intelligent Transportation Systems，2017，18（8）：1993-2016.

[105]MOOSBAUER S，KONIG D，JAKEL J，et al. A benchmark for deep learning based object detection in maritime environments[C]// IEEE Conference on Computer Vision and Pattern Recognition Workshops. Long Beach：IEEE，2019.

[106]GUNDOGDU E，SOLMAZ B，YUCESOY V，et al. Marvel：a large-scale image dataset for maritime vessels[C]// Asian Conference on Computer Vision. Taipei：Springer，2016.

[107]SOLOVIEV V，FARAHNAKIAN F，ZELIOLI L，et al. Comparing CNN-based object detectors on two novel maritime datasets[C]// IEEE International Conference on Multimedia Expo Workshops. London：IEEE，2020.

[108]STECCANELLA L，BLOISI D D，CASTELLINI A，et al. Waterline and obstacle detection in images from low-cost autonomous boats for environmental monitoring[J]. Robotics and Autonomous Systems，2020，124：103346.

[109]BOVCON B，MUHOVIC J，PERS J，et al. The MaSTr1325 dataset for training deep USV obstacle detection models[C]// IEEE/RSJ International Conference on Intelligent Robots and Systems. Macau：IEEE，2019.

[110]RAZA M, PROKOPOVA H, HUSEYNZADE S, et al. SimuShips–A High Resolution Simulation Dataset for Ship Detection with Precise Annotations[EB/OL]. [2022-09-22]. https://arxiv.org/abs/2211.05237.

[111]BOVCON B，MUHOVI J，VRANAC D，et al. MODS–a USV-oriented object detection and obstacle segmentation benchmark[J]. IEEE Transactions on Intelligent Transportation Systems，2022，23（8）：13403-13418.

[112]EVERINGHAM M，VAN GOOL L，WILLIAMS C K I，et al. The pascal visual object classes（VOC）challenge[J]. International Journal of Computer Vision，2010，88（2）：303-338.

[113]LIN T Y，MAIRE M，BELONGIE S，et al. Microsoft COCO：common objects in context[C]// European Conference on Computer Vision. Zurich：Springer，2014.

[114]GONG S X，LIU X M，JAIN A K. Mitigating face recognition bias via group adaptive classifier[C]// IEEE Conference on Computer Vision and Pattern Recognition. Virtual Event：IEEE，2021.

[115]CHU X G，ZHENG A L，ZHANG X Y，et al. Detection in crowded scenes：one proposal，multiple predictions[C]// IEEE Conference on Computer Vision and Pattern Recognition. Seattle：IEEE，2020.

[116]XIE S，GIRSHICK R，DOLLÁR P，et al. Aggregated residual transformations for deep neural networks[C]// IEEE Conference on Computer Vision and Pattern Recognition. Honolulu，IEEE，2017.

[117]GAO S H，CHENG M M，ZHAO K，et al. Res2Net：a new multi-scale backbone architecture[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2021，43（2）：652-662.

[118]ZHU X Z，HU H，LIN S，et al. Deformable ConvNets V2：more deformable，better results[C]// IEEE Conference on Computer Vision and Pattern Recognition Workshops. Long Beach：IEEE，2019.

[119]GUO C X，FAN B，ZHANG Q，et al. AugFPN：improving multi-scale feature learning for object detection[C]// IEEE Conference on Computer Vision and Pattern Recognition. Seattle：IEEE，2020.

[120]ZONG Z F，CAO Q G，LENG B. RCNet：reverse feature pyramid and cross-scale shift network for object detection[C]// ACM Conference on Multimedia. Virtual Event：ACM，2021.

[121]CHEN Z H，YANG C H，LI Q F，et al. Disentangle your dense object detector[C]// ACM Conference on Multimedia. Virtual Event：ACM，2021.

[122]FENG C J，ZHONG Y J，GAO Y，et al. TOOD：task-aligned one-stage object detection[C]// IEEE/CVF International Conference on Computer Vision. Virtual Event：IEEE，2021.

[123]CAO Y H，CHEN K，LOY C G，et al. Prime sample attention in object detection[C]// IEEE Conference on Computer Vision and Pattern Recognition. Seattle：IEEE，2020.

[124]KIM K，LEE H S. Probabilistic anchor assignment with IoU prediction for object detection[C]// European Conference on Computer Vision. Glasgow：Springer，2020.

[125]陈卫，乔延利，孙晓兵，等. 基于偏振辐射图融合的水面太阳耀光抑制方法[J]. 光学学报，2019，39（5）：382-390.

[126]QIAN R，TAN R Y，YANG W H，et al. Attentive generative adversarial network for raindrop removal from a single image[C]// IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City：IEEE，2018.

[127]KENDALL A，GAL Y，CIPOLLA R. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics[C]// IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City：IEEE，2018.

[128]WANG W Y，TRAN D，FEISZLI M. What makes training multi-modal classification networks hard? [C]// IEEE Conference on Computer Vision and Pattern Recognition. Seattle：IEEE，2020.

[129]MA J Y，YU W，LIANG P W，et al. FusionGAN：a generative adversarial network for infrared and visible image fusion[J]. Information Fusion，2019，48：11-26.

[130]READING C，HARAKEH A，CHAE J，et al. Categorical depth distribution network for monocular 3D object detection[C]// IEEE Conference on Computer Vision and Pattern Recognition. Virtual Event：IEEE，2021.

[131]ZHANG Y A, CHEN J X, and HUANG D. CAT-Det: contrastively augmented transformer for multi-modal 3D object detection[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans: IEEE, 2022.

[132]LIU H, NIE J, LIU Y J, et al. A multi-modality sensor system for unmanned surface vehicle[J]. Neural Processing Letters, 2020, 52(2): 977-992.

[133]KHANDELWAL S, GOYAL R, SIGAL L. UniT: unified knowledge transfer for any-shot object Detection and segmentation[C]// IEEE Conference on Computer Vision and Pattern Recognition. Virtual Event: IEEE, 2021.

[134]SHAO F F, CHEN L, SHAO J, et al. Deep learning for weakly-supervised object detection and localization: a survey[J]. Neurocomputing, 2022, 496: 192-207.

（责任编辑：肖楚楚）