

doi: 10.7690/bgzdh.2022.10.001

基于改进 YOLOv4-tiny 的舰面多目标检测算法

汪 丁¹, 黄 葵¹, 朱兴动², 范加利¹, 王 正¹

(1. 海军航空大学青岛校区, 山东 青岛 266041; 2. 海军航空大学, 山东 烟台 264001)

摘要: 针对舰面多目标的检测问题, 提出一种改进 YOLOv4-tiny 的舰面多目标检测算法。在卷积神经网络中引入卷积注意力模块(convolutional block attention module, CBAM), 通过混合通道特征和空间特征来关注舰面目标和抑制背景特征, 提高网络的抗背景干扰能力; 针对目标尺度变化加入空间金字塔池化结构(spatial pyramid pooling, SPP)以融合不同尺度的特征, 提高对不同大小目标的检测能力; 使用 Mish 激活函数替代 Leaky ReLU 激活函数以获得更好的泛化能力。实验结果表明: 5 类舰面目标的平均检测精度为 92.22%, 接近 YOLOv4 算法的 96.48%, 而检测速度(frames per second, FPS)达到了 42.5 帧/s, 远高于 YOLOv4 的 18 帧/s; 该算法能较好地平衡准确率和速度的关系, 可以对舰面目标进行实时检测。

关键词: 卷积神经网络; 注意力机制; 目标检测; 空间金字塔池化; 舰面目标

中图分类号: TP391.4 **文献标志码:** A

Shipboard Multi-target Detection Algorithm Based on Improved YOLOv4-tiny

Wang Ding¹, Huang Kui¹, Zhu Xingdong², Fan Jiali¹, Wang Zheng¹

(1. Qingdao Campus, Naval Aviation University, Qingdao 266041, China;

2. Naval Aviation University, Yantai 264001, China)

Abstract: An improved YOLOv4-tiny multi-target detection algorithm is proposed to solve the problem of multi-target detection on shipboard. A convolutional block attention module (CBAM) is introduced into the convolutional neural network to focus on shipboard targets and suppress background features by mixing channel features and spatial features, so as to improve the anti-background interference ability of the network; an spatial pyramid pooling (SPP) structure is added according to the change of target scale to fuse features of different scales, so as to improve the detection ability of targets of different sizes; Mish activation function is used instead of Leaky ReLU activation function for better generalization ability. The experimental results show that the average detection accuracy of five kinds of shipboard targets is 92.22%, which is close to the 96.48% of YOLOv4 algorithm, and the detection speed frames per second (FPS) reaches 42.5 frame/s, which is much higher than the 18 frame/s of YOLOv4 algorithm. The algorithm balances the relationship between accuracy and speed, and can detect the target on the warship in real time.

Keywords: convolutional neural network; attention mechanism; target detection; spatial pyramid pooling; shipboard target

0 引言

航母飞行甲板作为航母保障作业的重要场所, 被称为世界上最危险的 4.5 英亩。飞行甲板作业在存在较大风险的同时, 也成为影响舰载机出动架次率的重要因素。甲板上舰面目标类型复杂, 数量众多, 加上复杂的电磁环境, 在狭窄的空间约束下完成起飞、着舰以及驻留转运, 对舰面目标的空间位置感知成了一个难点问题。

1 相关知识

舰面目标类型众多, 包括各型舰载飞机、运输牵引设备以及作业人员等, 且存在相互遮挡等问题, 传统的技术手段难以达到性能要求。在航母发展早期, 飞行甲板的调度一直依赖于人工通信。二战时,

美国在航母上引入了占卜板(ouija board)来协助舰载机的调度和转运^[1], 理论上, 占卜板上飞机和设备的位置及位置都应当与真实的飞行甲板和机库的情况一致, 但当时的信息融合能力十分有限, 不可避免地会有时间上的延迟和目标的误判。近年来, 计算机视觉以及神经网络的快速发展, 已在多个领域得到广泛的应用, 可为解决调度、转运以及态势感知等问题提供信息支持。

目前卷积神经网络在目标检测上得到了大量应用, 可分为单阶段方法和两阶段方法^[2]2 大类: 1) 单阶段方法直接进行分类和回归, 是一种端到端的检测方式, 如 SSD 以及 YOLO 系列算法, 这类方法检测速度快, 实时性高; 2) 两阶段方法首先生成目标框, 再对目标框进行分类和回归实现目标的类型

收稿日期: 2022-06-01; 修回日期: 2022-07-28

基金项目: 军内科研基金

作者简介: 汪 丁(1997—), 男, 湖南人, 硕士, 从事计算机视觉研究。E-mail: 978032666@qq.com。

识别,如 Faster R-CNN 等算法。模式精度较高,但实时性不佳。文献[3]和文献[4]将神经网络应用到舰载机目标检测上,通过改进 Faster R-CNN 网络模型,使检测准确率分别达到了 93.46%和 95.9%,但其实时性较差,不到 10 帧/s,无法达到动态使用的要求。文献[5]通过对 YOLOv3 算法进行改进,实现了对舰面 5 类目标的有效检测,准确率达到 95.24%,并将帧率提高到了 20 帧/s,实现了初步的目标跟踪能力。文献[6]通过对 YOLOv4 算法进行改进,在加入高精度定位算法后,其定位的速度也不到 10 帧/s。

随着目标检测算法的不断更新迭代,算法的准确率和实时性也在不断提升,但两者一直是一个矛盾体。随着准确率的不断提升,目标检测模型体积不断增加,其实时性也一直难以得到质的提升。YOLO 系列算法自从 YOLOv3 以来得到了快速的发展,在各个领域得到了广泛应用。2020 年 4 月 23 日, YOLOv4 版本由 AlexeyAB^[7]正式发布,通过一系列的改进,新的算法实时性和准确率都得到了提升。同年 6 月, YOLOv4-tiny 版本^[8]正式发布,相比于 v4 版本,模型得到了进一步简化,其速度得到了巨大提升。

由于舰面目标的运动灵活且具有一定的速度, YOLOv4 算法的实时性依旧无法实现目标实时的动态检测,无法为舰面目标的实时跟踪提供支撑。笔者通过对 YOLOv4-tiny 算法进行改进,在得到与 YOLOv4 相接近的精度同时实现更高速度的检测。

2 YOLOv4-tiny 网络

2.1 YOLOv4-tiny 网络结构

如图 1 所示, YOLOv4-tiny 网络主要由 CSPDarknet53-tiny、金字塔模块(feature pyramid network, FPN)以及 YOLOHead 3 个主要结构组成。CSPDarknet53-tiny 主要用于图像特征提取,主要组成部件是 CBL 模块和 CSP 模块。CBL 模块是由卷积运算、批处理以及激活函数组成。而 CSP 模块是 YOLOv4 算法 CSPNet 的简化版,分为 2 部分:1) 残差边堆叠,2) 直接构成残差边,并将 2 部分结合进行一系列卷积运算后输出。该结构可大幅降低运算量且保证精度不出现较大的下降。YOLOHead 在结构上与 YOLOv4 及 YOLOv3 都是相同的,利用融合的特征进行预测,最终形成 13×13 和 26×26 两类尺度。

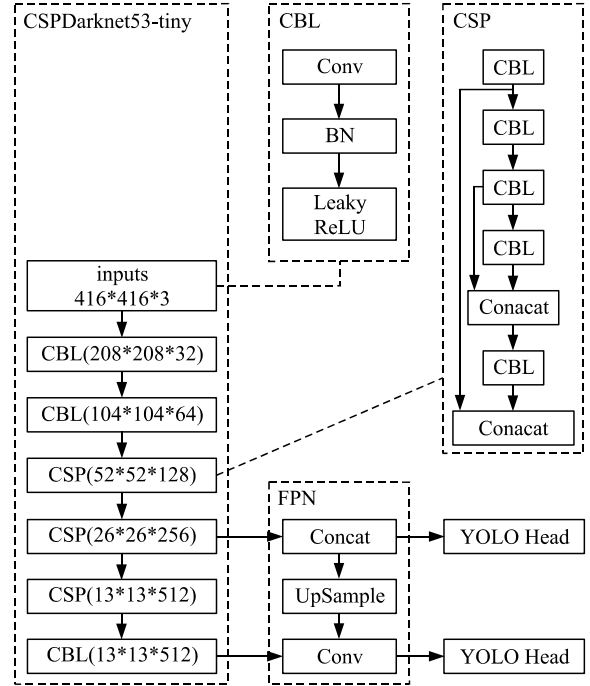


图 1 YOLOv4-tiny 网络结构

2.2 YOLOv4-tiny 原理及流程

YOLOv4-tiny 算法通过 2 个 YOLOHead 分别输出大小为 13×13 以及 26×26 两类尺度特征。其中 13×13 表示将图片分为 13×13 个网格,每个网格点负责一个区域的目标类型的分类与预测。预测结果为 $S \times S \times (w, h, x, y, C_i, \text{Class})$, 其中: $S \times S$ 为输出的两类划分单元格的尺度; Class 为 5 类目标的分类结果; w, h 为预测框的长和宽; x, y 为预测框中心的坐标; C_i 为第 i 个单元格的置信度:

$$C_i = P_o \times \text{IoU}_{\text{pred}}^{\text{truth}} \quad (1)$$

式中: P_o 为目标存在的概率; $\text{IoU}_{\text{pred}}^{\text{truth}}$ 为真实框与预测框的 IoU 值。由于 IoU 值对目标物体的尺寸不敏感,无法直接优化没有重叠的部分,而利用 CIoU 来优化边界框损失函数是更好的方法。CIoU 在计算过程中包括了距离、相交面积、尺度以及惩罚项等元素,使得其计算结果变得更加稳定,回归效果也更佳,计算公式为:

$$\text{CIoU} = \text{IoU} - (\rho^2(b, b^{gt}) / c^2) - \alpha v; \quad (2)$$

$$\alpha = v / (1 - \text{IoU} + v); \quad (3)$$

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

式中: $\rho^2(b, b^{gt})$ 为先验框和真实框的中心间距; c 为先验框和真实框相交域的对角线距离; α 和 v 分别为用来衡量方差与偏差以及长宽比的参数。

3 改进方法

3.1 引入注意力机制

近年来，注意力模型 (attention model, AM)^[9] 已成为神经网络中的一个重要概念，已从自然语言领域发展到了视觉领域的应用。舰面目标检测过程中，背景与目标之间存在着较为严重的干扰，可以

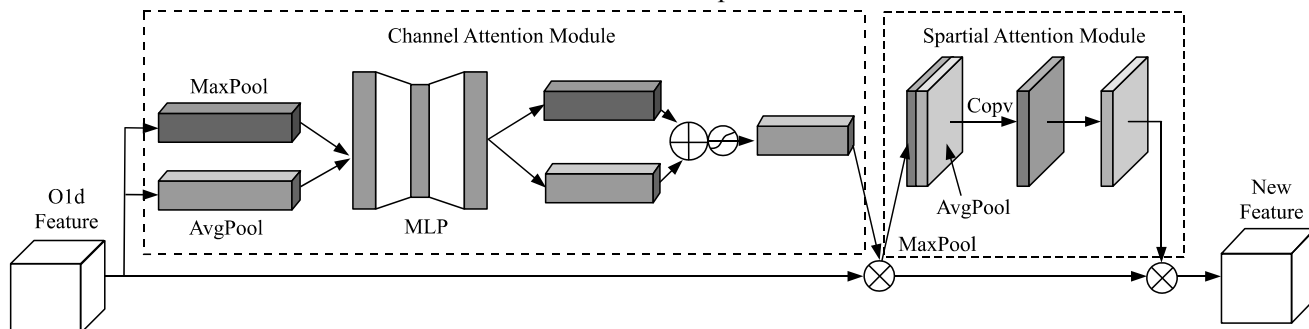


图 2 CBAM 注意力机制模块

在 CAM 模块中，主要通过池化和卷积来分别获取 2 个通道的特征。首先将卷积得到的原特征 O 采用 MaxPool 和 AvgPool 2 种方法进行池化，分别得到 $AvgPool(O)$ 和 $MaxPool(O)$ ，将其输入到多层感知机 (multi-layer perception, MLP) 中，并将输出的对应特征相加后通过激活函数，得到通道注意力权重分布 M_c ，计算公式如下：

$$M_c(O) = \sigma(\text{MLP}(\text{AvgPool}(O)) + \text{MLP}(\text{MaxPool}(O))) \quad (5)$$

式中： O 为卷积后的原特征； σ 为 Sigmoid 激活函数。

在 SAM 模块中，将 CAM 模块得到的特征 M_c 与原特征 O 相乘后作为输入，同样利用 2 种池化操作后堆叠得到 2 个通道的特征图，并利用 7×7 卷积核进行卷积，再通过激活函数得到空间注意力权重分布 M_s ，作为新的特征继续参与下层卷积操作。计算公式如下：

$$M_s = \sigma(f^{7 \times 7}([\text{AvgPool}(F)]; [\text{MaxPool}(F)])) \quad (6)$$

式中： F 为 M_c 和原特征 O 相乘的结果； $f^{7 \times 7}$ 为进行卷积核大小为 7×7 的卷积操作。

在 SAM 和 CAM 模块的组合方式和池化效果上，通过实验表明，2 个子模块在先通道后空间的串行顺序时效果最好^[10]，而并行的池化操作可以更好地保留特征信息，防止在池化过程中出现特征丢失。与 SE (squeeze-and-excitation)^[11] 模块相比，CBAM 注意力模块^[12]既考虑了不同通道像素的重要性，又考虑到同一通道不同位置像素的重要性。在提升模型性能的同时计算量较小，对模型性能影响较小。

通过引入注意力机制，更好地获取舰面目标的特征，将其与舰面背景剥离开来，增加目标检测算法的抗干扰能力。

如图 2 所示，卷积注意力模块 (CBAM) 是一个卷积注意力机制模块，包括通道注意力模块 (channel attention module, CAM) 和空间注意力模块 (spatial attention module, SAM) 2 个子模块。

3.2 多尺度目标优化

在舰面目标检测过程中，存在的难点是：1) 作业人员等小目标在距镜头较远时图像像素信息较少，因此难以达到较高的准确率；2) 由于飞机等大型目标过于靠近镜头或者遮挡导致图像中只有其部分特征，以至检测框位置出现偏差。舰面目标大小差异较大，加之甲板长度导致的远近不一，所以难以有较全面目标检测能力。

通过实验对比发现：作业人员等小目标检测召回率明显低于其他 4 类目标，需要针对这类小目标进行优化。在先验框设定上，锚框的大小也会影响到模型的收敛和精度^[13]，需要根据目标大小来进行设定，聚结类果如图 3 所示。

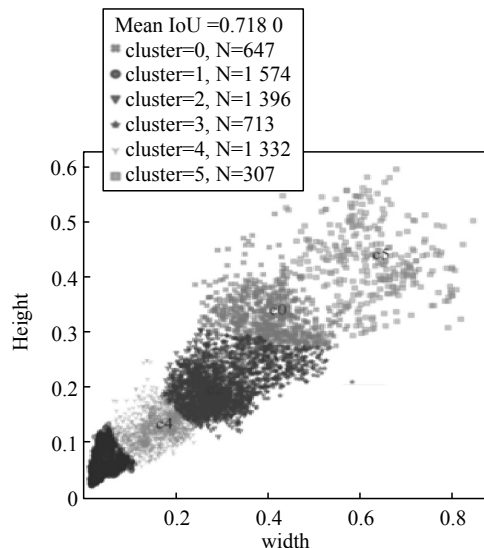


图 3 聚结类果

通过 k-means 聚类算法调整数据集中目标的先验框大小, 并将输出特征图由 13×13 和 26×26 改为 19×19 和 38×38, 通过增大先验框数值以获取更细分的网格^[14], 增强小目标的检测效果。同时, 笔者采用离线数据增强方法, 提高小目标在数据集中的比例, 提高模型对小目标的检测能力。

针对舰面目标在甲板上来回移动而导致的目标尺度变化带来的检测性能下降, 在 FPN 前加入空间金字塔池化结构(SPP), 通过融合多个尺度的目标特征信息, 提高模型对不同大小目标的检测能力^[15], 其结构如图 4 所示。

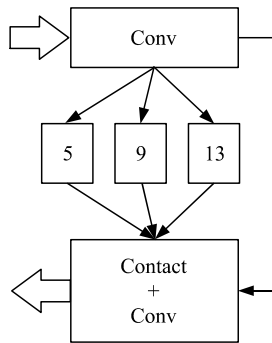


图 4 空间金字塔池化结构

在 SPP 结构中, 采用 3 个尺度大小的池化核进行最大化池化操作, 可以实现舰面目标在远距离或近距离时, 保证有相应的尺度与目标大小相接近, 从而防止由于较大尺寸比而产生的目标丢失问题。同时, SPP 结构也可以让相机获取到尺度不同的图像在模型上顺利运行, 提高了模型宽容度。

3.3 激活函数改进

在 YOLOv4-tiny 算法中, 为了追求极限的速度而采用 Leaky ReLU 激活函数:

$$\text{LeakyReLU} = \begin{cases} x & x \geq 0 \\ p * x & x < 0 \end{cases} \quad (7)$$

式中: p 为由用户定义的超参数, 可有效防止梯度弥散。

在实时性满足需要的前提下, 考虑利用 Mish 激活函数替代 Leaky ReLU 激活函数, 牺牲一部分速度来换取更高的准确率^[16]。文献^[17]表明 Mish 激活函数在最终准确率上相比于 ReLU 激活函数提高了 2.1%。Mish 激活函数的公式如下:

$$f(x) = x \tanh(\ln(1 + e^x)) \quad (8)$$

Mish 激活函数是一条光滑的曲线, 如图 5 所示。Mish 激活函数避免了像 ReLU 激活函数一样的激活状态突变的问题, 同时与 ReLU 和 Swish 相比,

Mish 的极小值范围更大, 提高了泛化能力。Mish 函数的优点是由其曲线的特性决定的。由于没有上界而避免了饱和, 因而不会出现由于梯度接近 0 而导致的训练缓慢。同时存在下界则可以产生较好的正则化效果, 防止训练过程中出现过拟合。

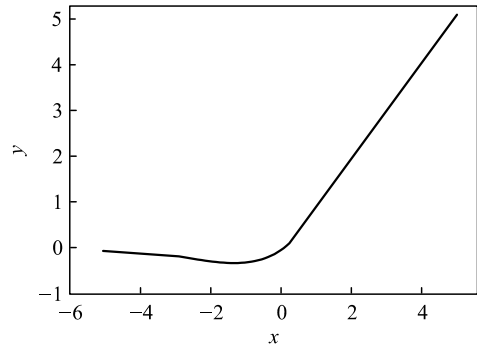


图 5 Mish 函数

4 实验结果

4.1 实验环境

实验环境根据比例模型搭建, 通过工业相机在多个角度进行拍摄, 总共采集了 2 000 张左右不同角度和状态下舰面目标的图像, 得到 PASCAL VOC 数据集格式, 训练集、验证集、测试集比例为 6:2:2。实验具体环境配置及参数设定如表 1 所示。

表 1 实验环境及参数

参数	配置信息
CPU	AMD 4 800H
内存	16 G
GPU	NVIDIA 2 060
显存	8 G
系统	Ubuntu16.04
深度学习框架	TensorFlow1.14
加速环境	CUDA9.0, CUDNN7.0
迭代轮数	100
优化器	Adam
Batchsize	32

4.2 评价指标

为评价网络的有效性, 采用平均准确率均值 (mean average precision, mAP)、精确率 (precision, P)、召回率 (recall, R)、检测帧率 (FPS) 以及模型体积作为实验的评价标准, 其中评价指标计算量如表 2 所示。

表 2 评价指标

实际结果	预测为正 (Positive)	预测为负 (Negative)
实际为正 (Ture)	TP	FN
实际为负 (False)	FP	TN

1) Precision 表示网络检测到正样本数量占检测到所有样本的比率, Recall 指网络检测到的正样本

数量占标记真实样本数量的比率。计算公式如下：

$$\text{Precision} = TP / (TP + FP); \quad (9)$$

$$\text{Recall} = TP / (TP + FN)。 \quad (10)$$

2) mAP 是一种多标签图像分类任务的评价方法，通常是用准确率和召回率形成的 PR (precision-recall) 所围成的最大面积 AP 来衡量。所有类的 AP 的均值就是 mAP。AP 和 mAP 的计算公式为：

$$\text{AP} = \int_0^1 P(R) dR; \quad (11)$$

$$\text{mAP} = \sum_{i=1}^N \text{AP}_i / N。 \quad (12)$$

式中： $P(R)$ 为准确率-召回率曲线；0~1 为阈值变化范围。

3) FPS 用来描述每秒中检测的图像帧数。

4) 模型体积，即训练得到的模型所占用的存储空间大小。

4.3 实验结果与分析

在经过 100 轮迭代训练后，得到模型与训练记录，损失函数如图 6 所示。

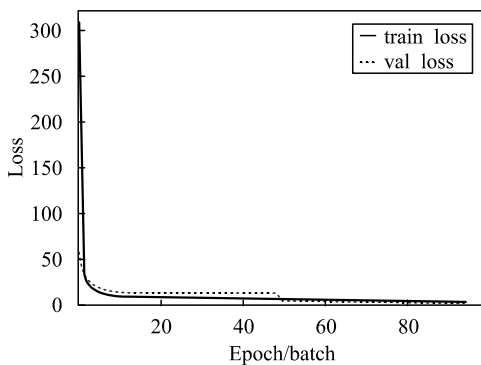


图 6 损失函数

选取最优轮次模型在测试集上进行测试来评估算法的性能，在测试集中选取 250 张图像送入到模

型中进行检测，检测速度的波动如图 7 所示。其中，第 1 帧图像由于模型加载影响了性能，从第 2 帧开始模型开始稳定运行，帧率波动范围小，平均帧率在 42.5 帧/s，可以满足实时目标检测的要求。

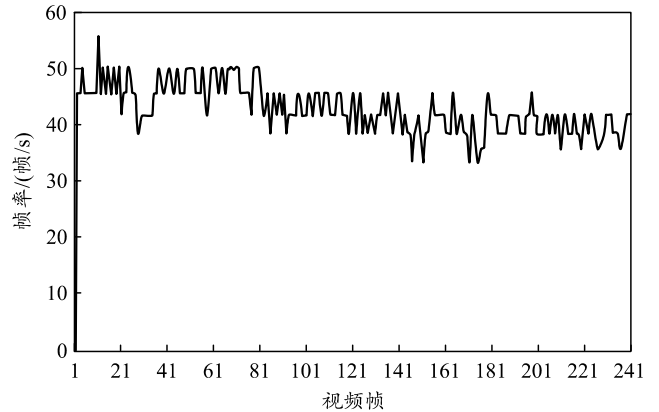


图 7 模型检测速率

对舰面目标的检测基本能够实现在密集目标情况下实现较高精度的检测。如图 8 所示，在暗光环境下仍有较好的效果。

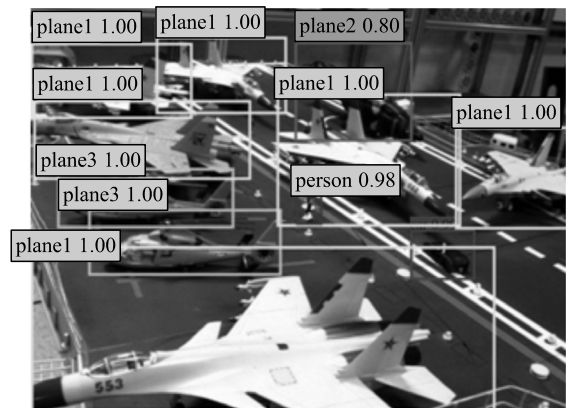


图 8 目标检测效果

为更好地评估算法性能，分别将 YOLOv3 网络、YOLOv4 网络、Faster R-CNN 进行对比试验，测试结果如表 3 所示。

表 3 算法性能对比

算法	Plane1 (AP)/%	Plane2 (AP)/%	Plane3 (AP)/%	Car (AP)/%	Person (AP)/%	平均准确率 (mAP)/%	帧率 (FPS)/(frame/s)	模型体积 /MB
Faster R-CNN	98.00	95.00	95.00	93.00	89.00	94.00	8.0	1 126.4
改进的 YOLOv3	98.00	99.00	98.00	93.00	88.00	95.24	15.6	985.4
YOLOv4	100.00	99.00	99.00	94.00	90.00	96.48	18.2	245.0
YOLOv4-tiny	95.00	90.00	89.00	86.00	50.00	82.00	55.2	22.0
改进的 YOLOv4-tiny	100.00	99.00	99.00	97.00	66.00	92.22	42.5	23.1

通过实验对比可以看出，Faster R-CNN 算法在精确度上具有较高的优势，但由于是两阶段的检测方法，模型体积巨大，且检测速度明显较慢。而文献[5]中改进的 YOLOv3 和 YOLOv4 算法精度和实时性得到了再次提升，模型体积也得到了进一步优

化，达到了基本能用的水平，但仍然无法满足运动目标高实时性的需要。改进的 YOLOv4-tiny 算法在准确率上略低于其他类型，但实时性却得到了巨大提升，其已经超过了正常视频的 30 帧/s，依旧达到了较好的效果，完全可以胜任高实时性的检测任务。

在模型体积上,不到 YOLOv4 体积的十分之一,非常适合在移动终端上进行部署。从 5 类舰面目标的检测效果分析,除了作业人员外其他类型目标的准确率都能够达到 95%以上。由于作业人员视野相对开阔,且位置相对固定,因此 66%左右的准确率可以为舰面提供基本的辅助手段。

5 结论

笔者提出了一种改进的 YOLOv4-tiny 算法。通过引入 CBAM 注意力机制模块和 SPP 结构、对先验锚框和激活函数优化以及数据增强,提高了舰面目标的检测精度和模型泛化能力,尤其是在小目标的检测上得到了较大提升。实验结果表明:在平均帧率达到 42.5 帧/s 时平均检测精度提升到了 92.22%,即改进算法可以在高帧率下实现对舰面目标的有效检测,且具有较好的鲁棒性。

参考文献:

- [1] 杨放青. 航母飞行甲板作业能力分析 with 优化研究[D]. 哈尔滨: 哈尔滨工程大学, 2018.
- [2] WEN B, WU M. Study on Pedestrian Detection Based on an Improved YOLOv4 Algorithm: 2020 IEEE 6th International Conference on Computer and Communications (ICCC)[C]. IEEE, 2020.
- [3] 朱兴动, 田少兵, 黄葵, 等. 基于深度卷积神经网络的舰载机目标检测[J]. 计算机应用, 2020, 40(5): 1529-1533.
- [4] 范加利, 田少兵, 黄葵, 等. 基于 Faster R-CNN 的航母舰面多尺度目标检测算法[J]. 系统工程与电子技术, 2021(8): 1-10.
- [5] 朱兴动, 汪丁, 范加利, 等. 复杂场景下基于增强 YOLOv3 的舰面多目标检测[J]. 计算机工程与应用, 2021(4): 1-9.
- [6] 朱齐丹, 李小铜, 郑天昊. 舰载机位姿实时视觉测量算法研究[J]. 智能系统学报, 2021(7): 1-10.
- [7] BOCHKOVSKIY A, WANG C Y, LIAO H. YOLOv4: Optimal Speed and Accuracy of Object Detection[J/OL]. arXiv preprint, 2020[2021-09-07]. <https://arxiv.org/abs/2004.10934>.
- [8] WANG C Y, BOCHKOVSKIY A, LIAO H. Scaled-YOLOv4: Scaling Cross Stage Partial Network[J/OL]. arXiv preprint, 2020[2021-09-07]. <https://arxiv.org/abs/2011.08036>.
- [9] GAO C, CAI Q, MING S. YOLOv4 Object Detection Algorithm with Efficient Channel Attention Mechanism: 2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE)[C]. ResearchGate, 2020.
- [10] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional Block Attention Module[C]. European Conference on Computer Vision, 2018.
- [11] JIE H, LI S, GANG S, et al. Squeeze-and-Excitation Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.
- [12] ALSANABANI A A, SAEED S A, AL-MKHLAFI M, et al. A Low Cost and Real Time Vehicle Detection Using Enhanced YOLOv4-Tiny: 2021 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)[C]. IEEE, 2021.
- [13] DENG H F, CHENG J H, LIU T, et al. Research on Iron Surface Crack Detection Algorithm Based on Improved YOLOv4 Network[J]. Journal of Physics: Conference Series, 2020.
- [14] 周捷, 徐光辉, 朱东林, 等. 改进 YOLOv4-tiny 的雨雪道路环境下的实时目标检测[J]. 信号处理, 2021, 37(8): 1550-1558.
- [15] PRASETYO E, SUCIATI N, FATICHAH C. Yolov4-tiny and Spatial Pyramid Pooling for Detecting Head and Tail of Fish: 2021 International Conference on Artificial Intelligence and Computer Science Technology (ICAICST)[C]. ResearchGate, 2021.
- [16] 张欣, 张永强, 何斌, 等. 基于 YOLOv4-tiny 的遥感图像飞机目标检测技术研究[J]. 光学技术, 2021, 47(3): 344-351.
- [17] MISRA D. Mish: A Self Regularized Non-Monotonic Neural Activation Function[J/OL]. arXiv preprint, 2019[2021-09-07]. <https://arxiv.org/abs/1908.08681v1>.