

DOI:10.11918/202201069

基于 SimAM 注意力机制的近岸船舶检测方法

郭玲,于海雁,周志权

(哈尔滨工业大学(威海)信息科学与工程学院,山东威海264209)

摘要: 基于视觉图像的船舶目标检测中由于图像背景复杂,无干扰较多,导致船舶目标检测的难度增大。并且多类别船舶检测数据集现有数量较少且存在样本不均衡的问题使得船舶目标检测性能较低。针对复杂背景干扰检测,本文通过引入 SimAM 注意力机制对 YOLOv3 模型进行改进,利用该机制加强船舶目标在提取特征中的权重并抑制背景干扰权重,从而提升模型检测性能;同时,采用强实时数据增强以改善样本尺度分布不均衡的问题,结合迁移学习提升在样本数量受限情况下的船舶检测精度。提取特征的可视化结果显示改进模型对无关背景特征干扰进行了抑制,增强了模型对于船舶特征的提取能力。在 SeaShips 数据集上,提出的改进模型在不引入额外可学习参数的情况下 mAP. 5、mAP. 75 分别达到了 96.93%、71.49%,检测速度达到了 66 frame/s,在检测精度与运行效率方面保持了均衡。与 Saliency-aware CNN、eYOLOv3 相比更有效地优化了目标特征,使得 mAP. 5 分别提高了 9.53%、9.19%。改进模型在新加坡海事数据集上在船舶类型目标检测的 mAP. 5 达到了 81.81%,验证了模型具有较好的泛化能力。

关键词: 船舶检测;注意力机制;数据增强;改进 YOLOv3;迁移学习

中图分类号: TP391.4 文献标志码: A 文章编号: 0367-6234(2023)05-0014-08

Nearshore ship detection method based on SimAM attention mechanism

GUO Ling, YU Haiyan, ZHOU Zhiqian

(School of Information Science and Engineering, Harbin Institute of Technology (Weihai), Weihai 264209, Shandong, China)

Abstract: Due to the complex background of ship targets and much irrelevant interference in visual images, it is difficult to conduct ship detection. In addition, there are few datasets for multi-category ship detection and the samples are often unbalanced, which makes the ship target detection performance degraded. Considering the ship detection background interference, an improved YOLOv3 model was proposed by introducing SimAM attention mechanism, which was used to enhance the weight of the ship target in the extracted features and suppress the weight of background interference, thus improving the model detection performance. Meanwhile, strong real-time data augmentation was applied to improve the unbalanced distribution of sample scales, and transfer learning was combined to improve the ship detection accuracy in the condition of a restricted number of samples. The visualization results of extracted features show that the improved model could suppress irrelevant background features, and the abilities of feature extraction and target localization were enhanced. Without introducing additional learnable parameters, the proposed model achieved 96.93% and 71.49% for mAP. 5 and mAP. 75 on the SeaShips dataset, and detection speed reached 66 frames per second, indicating a good balance between detection accuracy and efficiency. The improved model optimized the target features more effectively compared with the Saliency-aware CNN and eYOLOv3 models, resulting in an improvement of mAP. 5 by 9.53% and 9.19%. The mAP. 5 for ship type target detection on Singapore Maritime Dataset reached 81.81%, indicating that the proposed model has good generalization performance.

Keywords: ship detection; attention mechanism; data augmentation; improved YOLOv3; transfer learning

船舶检测在港口管理、海洋监测、航行安全等诸多领域都具有重要的应用价值,对提高海上交通管理效率、保障海域安全有着重要的作用。在过去的几年里,利用图像信息进行船舶检测的技术在不断

地发展,根据图像生成源,基于图像的船舶目标检测方法大致分为三类:遥感图像、雷达图像和视觉图像船舶检测。从卫星遥感图像与雷达图像中自动提取目标进行船舶的检测虽然取得了一些成果^[1-2],但

收稿日期: 2022-01-17;录用日期: 2022-04-20;网络首发日期: 2022-08-29

网络首发地址: <https://kns.cnki.net/kcms/detail/23.1235.T.20220826.1508.010.html>

基金项目: 山东省重点研发计划重大科技创新工程(2020JZZY010705)

作者简介: 郭玲(1997—),男,硕士研究生;周志权(1973—),男,教授,博士生导师

通信作者: 于海雁, haiyanu@hit.edu.cn

由于处理流程较长,难以进行实时检测,实现成本高,在具体应用中存在很多不足。视觉图像中船舶目标分辨率高,细节、颜色特征丰富,获取成本低,对于许多特殊的应用,如安全导航、无人船等需要准确和实时的船舶检测(船舶定位和分类),视觉图像更适合实时船舶目标检测,因此视觉图像中的船舶检测受到越来越多的研究关注^[3]。Zhang 等^[4]通过离散余弦变换区域特征进行水平线检测、背景建模、背景减影实现了大波浪情况下视觉图像船舶检测,但受限非学习的特征提取方式,背景存在相似物体时检测效果较差。Kim 等^[5]结合贝叶斯方法与 Faster R-CNN 网络通过深度学习方法在自建数据集的船舶检测任务上达到了 93.92% 的较高平均精度。Lee 等^[6]采用增加 passthrough 改进 YOLOv2 在新加坡海事数据集^[7]上实现了对快艇、帆船等 10 类目标的实时检测。Shao 等^[8]发布了名为 SeaShips 的视觉图像船舶目标检测公共数据集,涵盖了 6 种常见船舶类型(矿石船、散货船、杂货船、集装箱船、渔船和客船)31 455 张,公开部分 7 000 张,为视觉图像船舶检测提供了良好的数据基础。基于 SeaShips 数据集,Shao 等^[3]提出了 Saliency-aware CNN,利用海岸线分割的方法减少背景干扰、缩小检测区域,结合显著性图增强船舶目标定位准确性以提高检测平均精度但由于引入显著图计算检测效率有所下降。Liu 等^[9]在 YOLOv3 模型基础上改善了损失函数,加入了不确定性边框回归来提升船舶目标定位能力,提出的 eYOLOv3 在小目标检测上有所改善,使得在 SeaShips 数据集上检测平均准确率提高到 87.74%,但总体提升有限。通过海岸线分割或进行后处理的方式间接减少背景干扰,提升效果有限,为了直接有效地优化船舶特征、抑制背景干扰特征,本文在简洁高效的 YOLOv3 模型基础上引入了 SimAM 无参 3D 注意力机制对相关特征进行优化以提升检测性能,并在 SMD 数据集上进行了泛化性能验证。视觉图像船舶检测不同于常见的人脸检测、车辆检测等问题,船舶图像背景复杂,存在多种干扰信息,包括光照因素、云雾天气、水面波动、船舶目标小,使得提取船舶目标变得困难,增加了处理时间甚至造成了大量的漏检误检等现象^[10]。此外,数据集稀缺,样本不均衡,使得训练高精度的模型难度较大,现有文献较少考虑此方面问题,本文通过采用强实时数据增强结合迁移学习的方法进行了改善。在目标检测领域,已经涌现出多种基于卷积神经网络的目标检测算法^[11-14],在速度和精度上大大超过传统的目标检测算法。基于深度学习方法进行

船舶目标检测已经成为当前主流的研究方向^[3,9-10],在数据量少、样本不均衡情况下实现从复杂真实场景视觉图像中通过端到端神经网络准确、高效地检测船舶目标面临着巨大的挑战也具有重要意义。本文的主要贡献包括 4 个方面:

1) 使用强实时数据增强改善了数据不均衡的问题,通过迁移学习的方式使得在小样本情况下仍能获得较好的船舶检测性能。

2) 针对船舶目标检测易受背景干扰的问题在 YOLOv3 的基础上引入 SimAM 注意力机制对模型进行改进,提高了船舶目标检测的精度,同时保持了高效的检测效率,并通过实验对比验证了改进模型的有效性。

3) 利用特征图可视化的方法对改进模型精度提升做出了阐释。

4) 将改进模型应用于不同数据集与基线模型对比的实验结果表明提出的改进模型具有较好泛化能力。

1 引入注意力机制的改进 YOLOv3 船舶检测模型

主流的卷积神经网络目标检测模型分为一阶段类型与两阶段类型两大类。YOLOv3 是一种一阶段的检测模型,它将整个检测流程直接作为一个回归及分类问题同时进行处理,与 Faster RCNN 为代表的两阶段的目标检测模型相比,无需进行候选区域提取再进行分类和回归,因此检测速率远高于两阶段的目标检测模型,在实际生产应用中 YOLOv3 凭借较高的效率获得了广泛的应用。YOLOv3 模型结构简洁未集成大量技巧且结构设计成熟适合作为模型改进的基线。较新发表的 YOLOv4^[15]在 YOLOv3 的基础上集成了大量训练技巧在不同情景的特定任务上可能存在冲突,造成分析上的困难,并且与 YOLOv3 相比并未有理论上的突破,因此大多数如 YOLOX^[16]、YOLOR^[17]等基于 YOLO 框架改进的模型均在 YOLOv3 的基础上进行优化。为了面向船舶检测任务更好地分析与优化模型并使得模型能够兼顾检测精度与检测效率,改进模型以 YOLOv3 作为基线。

YOLOv3 模型由 Darknet-53 主干网络、特征金字塔网络与检测头三部分组成。Darknet-53 引入了残差连接使得网络深度堆叠更深以获得更强的特征表达能力,原始输入图像通过 Darknet-53 主干网络进行特征的提取,为了提高对不同尺寸目标的预测精度,YOLOv3 主干网络中将降采样率为 8、16、32

三个不同分辨率下提取到的特征组成特征金字塔结构并进行高低层特征融合,融合后的特征送入检测头中进行分类与回归。检测训练的损失函数如式(1)~(5)所示由边界框损失、分类损失、目标损失加权组成。其中 s 为格点尺寸; B 为 box; $1_{i,j}^{obj}$ 为若在 i, j 处的 box 有目标,其值为 1, 否则为 0; $1_{i,j}^{noobj}$ 为若在 i, j 处的 box 无目标,其值为 1, 否则为 0; c_f 为存在目标的置信度。

$$l_{\text{box}} = \lambda_{\text{coord}} \sum_{i=0}^{s^2} \sum_{j=0}^B 1_{i,j}^{obj} (2 - w_i * h_i) [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + (w_i - \hat{w}_i)^2 + (h_i - \hat{h}_i)^2] \quad (1)$$

$$l_{\text{cls}} = \lambda_{\text{class}} \sum_{i=0}^{s^2} \sum_{j=0}^B 1_{i,j}^{obj} \sum_{c \in \text{classes}} \text{BCE}_i(c) \quad (2)$$

$$l_{\text{obj}} = \lambda_{\text{noobj}} \sum_{i=0}^{s^2} \sum_{j=0}^B 1_{i,j}^{noobj} \text{BCE}_i(c_f) + \lambda_{\text{obj}} \sum_{i=0}^{s^2} \sum_{j=0}^B 1_{i,j}^{obj} \text{BCE}_i(c_f) \quad (3)$$

$$\text{BCE}_i(x) = -[p_i(x) \log(\hat{p}_i(x)) + (1 - p_i(x)) \log(1 - \hat{p}_i(x))] \quad (4)$$

$$\text{loss} = l_{\text{box}} + l_{\text{cls}} + l_{\text{obj}} \quad (5)$$

在船舶检测任务中船舶目标所处背景相对于通用目标检测而言通常更加复杂多变。如图 1 所示背景中水面区域存在波动、天空中不规则的云和陆地区域的建筑等无关特征,对船舶目标的检测造成了很大干扰。与通用目标检测场景中目标常处于图像的中心位置不同,航行中的船舶目标存在较大的位置变化和朝向变化,这对于模型的定位与检测能力提出了更高要求。



图 1 船舶检测背景干扰

Fig. 1 Ship detection background interference

为了降低复杂背景的干扰,传统的检测方法常采用霍夫变换等方法提取海岸线的方式将水面区域划分出来进行进一步检测^[3],该方法处理效率较低且在海岸线变化大时分割效果较差。受通用目标检测中为了提升模型特征表示能力在主干网络中引入非 3D 的带参注意力机制的 Attention-YOLO^[18] 的启

发,针对船舶目标检测情景中干扰影响的问题,为了提高模型的干扰抑制能力,本文提出的改进模型在特征金字塔网络中引入了 3D 的无参 SimAM 注意力机制^[19],相比于 3D 带参注意力机制在不引入额外可学习参数情况下能够更全面高效地评估特征权重,以增强船舶特征,减弱背景干扰特征,提高了模型的干扰抑制能力并通过直接有效的方式对船舶目标特征优化,减小了对变化较大的海岸线等先验信息提取的依赖。SimAM 建立在视觉神经科学理论的基础之上,具有更多信息的神经元,与其相邻神经元相比表现更加显著,并会对相邻神经元产生空间抑制现象,在处理视觉相关任务时这些神经元带有更多关键信息应该赋予更高的权重,在船舶检测任务中这些网络的神经元往往负责提取出船舶目标的关键特征需要进行增强。如图 2 所示本文在 YOLOv3 模型的特征金字塔结构中引入了 SimAM 注意力机制对主干网络提取出的特征进行优化,通过端到端的方式减少背景噪声对检测的干扰,相比海岸线划分对水域形状变化更有鲁棒性。如式(6)~(8)所示 SimAM 通过定义线性可分性的能量函数对网络中的每个神经元进行评估,其中 t 为目标神经元, x 为相邻神经元, λ 为超参数, e_i^* 能量越低表明神经元与相邻的区分度越高,神经元的重要程度也越高,如式(9)所示通过 $1/e_i^*$ 对神经元根据重要性进行加权。与 SENet^[20]、CBAM^[21] 关注于通过池化、全连接层手工设计注意力模块不同,依据神经科学理论中的能量函数评估各处特征的重要性提出的 SimAM 更具有可解释性,无需引入可学习参数。且 SimAM 是一种 3D 注意力机制与 SENet 关注于通道维度重要性、CBAM 依次关注于通道、空间维度重要性相比直接评估了各个独立神经元的重要性,在保持运算高效性的同时能够更全面地评估各神经元的重要性进行加权。

$$e_i^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda} \quad (6)$$

$$\hat{\mu} = \frac{1}{M} \sum_{i=1}^M x_i \quad (7)$$

$$\hat{\sigma}^2 = \frac{1}{M} \sum_{i=1}^M (x_i - \hat{\mu})^2 \quad (8)$$

$$\tilde{X} = \text{sigmoid}\left(\frac{1}{E}\right) \odot X \quad (9)$$

在提出的改进船舶目标检测模型中 SimAM 模块通过能量函数的闭式解高效地评估 Darknet-53 提取特征的重要性,对具有更多有效信息的神经元输出进行增强,并有效地抑制其他无关特征的干扰,使得网络具有更强的特征表示能力,在降低背景干扰的同时也提高了模型的目标定位能力。

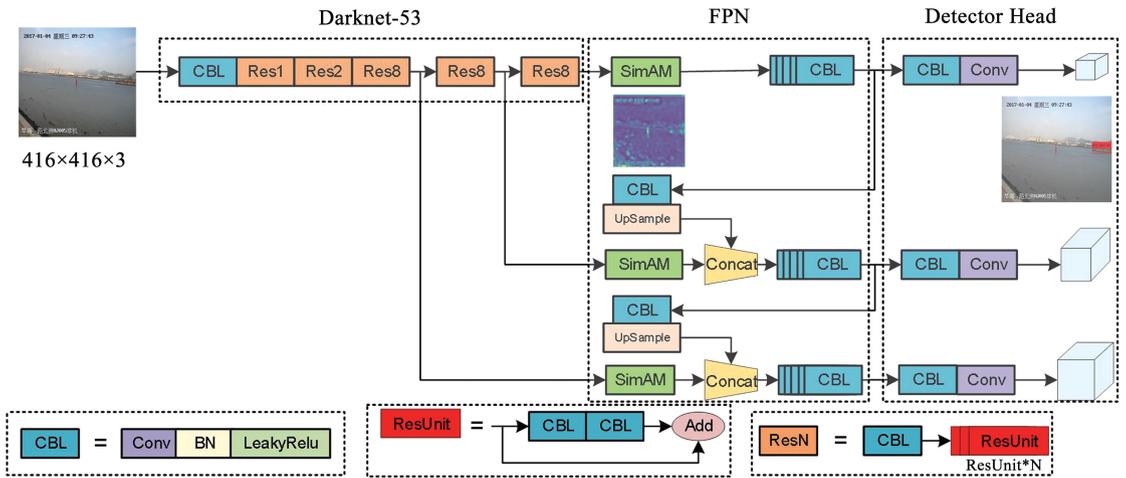


图 2 引入 SimAM 注意力的改进网络结构

Fig. 2 Improved network structure based on SimAM attention

2 数据增强与迁移学习

实验使用的数据集为 SeaShips7000, 该数据集是由真实场景监控视频中选取的图像组成的视觉图像船舶检测的公开数据集, 它涵盖了 6 种常见的船舶类型, 充分考虑了目标尺寸变化、背景变化、视角变化、光照变化等重要的变差因素, 能够反映实际场景中的船舶目标分布情况。SeaShips 数据集公开部分为 7 000 张, 其中训练集 1 750 张、验证集 1 750 张、测试集 3 500 张, 如图 3 所示各集合中船舶目标样本分布接近, 可使得训练出的模型具有较好的泛化性。

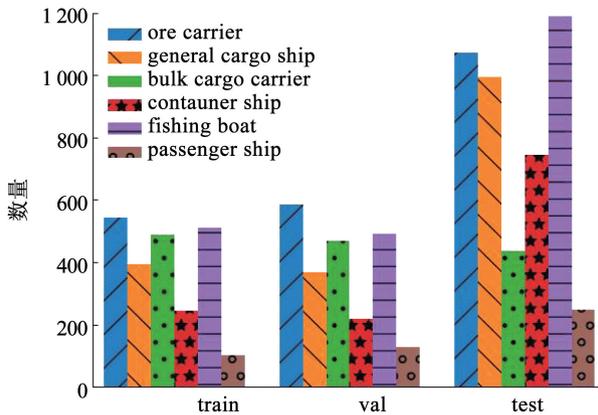


图 3 SeaShips7000 数据分布

Fig. 3 Data distribution of SeaShips7000

训练所使用的训练集仅包含 1 750 张训练数据, 对于具有几十兆参数的大型目标检测模型而言少量的数据难以使得模型训练收敛, 会使得对学习率、mini batch size 的设置要求很严格, 因此为了使得模型能够得到更好训练, 采用合适的数据增强方法对数据的进行扩充。如图 4 所示按照 COCO 数据

集中尺寸分类标准对训练集船舶目标进行分布分析, 训练集中不仅存在样本数量不足的问题, 样本的尺度分布也存在不均衡的问题, 小尺寸目标占比较小, 中尺寸与大尺寸目标数量占主导地位, 会使得模型忽略小尺寸船舶目标的检测产生漏检导致性能下降。

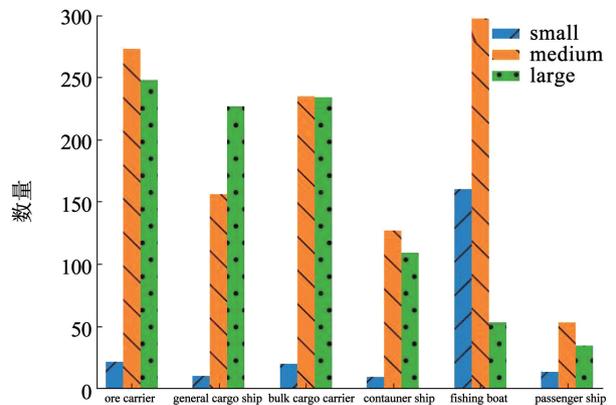


图 4 SeaShips7000 训练集目标尺寸分布

Fig. 4 Object scale distribution of SeaShips7000 training set

为了使训练数据数量得到扩充并改善尺寸分布不均衡的问题, 提出的模型采用强实时数据增强方法, 在训练时每个 epoch 都进行随机数据增强。对输入的图像进行随机纵横比变换、随机尺寸变换相结合的方式改善目标尺寸分布不均衡的问题, 进行随机平移使得目标位置分布多样化以提高模型的定位能力, 并通过随机色域变换与随机翻转相结合的方式提高模型的鲁棒性与泛化能力。如图 5 所示为增强数据可视化结果, 如图 6 所示增强后的训练集数据在尺寸分布上得到了改善, 各类别船舶目标中小尺寸目标的比例得到了提升。



(a) 尺度变换 (b) 色域变换 (c) 水平翻转

图 5 增强数据可视化

Fig. 5 Visualization of augmented data

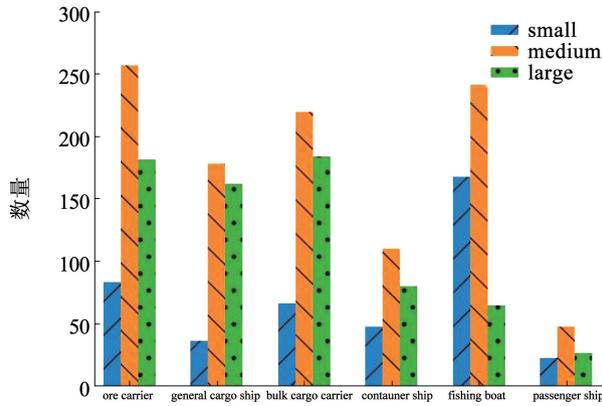
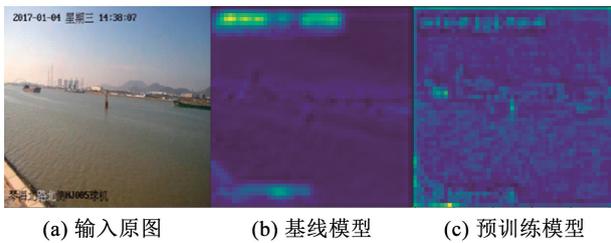


图 6 增强数据目标尺寸分布

Fig. 6 Object scale distribution of augmented data

在数据量较少的情况下,如图 7(b)所示直接从头开始训练的基线模型无法很好地训练出模型的特征提取能力,对于船舶目标的特征响应较弱,提取细节的能力不足。在 COCO 2017 数据集上先对模型预训练使得模型具备较好的特征提取能力,将获得的权重在 SeaShips7000 数据集上通过迁移学习的方式先保持 Darknet-53 主干网络的参数不变,训练 30 个 epoch 后再降低学习率与 mini batch size 进行参数微调以提高在船舶检测任务上的性能。如图 7(c)所示通过迁移学习的方式使得模型的特征提取能力得到了增强,对于船舶目标特征提取效果更加突出。



(a) 输入原图 (b) 基线模型 (c) 预训练模型

图 7 模型特征可视化对比

Fig. 7 Comparison of model feature visualization

3 实验与分析

实验中实验设置与性能评价指标进行介绍,通过应用不同注意力机制改进结合实时数据增强效果

与提出的 SimAM 改进模型进行横向对比,与在 SeaShips 数据集上其他改进模型性能进行纵向对比,最后在新加坡海事数据集上进行了模型的泛化性能验证。

3.1 实验设置

实验在 Ubuntu 20.04 平台上进行,CPU 为 Intel Core i7-10700F@2.90 GHz,使用的 GPU 为 Nvidia GeForce RTX 3070,采用 CUDA11.2 加速库进行并行加速,基于 python3.7 的 PyTorch 深度学习框架实现。为了保证对比实验的公平性,各模型输入尺寸均为 416×416 并在 SeaShips7000 上训练 100 个 epoch,进行相同的数据增强,使用 Adam 优化器进行参数的优化。作为基线的 YOLOv3 原始模型学习率为 0.001,学习率衰减为 0.92,mini batch size 设置为 8,其他模型都使用了预训练权重进行迁移学习,前 30 个 epoch 保持了主干网络参数不变,学习率为 0.001,学习率衰减为 0.92,mini batch size 设置为 8,后 70 个 epoch 学习率为 0.000 1,学习率衰减为 0.92,mini batch size 设置为 4。

3.2 评价指标

I_{ou} (intersection over union) 为预测框和真实框的交集和并集之比如式 (10) 所示,用于评价预测框的精准程度以判定预测是否正确。 I_{ou} 越大说明预测框的与真实框重合程度越高,预测框质量越高。

$$I_{ou} = \frac{B \cap G}{B \cup G} \quad (10)$$

对于目标检测中的分类问题,可以根据目标的真实类别和预测类别的组合,将样本划分真正例 (true positive)、真负例 (true negative)、假正例 (false positive)、假负例 (false negative) 4 种情况,令 T_p 、 T_n 、 F_p 、 F_n 分别为其对应的样本数量,则在给定 I_{ou} 阈值下,可以根据式 (11) 和式 (12) 计算样本预测的查准率 P 和查全率 R 。 P 为被正确检测的样本数量占所有被检测到的样本数量的比例, R 为被正确检测的样本数量占应该被检测到的样本数量的比例。查准率和查全率相互制约,为了综合评价算法有效性,采用平均精度 A_p (average precision) 进行评价,以 R 为横坐标, P 为纵坐标,可以得到一条 $P-R$ 曲线,如式 (13) 所示 A_p 为该曲线所包围的面积。

$$P = \frac{T_p}{T_p + F_p} \quad (11)$$

$$R = \frac{T_p}{T_p + F_n} \quad (12)$$

$$A_p = \int_0^1 P dR \quad (13)$$

多类别检测时,如式 (14) 所示每一个类别 i 都

对应一个 A_p 值,一共有 n 种类别。 m_{AP} (mean average precision) 为所有类别 A_p 值的平均值, I_{ou} 阈值为 0.5 时得到 mAP.5, I_{ou} 阈值为 0.75 时对于定位要求更严格,可得到 mAP.75 用于反映模型的定位能力。

$$m_{AP} = \frac{\sum_{i=1}^n A_{p,i}}{n} \quad (14)$$

除了评价精度,为了评价模型的检测效率,还采用每秒帧数 (frame/s) 评估模型的速度, frame/s 表示检测器一秒钟能够检测的帧数,30 frame/s 为进行实时检测的最低标准。

3.3 实验结果与分析

如表 1 所示通过强实时数据增强的方式在 SeaShips7000 上对 YOLOv3 进行训练作为基线,基线模型在训练后在测试集上的 mAP.5 达到了 91.69%,在矿砂船、散货船、杂货船、集装箱船这些

大型船舶目标上检测 AP.5 达到了 93% 以上,但由于数据量较少训练出的模型细节特征提取能力不足导致在渔船、客船这类小型船舶目标上检测性能较差,模型的定位能力较弱, mAP.75 仅有 49.64%。在 COCO 2017 上预训练后模型的特征提取能力与定位能力都得到了加强,在小型船舶目标的检测性能提升到 93% 以上, mAP.5 由 91.69% 提升到 96.47% 得到了较大提升,并且 mAP.75 提升到 69.89% 表明了目标定位的性能得到很大的改善。提出的改进模型通过 SimAM 注意力机制调整目标与干扰的权重信息,抑制了背景特征的干扰,增强了船舶目标特征的强度,有效改善了原始模型在小目标检测性能,与预训练模型相比易受背景干扰影响的小型目标渔船、客船的 AP.5 分别提高了 0.37% 与 1.58%,模型的目标定位能力也得到了提升,与预训练模型相比 mAP.75 提高了 1.6%。

表 1 改进模型性能分析

Tab.1 Performance analysis of improved models

%

模型	mAP.5	mAP.75	AP.5					
			ore carrier	bulk cargo carrier	general cargo ship	container ship	fishing boat	passenger ship
baseline	91.69	49.64	94.77	93.05	93.15	97.07	89.34	82.76
pretrain	96.47	69.89	97.02	97.45	97.44	98.85	94.92	93.13
pretrain + SimAM	96.93	71.49	96.49	98.68	97.39	99.02	95.29	94.71

为了进一步说明改进模型的有效性,通过主干网络提取出的特征图可视化对提出改进模型性能提升原因做出了阐述。如图 8(c) 所示预训练模型特征相比于图 8(b) 中基线模型特征经过预训练后模型提取的特征更加丰富,对于细节特征的提取能力得到了很大增强,使得模型的 mAP.5 与 mAP.75 均得到了较大提升。图 8(d) 中提出的 SimAM 改进模型特征与预训练模型特征相比,在不同的场景下船舶目标处特征响应强度显著高于预训练模型特征,这主要因为在注意力机制作用下船舶特征的权重被加强,使得改进模型更加关注于检测的船舶目标特征。在增强了船舶目标特征提取能力的同时,对于水面波动与天空中云影等无关特征在改进模型中也得到了有效的抑制。通过引入 SimAM 注意力机制使得船舶检测任务中背景干扰特征强度得到抑制、船舶目标特征得到增强,从而实现了改进模型的检测性能提升。

如表 2 所示对于不同注意力机制进行了对比实验,引入 SimAM 注意力机制的改进模型相比于其他类型的注意力机制^[18]能够直接评估神经元的 3D 权重,因此对于模型定位能力的提升最高 mAP.75 达

到了 71.49%,在不引入额外学习参数的情况下对检测精度进行了有效提升。并且 SimAM 具有计算量小的优势,在提升检测精度的同时能保持较高的运行效率,达到了 66.16 frame/s。综合来看 SimAM 改进模型在模型参量、检测精度提升与运行效率方面得到了较好的均衡。

在 SeaShips 数据集上,如表 3 所示提出的模型与 Saliency-aware CNN 通过海岸线提取减少背景干扰区域的方式相比,更少受到水域形状变化影响,且与额外计算显著图改善定位相比计算代价更小,因此使用 SimAM 注意力机制抑制背景干扰、提升模型特征提取能力更加有效,在提升检测精度的同时保持了更高的检测效率。Liu 等^[9]提出的不确定性边框回归与改进损失函数相结合的 eYOLOv3 模型与 Saliency-aware CNN 相比在渔船小目标上提升效果显著,但该方法主要通过间接的后处理方式改善检测效果,与提出的改进模型可直接抑制无关干扰、优化目标特征相比,未能直接改善模型特征提取能力,导致总体提升效果一般。与检测精度较高的两阶段 Faster R-CNN 相比,改进模型采用了更高效的一阶段 YOLOv3 为基线,因此维持了较高的检测效率。在 SeaShips 数据集上 mAP.5 检测性能对比表

明了提出的引入 SimAM 注意力机制的改进模型通过预训练迁移学习与强实时数据增强结合能够有效地提高网络特征提取与目标定位能力,与 Faster R-

CNN、Saliency-aware CNN、eYOLOv3 相比在各类大尺寸与小尺寸船舶目标的检测精度与效率上均得到了提高。

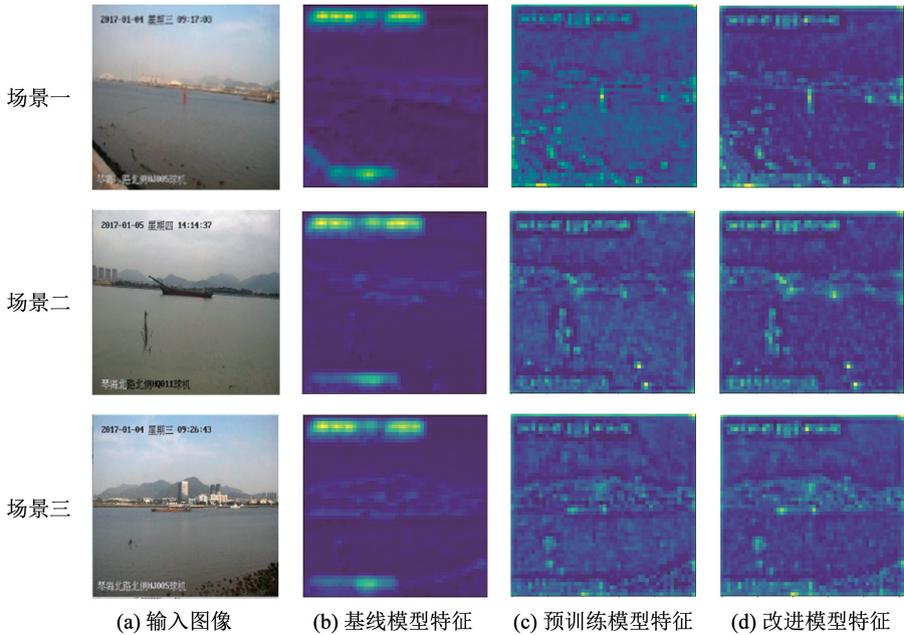


图 8 不同场景下特征可视化对比

Fig. 8 Comparison of feature visualization in different scenes

表 2 不同注意力机制改进效果横向对比

Tab. 2 Comparison of improvement effect of different attention mechanisms

方法	mAP. 5/%	mAP. 75/%	参数量/ 10^6	效率/($\text{frame} \cdot \text{s}^{-1}$)
baseline	91.69	49.64	61.55	66.44
pretrained	96.47	69.89	61.55	66.15
SimAM	96.93	71.49	61.55	66.16
CBAM	96.20	71.01	61.72	62.43
channel	96.99	70.95	61.72	65.80
spatial	96.93	70.30	61.55	65.29

表 3 改进模型性能纵向对比

Tab. 3 Performance comparison of improved models

模型	mAP. 5/%	效率/ ($\text{frame} \cdot \text{s}^{-1}$)	AP. 5/%					
			ore carrier	bulk cargo carrier	general cargo ship	container ship	fishing boat	passenger ship
Faster R-CNN	90.10	6	89.40	90.30	90.70	90.90	88.80	90.60
Saliency-aware CNN	87.40	49	88.10	87.60	91.70	90.30	78.30	88.60
eYOLOv3-608	87.74	30	90.61	85.21	87.95	91.07	89.51	82.09
YOLOv3 + SimAM	96.93	66	96.49	98.68	97.39	99.02	95.29	94.71

3.4 模型泛化性测试

为了验证本文提出 SimAM 改进模型的泛化能力,在新加坡海事数据集上进行了模型泛化性的验证,使用的数据由岸上采集视频中抽取的图像组成,其中包含 6 种类别的船舶,使用的图像数据中训练集 3 926 张、验证集 1 962 张和测试集 2 508 张。将在 SeaShips 数据集上预训练的模型在新加坡海事数

据集上进行迁移学习,如表 4 所示提出的改进模型在 Ferry、Sailboat、Vessel/ship 类别船舶目标检测上达到了 88% 以上,与基线模型 YOLOv3 相比在船舶目标类别检测上 mAP. 5 由 70.23% 提升到 81.81% 有了较大改善,验证了提出的改进模型在船舶检测任务上具有较好的泛化能力。

表4 新加坡海事数据集上模型泛化性能分析

Tab. 4 Generalization performance of model on Singapore Maritime Dataset

类别	AP. 5	
	YOLOv3 + SimAM	YOLOv3
Boat	64.23	49.33
Ferry	88.69	76.76
Kayak	70.28	67.84
Sailboat	100.00	76.97
Speedboat	71.35	56.76
Vessel/ship	96.28	93.71
mAP. 5	81.81	70.23

4 结 论

针对船舶目标检测任务中复杂背景特征对船舶目标检测造成干扰的问题,提出了引入 SimAM 注意力机制的 YOLOv3 改进模型,通过注意力机制对背景特征的干扰进行了抑制,增强了模型对船舶目标特征提取能力,提高了模型定位准确性。在样本数量受限情况下,结合迁移学习与强实时数据增强的方法,改善了船舶目标尺寸分布不均、小目标检测性能差的问题。提出的改进模型在不引入额外可学习参数的情况下,在 SeaShips 数据集上与 Saliency-aware CNN、eYOLOv3 相比提高检测精度的同时保持了较高检测效率。与基线模型在 SMD 上的检测性能对比,验证了提出的改进模型在船舶检测任务上具有较好的泛化能力。

参考文献

[1] LI Linhao, ZHOU Zhiqiang, WANG Bo, et al. A novel CNN-based method for accurate ship detection in HR optical remote sensing images via rotated bounding box [J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59 (1): 686. DOI: 10.1109/TGRS.2020.2995477

[2] LIU Tao, YANG Ziyuan, MARINO A, et al. PolSAR ship detection based on neighborhood polarimetric covariance matrix [J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59 (6): 4874. DOI: 10.1109/TGRS.2020.3022181

[3] SHAO Zhenfeng, WANG Linggang, WANG Zhongyuan, et al. Saliency-aware convolution neural network for ship detection in surveillance video [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 30(3): 781. DOI: 10.1109/TCSVT.2019.2897980

[4] ZHANG Yang, LI Qingzhong, ZANG Fengni. Ship detection for visual maritime surveillance from non-stationary platforms [J]. Ocean Engineering, 2017, 141: 53. DOI: 10.1016/j.oceaneng.2017.06.022

[5] KIM K, HONG S, CHOI B, et al. Probabilistic ship detection and classification using deep learning [J]. Applied Sciences, 2018, 8 (6): 936. DOI: 10.3390/app8060936

[6] LEE S J, ROH M I, OH M J. Image-based ship detection using deep learning [J]. Ocean Systems Engineering, 2020, 10 (4): 415. DOI: 10.12989/ose.2020.10.4.415

[7] 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. DOI: 10.1109/TITS.2016.2634580

[8] SHAO Zhenfeng, WU Wenjing, WANG Zhongyuan, et al. SeaShips: a large-scale precisely annotated dataset for ship detection [J]. IEEE Transactions on Multimedia, 2018, 20 (10): 2593. DOI: 10.1109/TMM.2018.2865686

[9] LIU R W, YUAN W, CHEN X, et al. An enhanced CNN-enabled learning method for promoting ship detection in maritime surveillance system [J]. Ocean Engineering, 2021, 235: 109435. DOI: 10.1016/j.oceaneng.2021.109435

[10] NIE Xin, YANG Meifang, RYAN W L. Deep neural network-based robust ship detection under different weather conditions [C]// Proceedings of IEEE Intelligent Transportation Systems Conference (ITSC). Piscataway: IEEE, 2019: 47. DOI: 10.1109/ITSC.2019.8917475

[11] REN Shaoqing, HE Kaiming, 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, 2016, 39(6): 1137. DOI: 10.1109/TPAMI.2016.2577031

[12] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 779. DOI: 10.1109/CVPR.2016.91

[13] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 7263. DOI: 10.1109/CVPR.2017.690

[14] REDMON J, FARHADI A. YOLOv3: an incremental improvement [Z/OL]. (2018-04-08). <https://arxiv.org/abs/1804.02767>

[15] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection [Z/OL]. (2020-04-23) <https://arxiv.org/abs/2004.10934>

[16] GE Zheng, LIU Songtao, WANG Feng, et al. YOLOX: exceeding YOLO series in 2021 [Z/OL]. (2021-07-18). <https://arxiv.org/abs/2107.08430>

[17] WANG C Y, YEH I H, LIAO H Y M. You only learn one representation: unified network for multiple tasks [Z/OL]. (2021-05-10). <https://arxiv.org/abs/2105.04206>

[18] 徐诚极, 王晓峰, 杨亚东. Attention-YOLO: 引入注意力机制的 YOLO 检测算法 [J]. 计算机工程与应用, 2019, 55(6): 13

XU Chengji, WANG Xiaofeng, YANG Yadong. Attention-YOLO: YOLO detection algorithm that introduces attention mechanism [J]. Computer Engineering and Applications, 2019, 55(6): 13. DOI: 10.3778/j.issn.1002-8331.1812-0010

[19] YANG Lingxiao, ZHANG Ruyang, LI Lida, et al. SimAM: a simple, parameter-free attention module for convolutional neural networks [C]// Proceedings of the 38th International Conference on Machine Learning. New York: PMLR, 2021: 11863.

[20] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7132. DOI: 10.1109/TPAMI.2019.2913372

[21] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C]// Proceedings of the 15th European Conference on Computer Vision. Cham: Springer, 2018: 3. DOI: 10.1007/978-3-030-01234-2_1