

张庆旭, 耿志卿, 程亚鹏, 等. 基于改进 YOLOv5s 的航拍图像目标检测[J]. 智能计算机与应用, 2025, 15(1): 24-31. DOI: 10.20169/j.issn.2095-2163.250104

基于改进 YOLOv5s 的航拍图像目标检测

张庆旭, 耿志卿, 程亚鹏, 苏嘉涛

(河北工程大学 信息与电气工程学院, 河北 邯郸 056038)

摘要: 针对航拍图像中的背景复杂、冗余信息过多、小目标检测不到的问题, 本文提出了一种改进 YOLOv5s 的目标检测算法 (GGS-YOLOv5)。首先, 在 Backbone 网络中加入了 GAM 注意力机制, 减少复杂背景的干扰, 抑制冗余信息, 侧重检测目标, 增强模型的特征提取能力, 还提出了一种新型结构 SPPFCSPC, 增强感受野的同时提高检测速度和精度; 其次, 在 Neck 网络中引入 GSConv 模块, 减少语义信息的丢失, 增强全局感知和特征融合能力; 最后, 更换损失函数为 SIOU, 添加角度惩罚成本, 有效地降低了自由度, 进一步提升模型的收敛速度以及检测精度。将本文提出的算法在 SeaDroneSee 数据集进行消融以及对比实验, 结果表明, 该算法比原 YOLOv5s 召回率提高了 4.9%, mAP 0.5 提高了 2.8%。

关键词: 航拍图像; GSConv; SPPFCSPC; 全局注意力机制; SIOU

中图分类号: TP391.4

文献标志码: A

文章编号: 2095-2163(2025)01-0024-08

Improve YOLOv5s object detection algorithm for aerial images

ZHANG Qingxu, GENG Zhiqing, CHENG Yapeng, SU Jiatao

(School of Information and Electrical Engineering, Hebei University of Engineering, Handan 056038, Hebei, China)

Abstract: Aiming at the problems of complex background, excessive redundant information and undetectable small targets in aerial images, an improved object detection algorithm (GGS-YOLOv5) of YOLOv5s is proposed. Firstly, the GAM attention mechanism is added to the Backbone network to reduce the interference of complex backgrounds, suppress redundant information, focus on detection targets, and enhance the feature extraction ability of the model, and a new structure, SPPFCSPC, is proposed to enhance the receptive field while improving the detection speed and accuracy. Secondly, the GSConv module is introduced in the Neck network to reduce the loss of semantic information and enhance global perception and feature fusion capabilities. Finally, the loss function is replaced with SIOU, and the angle penalty cost is added to effectively reduce the degree of freedom, further improve the convergence speed and detection accuracy of the model. The results of the algorithm ablation in the SeaDroneSee dataset and the comparative experimental results show that the proposed algorithm improves the recall rate by 4.9% and mAP 0.5 by 2.8% compared with the original YOLOv5s.

Key words: aerial image; GSConv; SPPFCSPC; global attention mechanism; SIOU

0 引言

航拍图像目标检测常用于航拍侦察、公路管理、海上救援等各个方面。而海上救援目标检测作为航拍图像目标检测的一个重要方面, 主要任务是准确地检测出海面上被困的船舶和人员, 方便派出救援人员。海上航拍采用空中装置对海面采取信息, 可以改良采集受限和目标遮挡的问题, 但是远距离和俯视角度拍摄物体, 图像会呈现出分辨率高的情况

下多目标聚集、小目标比重占比大以及目标尺寸多样等特点, 从而使准确率大幅降低。传统的目标检测方法耗费时间长、检测精度低, 无法取得良好的检测效果。

目前, 深度学习中目标检测分为 Two-Stage 和 One-Stage 两种类型的算法, 区别在于是否生成候选框。Two-Stage 的目标检测算法需要先获得样本的候选框, 这类算法的典型代表是基于区域候选网络的 R-CNN (Region-Convolutional Neural

基金项目: 河北省高等学校科学技术研究项目(QN2021034)。

作者简介: 张庆旭(2000—), 男, 硕士研究生, 主要研究方向: 计算机视觉; 程亚鹏(1999—), 男, 硕士研究生, 主要研究方向: 图像识别。

通信作者: 耿志卿(1982—), 男, 博士, 副教授, 主要研究方向: 计算机视觉, 图像识别。Email: gengzhiqin05@sina.com。

收稿日期: 2023-08-07

Networks)系算法,如Fast R-CNN、Faster R-CNN等;One-Stage的目标检测算法则直接将采用回归的方式对目标边框定位处理,典型算法有SSD(Single Shot MultiBox Detector)系列和(You Only Look Once)系列。文献[1]在利用深层语义信息自下而上地与浅层语义信息进行融合,从而丰富小目标的特征信息,有效地提高了对航拍小目标的检测能力;文献[2]在原特征提取网络结构上添加SE(Squeeze-and-Excitation)注意力机制,提高模型对有用信息的专注能力;文献[3]提出一种通道-全局注意力机制(CGAM),以增强对不同尺度目标的特征提取能力,抑制冗余信息的干扰。目前的算法在航拍目标检测领域上获得了较好的效果,但仍然存在两大不足:其一,面对信息复杂的背景时,对小目标检测的精度较低、对关键信息获取能力较差;其二,特征提取和融合能力较弱,小目标信息传递语义丢失过多,导致精度下降。

本文提出了一种改进的YOLOv5s检测算法(GGS-YOLOv5)。首先,在颈部网络中引入深度可分离卷积(DSC)以及标准卷积(SC),将数据分发(shuffle)标准卷积生成的信息(密集卷积操作)渗透到深度可分离卷积生成的信息的每个部分的GSConv(Generalized-Sparse Convolution),提高算法的特征融合能力,以及减少特征图的空间(宽度和高度)压缩和通道扩展中语义信息的丢失;其次,通过在骨干网络中加入GAM(Global Attention Mechanism)注意力机制,在减少信息弥散的情况下放大全局跨维度的交互特征,提高算法的特征提取能力以及在复杂背景下对关键信息的获取能力。最后,提出一种改进于SPPF(Spatial Pyramid Pooling Fast)的新型结构SPPFCSPC(Spatial Pyramid Pooling Fast, Cross Stage Partial Channel),扩大特征图感受野的同时进一步提升检测精度;最后,将原YOLOv5s网络中的损失函数*CIoU*(Complete-IoU)替换为*SIoU*(Soft Intersection over Union)。*SIoU*通过添加角度惩罚成本有效地降低了自由度,在提高训练速度的同时也提高了检测的准确性^[4]。

1 相关技术及理论

YOLOv5算法是目前应用较为广泛的一阶段目标检测模型,相较于其他算法,YOLOv5s具有较高的精度和较快的检测速度,同时参数量更少。

YOLOv5s目标检测模型结构主要由包括输入端(Input)、骨干(Backbone)、颈部(Neck)、输出端

(Head)4个部分。输入端,本文在数据预处理阶段使用Mosaic数据增强。Mosaic数据增强通过拼接4张随机选取的图片来改善小目标不均匀的状况,提高小目标的检测能力^[5];骨干部分包含Focus结构和C3结构,Focus结构能够减少参数量和计算量,加快训练速度;C3是对残差特征进行学习的主要模块,其结构分为两支,一支使用了多个Bottleneck堆叠和3个标准卷积层,另一支仅经过一个基本卷积模块,最后将两支进行连接操作。Neck网络中包含SPPF模块以及特征融合模块,SPPF模块将SPP(Spatial Pyramid Pooling)模块中并行的最大池化层改为串联,两者的计算结果一样,但SPPF比SPP计算速度快了两倍多;特征融合部分采用FPN(Feature Pyramid Networks)结构生成特征金字塔。FPN通过引入PANet(Path Aggregation Network)把低维度的特征信息向高维度传递,实现多尺寸特征图的融合^[6]。输出端输出尺度分别为 20×20 、 40×40 、 80×80 。最后通过NMS(Non-Maximum Suppression)等后处理操作,对很多目标框进行筛选,消除冗余框,输出置信度得分最高的预测物体的类别。YOLOv5s算法的网络结构及各模块细化如图1所示,其中*c*代表通道数,BN1为BottleNeck1的缩写,BN2为BottleNeck2的缩写,*k*为卷积核大小。

2 算法改进

本文在YOLOv5s算法的基础上从Backbone网络、Neck网络、损失函数3个方面入手:在Backbone网络中C3模块以及SPPF模块后面融入GAM注意力机制,提高检测目标的关注度,增强小目标信息提取能力;另外,为了增强感受野,提升检测精度,融入了改进的SPPFCSPC结构;为了减少每次特征图的空间(宽度和高度)压缩和通道扩展导致的语义信息部分丢失,本文在Neck网络中引入GSConv模块,减少冗余重复信息;在模型预测部分,采用*SIoU*损失函数对预测框的回归进行优化,加快收敛速度。本文YOLOv5s算法的网络结构如图2所示,虚线框内为改进部分。

2.1 GAM注意力机制模块

本文使用数据集为航拍图像,航拍图像的目标普遍较小,覆盖区域大,信息量比较大,信息一般难以有效提取,在目标检测的卷积运算中,冗余信息会随着卷积运算次数的增加而越来越多,从而干扰到算法检测待检目标,增加了目标的误检率和漏检率,最终使得检测精度减小。本文在Backbone网络中引入GAM

注意力机制模块,可以减少冗余信息,聚焦并选择对任务有用的信息,使检测器对目标进行有区别的检测,提升模型的检测精度。GAM 全局注意力模块能够提升对海上航拍小目标的注意程度,降低冗余信息的干扰,提升模型检测精度。GAM 采用顺序通道注

意力子模块(Channel Attention, CA)-空间注意力子模块(Spatial Attention, SA)并重新设计了(Convolutional Block Attention Module, CBAM)从而在减少信息弥散的情况下也能放大大局跨维度的交互信息^[7]。GAM 注意力机制的实现过程如图 3 所示。

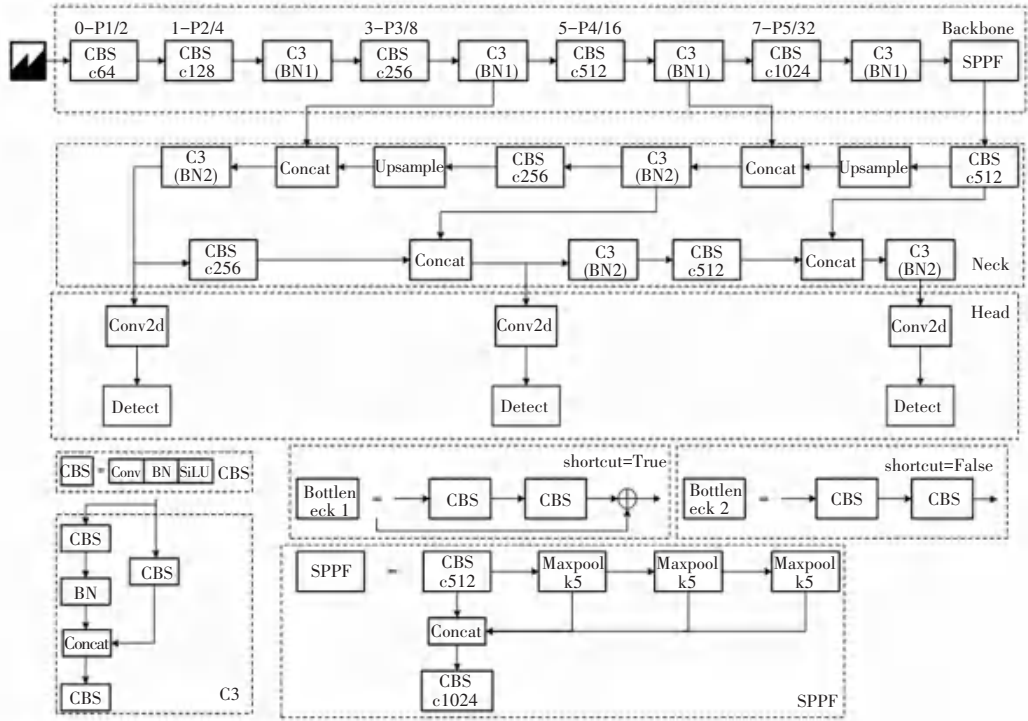


图 1 YOLOv5s 算法的网络结构及各模块细化图

Fig. 1 YOLOv5s algorithm network structure and each module refinement diagram

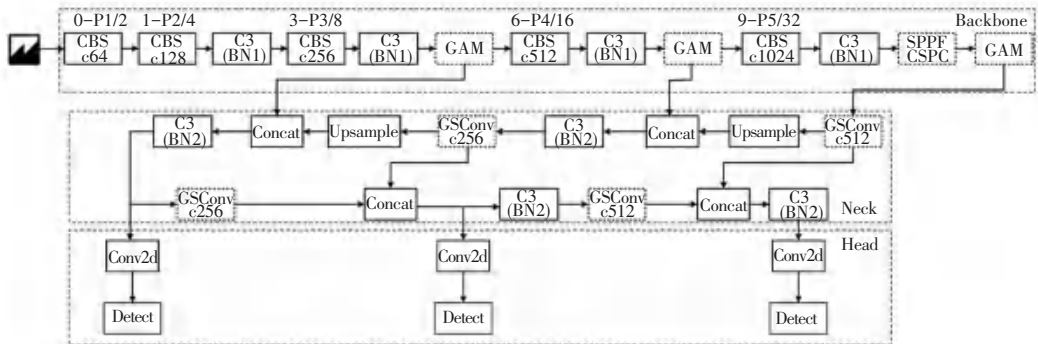


图 2 改进后的 YOLOv5s 算法的网络结构图

Fig. 2 Network structure diagram of the improved YOLOv5s algorithm

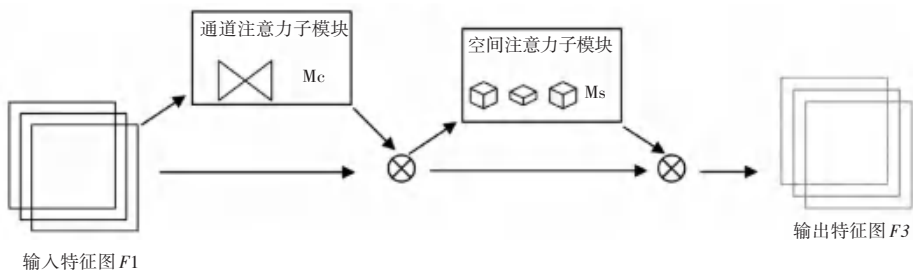


图 3 GAM 注意力机制的实现过程

Fig. 3 Implementation process of GAM attention mechanism

F_1 为给定输入特征映射 $F_1 \in R^{C*H*W}$, 中间状态 F_2 和 F_3 输出定义如下:

$$F_2 = M_c(F_1) \otimes F_1 \quad (1)$$

$$F_3 = M_s(F_2) \otimes F_2 \quad (2)$$

其中, M_c 和 M_s 分别为通道和空间注意力, \otimes

表示按元素进行乘法操作。

通道注意力子模块采取三维排列在 3 个维度上保留信息, 随后采取了两层的多层感知机 (MLP) 来放大跨维通道-空间依赖度, 从而提高特征提取能力。通道注意力子模块实现过程如图 4 所示。

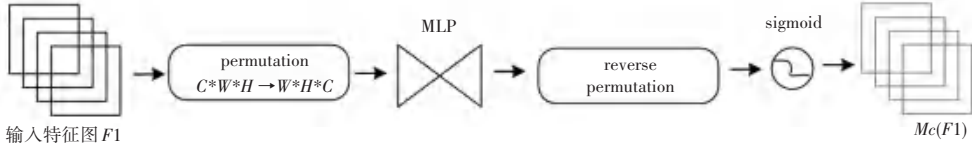


图 4 通道注意力子模块的实现过程

Fig. 4 Realization process of channel attention submodule

空间注意力主要是为了关注目标在输入图像的位置信息, 聚焦有效信息丰富的部分, 这里 GAM 创新性地采用了两个 7×7 的 Conv 对获取到的空间信息进行融合, 能够更多地捕获空间之间的关系。另外, GAM 使用了与 BAM (Bottleneck Attention

Module) 相同的通道注意力子模块的缩减比 r , 能够更有效地控制注意力模块的容量与计算开销^[8]。但 GAM 剔除了池化操作, 使得输出特征包含了 3 个维度的相互作用。空间注意力子模块的实现过程如图 5 所示。

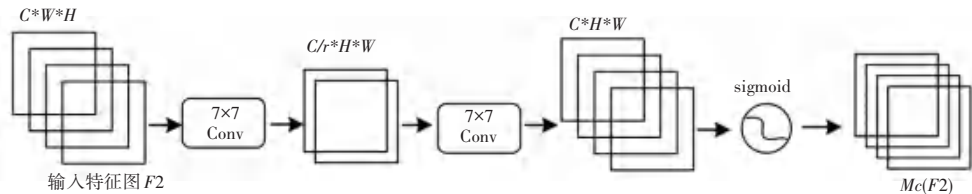


图 5 空间注意力子模块的实现过程

Fig. 5 Implementation process of spatial attention submodule

由于 Backbone 模块包含多次卷积运算, 冗余信息较多, 加入 GAM 模块后使模型更加关注检测目标的位置信息, 聚焦于有效信息丰富的部分, 并提高模型的特征提取能力, 从而提高模型的检测精度。

2.2 SPPFCSPC 结构

SPPFCSPC 结构是基于 YOLOv7 中 SPPCSPC 结构的改进, 而 SPPCSPC 结构是在 SPP 结构的基础上改进而来。SPP 对输入的特征图进行多尺度的空间金字塔池化, 从而得到多个不同尺度的特征

图^[9]。

SPPCSPC 结构在 SPP 上增加不同尺度卷积核的卷积层, 对 SPP 输出的特征图进行卷积操作, 从而可以捕捉不同尺度下的检测目标。本文借鉴了 YOLOv5-6.0 的 SPPF 结构对于 SPP 结构的改进, 将 SPPCSPC 结构中的 Maxpool 改为串联, 模型的计算量变小, 在提升计算速度的同时进一步提升了检测精度。SPPFCSPC 结构如图 6 所示。

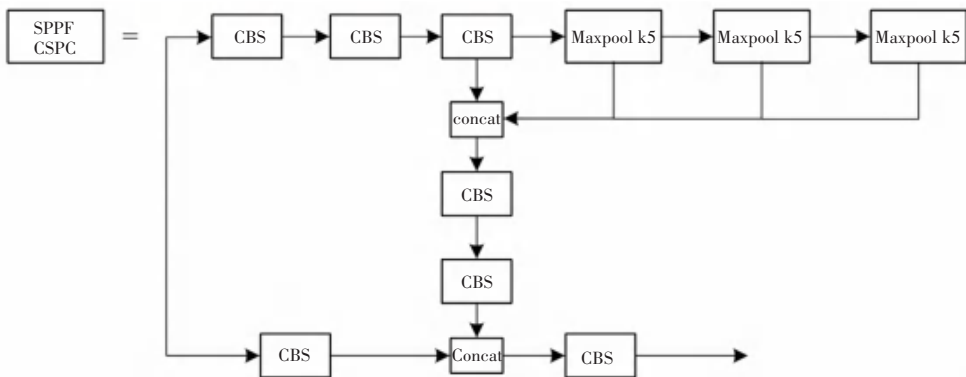


图 6 SPPFCSPC 结构

Fig. 6 SPPFCSPC structure

2.3 GSConv 模块

现代大多数架构通过增加模型参数的数量来构建强大的模型,一般会引入深度可分离卷积来减少卷积的参数,但由于深度可分离卷积层输入图像的通道信息在计算过程中是分离的,模型的能力可能会显著

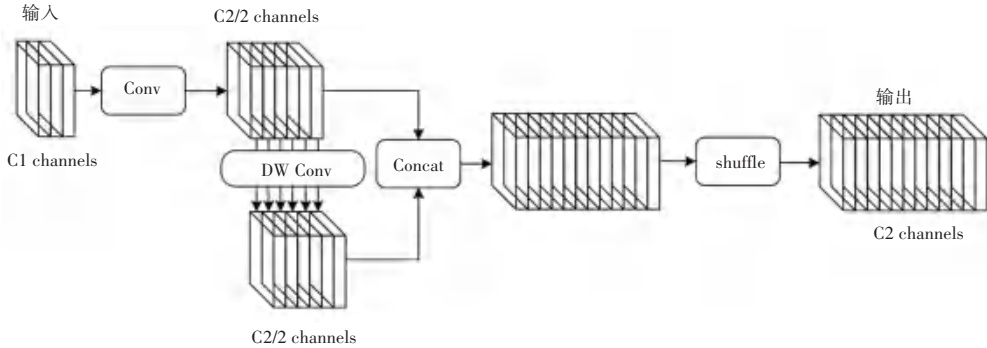


图7 GSConv 结构图

Fig. 7 GSConv structure diagram

仅在 Neck 网络中使用 GSConv,是因为如果把 YOLOv5s 模型中所有 Conv 都替换为 GSConv,模型的网络层会加深,网络结构层数加深会显著增加推理时间。另外,当经过 Backbone 网络的特征图输入到 Neck 网络时,特征图是通道维度达到最大,宽高维度达到最小的状态,不再需要变换。在 Neck 网络中使用 GSconv 处理这些特征图时,冗余信息少,不需要压缩,对增加的注意力模块提升效果更好。

2.4 损失函数 SIoU

YOLOv5s 用的损失函数是 $CIoU$,但是 $CIoU$ 考虑了预测的边界框和预测框之间的距离、重叠区域、纵横比等因素,没有考虑到方向不匹配的问题,可能会导致收敛速度较慢,最终得到一个较差的模型。而 $SIoU$ 损失函数重新定义了惩罚指标并且考虑了回归向量的角度,本文将 $CIoU$ 损失函数替换为 $SIoU$ 损失函数,算法在数据集上的检测精度有所提升。

$SIoU$ 损失函数一共包含 4 个部分:角度损失 (Angle cost),距离损失 (Distance cost),形状损失 (Shape cost), IoU 损失 (IoU cost)。 $SIoU$ 损失函数中各参数关系的示意图如图 8 所示。

角度损失公式:

$$\Lambda = 1 - 2 \times \sin^2\left(\arcsin\left(\frac{C_h}{\sigma}\right) - \frac{\pi}{4}\right) \quad (3)$$

其中, c_h 为预测框与真实框中心点的高度差, σ 为预测框与真实框中心点的距离。

距离损失公式:

$$\Delta = \sum_{t=x,y} (1 - e^{-\rho_t}) \quad (4)$$

下降。GSConv 模块将深度可分离卷积和标准卷积结合在一起,使用数据分发操作将标准卷积生成的信息密集卷积操作渗透到深度可分离卷积生成的信息的每个部分,降低计算复杂度的同时保持精度^[10]。本文引入 GSConv 模块,GSConv 结构如图 7 所示。

其中:

$$\rho_x = \left(\frac{b_{c_x}^{g_t} - b_{c_x}}{c_w}\right)^2 \quad (5)$$

$$\rho_y = \left(\frac{b_{c_y}^{g_t} - b_{c_y}}{c_h}\right)^2 \quad (6)$$

$$\gamma = 2 - \Lambda \quad (7)$$

其中, ρ_x 和 ρ_y 分别为宽度和高度的差值平方, γ 为角度损失的调整因子。

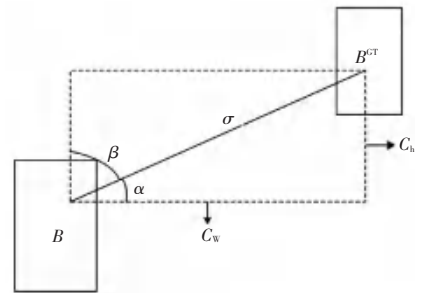


图8 $SIoU$ 损失函数中各参数关系的示意图

Fig. 8 Diagram of the relationship between parameters in the $SIoU$ loss function

当 α 趋于 0 时,距离损失的贡献度越来越低;当 α 趋于 45° 时,距离损失的贡献越来越大。

形状损失公式:

$$\Omega = \sum_{t=w,h} (1 - e^{-w_t})^\theta = (1 - e^{-w_w})^\theta + (1 - e^{-w_h})^\theta \quad (8)$$

其中:

$$w_w = \frac{|w - w^{g_t}|}{\max(w, w^{g_t})} \quad (9)$$

$$w_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})} \quad (10)$$

(w, h) 和 (w^{gt}, h^{gt}) 分别为预测框与真实框的宽高, 通过 θ 对形状损失进行控制, θ 的参数范围为 $[2, 6]$ 。

IoU 损失公式:

$$IoU = \frac{(b \cap b^{gt})}{(b \cup b^{gt})} \quad (11)$$

综上所述, $SIoU$ 损失定义如下:

$$Loss_{Siou} = 1 - IoU + \frac{\Delta + \Omega}{2} \quad (12)$$

3 实验分析

本文使用的操作系统为 Linux, GPU 为 NVIDIA Tesla T4 (16 G 显存), 深度学习框架 Pytorch。模型训练参数设置: 迭代次数 epoch = 100, 批量大小 batch_size = 64, 其余参数为初始值。

3.1 数据集与评价指标

本文使用 SeaDroneSee 数据集来训练和评估模型。SeaDronesSee 是一个大型数据集, 旨在帮助开发在海上场景中使用无人驾驶飞行器 (UAV) 进行搜索和救援 (SAR) 的系统。SeaDroneSee 数据集包含 5 个类别: 人 (swimmer)、船 (boat)、摩托艇 (jetski)、救生装置 (Life Saving Appliances, LSA)、浮标 (buoy)。随机选取训练集图片 8 930 张, 验证集图片 1 547 张, 测试集图片 3 750 张。数据集中小目标占比较高, 适合验证改进算法的性能。

本文采用精确度、召回率、平均精度均值、帧速 4 种评价指标对算法进行评价。

精确度 (Precision, P) 和召回率 (Recall, R), 精确度越高, 召回率越高, 模型和算法就越高效。计算公式如下:

$$P = \frac{T_p}{T_p + F_p} \quad (13)$$

$$R = \frac{T_p}{T_p + F_N} \quad (14)$$

其中, T_p (True Positive) 为正样本被正确识别的数量; F_p (False Positive) 为误检的负样本数量; F_N (False Negative) 为漏检的正样本数量。

平均精度均值 (mean Average Precision, mAP), 即 AP 的平均值, 计算公式如下:

$$mAP = \frac{1}{N} \sum_0^N \int_0^1 P_n(r) dr \quad (15)$$

其中, N 为检测目标类型的个数, P_n 表示某种类型的 AP 值。 $mAP@0.5$ 和 $mAP@0.5:0.95$ 评估模型的综合能力, $mAP@0.5$ 是 IoU 设为 0.5 时, 计算每一类型所有图片的 AP 值, 然后所有类型求平均; $mAP@0.5:0.95$ 表示 IoU 从 0.5 到 0.95, 步长 0.05 上的平均 mAP 。本文中若无特殊说明 mAP 默认为 $mAP@0.5$ 。

帧速 (Frames Per Second, FPS) 为模型每秒可以处理图片的数量, 单位为 f/s , 用来测试模型的实时性, FPS 越大代表模型的实时能力越好^[11]。

3.2 实验结果分析

为了验证算法改进的有效性, 本文在 SeaDroneSee 数据集上验证改进后的 YOLOv5s 算法, 本文进行了注意力机制模块对比实验、损失函数模块对比实验、消融实验、主流算法对比实验。

3.2.1 注意力机制模块对比实验

本文共选取了 4 个注意力机制进行对比, 见表 1。从表 1 中可以看出, 本文使用的 GAM 注意力机制与其他注意力机制相比, 召回率最高, mAP 值最高。

表 1 注意力机制对比实验

Table 1 Attentional mechanism comparison experiment

方法	Precision/%	Recall/%	mAP /%
+CA * 3 ^[12]	90.5	64.2	69.1
+NAM * 3 ^[13]	86.0	65.9	68.3
+CBAM * 3 ^[14]	84.2	66.9	69.4
本文 (GAM)	83.9	67.2	70.4

注意力机制 mAP 值曲线对比如图 9 所示, 可以看出本文使用的 GAM 注意力机制在训练轮次 epochs = 100 时, mAP 值高于其他 3 种注意力机制。

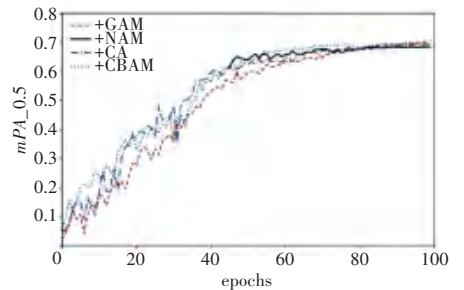


图 9 注意力机制 mAP 值曲线对比图

Fig. 9 Comparison of mAP value curves of attention mechanism

3.2.2 损失函数模块对比实验

为了验证本文所使用损失函数的有效性, 同样选取了 3 个损失函数进行对比, 见表 2。从表 2 中

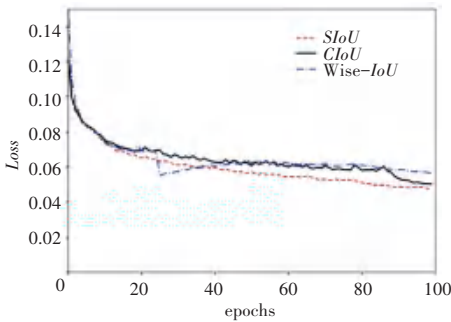
可以看出,本文使用的 $SIoU - loss$ 与其他损失函数相比,准确率最高,同时 mAP 值也是最高。

表2 损失函数对比实验

Table 2 Loss function comparison experiment

损失函数	Precision/%	Recall/%	mAP /%
+ $CIoU$ ^[15]	85.8	64.9	68.3
+Wise- IoU ^[16]	78.9	67.1	69.1
本文 ($SIoU$)	88.6	66.1	70.2

损失函数 $Loss$ 值曲线如图10所示,可以看出本文使用的 $SIoU - loss$, 曲线收敛更快,最终的损失值最低,均优于另外两种损失函数。

图10 损失函数 $Loss$ 值曲线对比图Fig. 10 Comparison of $Loss$ curve of loss function

3.2.3 消融实验

为了验证改进算法各模块的有效性,本文在 SeaDroneSee 数据集上进行消融实验,控制实验过程

中参数的一致性,依次添加 GSCConv 模块、SPPFCSPC 结构、GAM 注意力机制以及 $SIoU$ 损失函数,实验结果见表3。

表3 消融实验

Table 3 Ablation experiment

YOLOv5s	GSCConv	SPPFCSPC	GAM	$SIoU$	mAP /%
✓					68.3
✓	✓				69.7
✓	✓	✓			70.3
✓	✓	✓	✓		70.7
✓	✓	✓	✓	✓	71.1

从表3数据可看出,本文改进的4个模块均有效,使得整体 mAP 上升2.8%。

3.2.4 与主流算法对比实验

本文以近几年来在航拍图像检测常用的算法 Faster R-CNN、YOLOv7-tiny、YOLOX、TPH-YOLOv5s 以及 YOLOv5s 共5种算法作为对照算法与本文算法进行对比实验,实验结果见表4。从表中可以看出,与上述5种算法相比,GGs-YOLOv5算法在召回率、 $mAP@0.5$ 、 $mAP@0.5:0.95$ 上取得最佳结果,在每秒十亿次的浮点运算 ($GFLOPs$)、精度 ($Precision$)、帧速 (FPS) 上表现突出。综合来看,本文所提出的改进算法在航拍图像目标检测方面性能更优。

表4 与主流算法对比实验

Table 4 Comparison experiments with mainstream algorithms

算法	Precision/%	Recall/%	$mAP@0.5$ /%	$mAP@0.5:0.95$ /%	$GFLOPs/10^9$	$FPS/(1/s)$	模型大小/MB
Faster R-CNN ^[17]	26.1	37.9	29.2	14.1	370.2	12.2	110.9
YOLOv7-tiny ^[18]	93.5	53.8	65.8	34.0	13.9	87.2	23.1
YOLOX ^[19]	83.1	64.4	68.4	37.2	26.8	57.2	34.3
TPH-YOLOv5s ^[20]	85.2	63.2	66.6	34.1	36.6	46.5	15.6
YOLOv5s-6.0	85.8	64.9	68.3	36.5	15.8	90.9	13.7
本文算法	85.2	68.8	71.1	37.3	25.3	58.9	30.4

4 结束语

针对检测航拍图像小目标检测精度过低和召回率过低、漏检以及误检的问题,本文基于 YOLOv5s 目标检测算法提出了 GGS-YOLOv5 算法,在 Backbone 主干网络内嵌入 GAM 注意力机制,抑制噪声干扰,提高检测目标的关注度,并提出了一种新型结构 SPPFCSPC,提高检测速度和精度;在多尺度特征融合 Neck 网络中更换普通卷积 Conv 为

GSCConv 模块,提高算法的特征融合能力,减少特征图宽度和高度压缩,通道扩展中语义信息的丢失;最后,在预测阶段采用 $SIoU$ 损失函数,有效减少目标检测框的重叠,提高检测效率和精度。实验结果显示,GGs-YOLOv5 算法有着比原算法更加优越的检测能力,并且通过对比试验可以看出,在 SeaDroneSee 数据集上的综合能力优于其他算法,证明了本文模型改进的有效性。

但本文算法依然存在一些缺陷:改进后的算法

比原算法更加复杂,带来检测能力提升的同时推理时间和模型大小增加,不利于装载到移动设备上实时检测。将来会对算法进行轻量化改进,提升检测能力的同时降低模型大小,提高模型的实时性,进而实现在实时场景中的应用。

参考文献

- [1] 吴靖,韩祿欣,沈英,等. 基于改进 YOLOv4-tiny 的无人机航拍目标检测[J]. 电光与控制,2022,29(12):112-117.
- [2] 王浩雪,曹杰,邱诚,等. 基于改进 YOLOv4 的航拍图像多目标检测方法[J]. 电光与控制,2022,29(5):23-27.
- [3] 李坤亚,欧鸥,刘广滨,等. 改进 YOLOv5 的遥感图像目标检测算法[J]. 计算机工程与应用,2023,59(9):207-214.
- [4] GEVORGYAN Z. SIoU loss: More powerful learning for bounding box regression[J]. arXiv preprint arXiv, 2205.12740, 2022.
- [5] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv,2004.10934, 2020.
- [6] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway,NJ:IEEE, 2017: 2117-2125.
- [7] LIU Y, SHAO Z, HOFFMANN N. Global attention mechanism: Retain information to enhance channel-spatial interactions[J]. arXiv preprint arXiv,2112.05561, 2021.
- [8] PARK J, WOO S, LEE J Y, et al. Bam: Bottleneck attention module[J]. arXiv preprint arXiv,1807.06514, 2018.
- [9] HE K, ZHANG X, REN S, 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.
- [10] LI H, LI J, WEI H, et al. Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicles [J]. arXiv preprint arXiv,2206.02424, 2022.
- [11] SANCHEZ S A, ROMERO H J, MORALES D. A review: Comparison of performance metrics of pretrained models for object detection using the TensorFlow framework [C]// Proceedings of IOP Conference Series: Materials Science and Engineering. IOP Publishing, 2020: 012024.
- [12] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway,NJ:IEEE,2021: 13713-13722.
- [13] LIU Y, SHAO Z, TENG Y, et al. NAM: Normalization-based attention module[J]. arXiv preprint arXiv,2111.12419, 2021.
- [14] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module [C]//Proceedings of the European Conference on Computer Vision (ECCV). Piscataway,NJ:IEEE,2018: 3-19.
- [15] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression [C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020: 12993-13000.
- [16] TONG Z, CHEN Y, XU Z, et al. Wise-IoU: Bounding box regression loss with dynamic focusing mechanism [J]. arXiv preprint arXiv,2301.10051, 2023.
- [17] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149.
- [18] WANG C Y, BOCHKOVSKIY A, LIAO H. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[J]. arXiv preprint arXiv,2207.02696, 2022.
- [19] GE Z, LIU S, WANG F, et al. Yolox: Exceeding yolo series in 2021[J]. arXiv preprint arXiv,2107.08430, 2021.
- [20] ZHU X, LYU S, WANG X, et al. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway, NJ: IEEE,2021: 2778-2788.