

基于YOLO-SCAT的可见光图像无人机小目标检测算法

陈海永 刘博洋 晏行伟

Small object detection algorithm of UAV for visible light images based on YOLO-SCAT

CHEN Haiyong, LIU Boyang, YAN Xingwei

引用本文:

陈海永, 刘博洋, 晏行伟. 基于YOLO-SCAT的可见光图像无人机小目标检测算法[J]. 应用光学, 2025, 46(3): 505–514. DOI: 10.5768/JAO202546.0311004

CHEN Haiyong, LIU Boyang, YAN Xingwei. Small object detection algorithm of UAV for visible light images based on YOLO-SCAT[J]. Journal of Applied Optics, 2025, 46(3): 505–514. DOI: 10.5768/JAO202546.0311004

在线阅读 View online: <https://doi.org/10.5768/JAO202546.0311004>

您可能感兴趣的其他文章

Articles you may be interested in

基于IDOU-YOLO的红外图像无人机目标检测算法

Infrared image UAV target detection algorithm based on IDOU-YOLO

应用光学. 2024, 45(4): 723–731 <https://doi.org/10.5768/JAO202445.0402001>

基于改进YOLOv3的小目标检测算法

Small object detection algorithm based on improved YOLOv3

应用光学. 2024, 45(4): 732–740 <https://doi.org/10.5768/JAO202445.0402002>

基于嵌入式GPU的红外弱小目标检测算法

Infrared weak small target detection algorithm based on embedded GPU

应用光学. 2020, 41(5): 1089–1095 <https://doi.org/10.5768/JAO202041.0506004>

单光子探测三维点云与可见光图像融合处理算法研究

Fusion processing algorithm of single-photon detection for three-dimensional point cloud and visible light image

应用光学. 2021, 42(6): 1034–1039 <https://doi.org/10.5768/JAO202142.0602004>

复杂背景下红外图像弱小目标检测

Dim-small targets detection of infrared images in complex backgrounds

应用光学. 2021, 42(4): 643–650 <https://doi.org/10.5768/JAO202142.0402002>

无人机对地目标自动检测与跟踪技术

Automatic target detecting and tracking technology based on UAV ground target images

应用光学. 2020, 41(6): 1153–1160 <https://doi.org/10.5768/JAO202041.0601003>



关注微信公众号, 获得更多资讯信息

文章编号: 1002-2082 (2025) 03-0505-10

引用格式: 陈海永, 刘博洋, 晏行伟. 基于 YOLO-SCAT 的可见光图像无人机小目标检测算法 [J]. 应用光学, 2025, 46(3): 505-514.

CHEN Haiyong, LIU Boyang, YAN Xingwei. Small object detection algorithm of UAV for visible light images based on YOLO-SCAT[J]. Journal of Applied Optics, 2025, 46(3): 505-514.



在线阅读

基于 YOLO-SCAT 的可见光图像无人机 小目标检测算法

陈海永¹, 刘博洋¹, 晏行伟^{2,3}

(1. 河北工业大学 人工智能与数据科学学院, 天津 300401; 2. 国防科技大学 电子科学学院, 湖南 长沙 410073;

3. 天津先进技术研究院, 天津 300459)

摘 要: 非法使用民用低空小型无人机事件频发, 严重威胁国家和社会和谐发展。针对基于可见光的无人机小目标检测存在漏检、误检和检测平均精度不高等问题, 提出了 YOLO-SCAT (YOLO-SCConv ATT) 算法模型, 通过重构 ELAN (efficient layer aggregation networks) 结构, 减少空间和通道冗余特征; 同时引入注意力机制 CBAM (convolutional block attention module), 训练中在空间和通道维度加强特征提取, 提高模型检测平均精度。实验表明 YOLO-SCAT 模型的精确率、召回率、 F_1 分数、mAP@0.5 和 mAP@0.5:0.95 分别可达 94.4%、94.4%、94.4%、94.7% 和 52.9%, 证实了 YOLO-SCAT 模型提高了在复杂可见光场景下对无人机小目标的检测识别能力, 能更好地满足反无人机系统的实际需求。

关键词: 无人机检测; 小目标检测; 可见光图像; YOLOv7

中图分类号: TN29; TP391.4

文献标志码: A

DOI: 10.5768/JAO202546.0311004

Small object detection algorithm of UAV for visible light images based on YOLO-SCAT

CHEN Haiyong¹, LIU Boyang¹, YAN Xingwei^{2,3}

(1. School of Artificial Intelligence, Hebei University of Technology, Tianjin 300401, China; 2. College of

Electronic Science, National University of Defense Technology, Changsha 410073, China;

3. Tianjin Institute of Advanced Technology, Tianjin 300459, China)

Abstract: The frequent illegal use of civilian low-altitude small unmanned aerial vehicles (UAV) poses a serious threat to the harmonious development of the country and society. In response to the problems of missed detection, false detection, and low detection accuracy in small target detection of UAV based on visible light, a YOLO-SCConv ATT (YOLO-SCAT) algorithm model was proposed to reconstruct the ELAN (efficient layer aggregation networks) structure and reduce redundant spatial and channel features. At the same time, the attention mechanism CBAM (convolutional block attention module) was introduced to enhance feature extraction in spatial and channel dimensions during training, thereby improving the average detection accuracy of the model. The experimental results show that the accuracy, recall, F_1 score, mAP@0.5 and mAP@0.5:0.95 can reach 94.4%, 94.4%, 94.4%, 94.7% and 52.9%, respectively, which proves that the YOLO-SCAT model

收稿日期: 2024-01-29; 修回日期: 2024-07-22

基金项目: 国家自然科学基金 (U21A20482, 62073117); 国家重点研发计划项目 (2022YFB3303800); 河北省自然科学基金 (F2022202064)

作者简介: 陈海永 (1980—), 男, 博士, 教授, 主要从事图像处理、机器视觉、模式识别研究。E-mail: haiyong.chen@hebut.edu.cn

通信作者: 晏行伟 (1985—), 男, 博士, 副教授, 主要从事多源信息融合、无源定位研究。E-mail: yanxingwei@nudt.edu.cn

improves the detection and recognition ability of small targets in complex visible light scenes, and can better meet the practical needs of anti-drone systems.

Key words: UAV detection; small target detection; visible light images; YOLOv7

引言

微小型民用无人机被广泛应用于旅行航拍、农作物生长监测、线路巡航等,但也带来一系列问题,近年来无人机引起的安全事故不断攀升。多地发生“黑飞无人机”事件,无人机的探测研究成为国内外学术界关注焦点。国内外无人机探测技术手段分为雷达探测、无线电信号监测、光电识别跟踪、声音监测^[1]等,其中光电识别包括红外识别和可见光识别^[2],广泛应用于各种反无人机系统中,其中红外识别设备成本较高,其应用受到一定限制,而可见光识别设备成本较低,容易普及使用。

传统无人机检测方法通常采用基于手工人为提取特征和传统机器学习的方法^[3],王晓^[4]设计出一种实时检测方法,方法包含运动信息获取、背景分析、改进视觉显著性 SR(spectral residual)算法^[5] 3个步骤。邵盼愉^[6]设计出一种方向梯度直方图特征提取方式,并结合支持向量机分类器的算法对无人机目标进行检测。传统无人机检测算法在现实场景下应用稳定性差,当场景过于复杂或检测目标出现遮挡情况,检测效果大打折扣,容易出现漏检现象。随着深度学习技术的快速发展,基于深度学习的目标检测算法成为研究热点。陶磊等人^[7]基于 YOLOv3算法^[8]进行无人机定位识别,根据输出视频帧中存在无人机的位置信息,利用 PID(proportion-integral-differential)算法调节摄像头中心朝向追踪无人机,通过多摄像参数解出实际坐标,从而进行定位与检测。薛珊等人^[9]基于 YOLOX-S^[10]目标检测算法,设计自适应空间和通道融合模块,通过增强底层细节特征和高层语义特征融合效果来提高复杂场景下对无人机的检测能力。虽然基于深度学习的目标检测算法在无人机检测领域取得了较好的效果,但是由于复杂场景的影响,仍存在平均精度不高、漏检和误检严重等问题。

基于视觉信息的无人机检测仍是一个具有挑战性的问题,主要存在两个难点:一是无人机的尺寸非常小(尤其是远距离时),缺乏足够的判别特征;二是复杂的背景也会对无人机检测造成较大干扰。针对无人机小目标易受复杂场景和天气影响,而出现无人机漏检、误检和平均精度不高的问题,本文以 YOLOv7 算法^[11]作为可见光无人机小

目标检测的基准网络,引入 SCConv 卷积^[12]进行 ELAN(efficient layer aggregation networks)模块重构,通过减少卷积神经网络中特征之间的空间和通道冗余,从而降低模型参数量,提高模型对特征的提取能力,削弱对冗余特征的关注,降低计算复杂度。同时引入 CBAM(convolutional block attention module)注意力机制^[13],通过动态调整网络中不同空间位置和通道维度的相应特征,提高网络的特征提取能力。将 SCConv 模块能够减少空间和通道冗余特征的特性与 CBAM 注意力机制在空间和通道维度加强小目标特征学习的特性有机结合起来,实验证明改进后的 YOLO-SCAT(YOLO-SCConv ATT)网络有效提高了对复杂场景下可见光无人机小目标的检测能力。

1 检测算法

1.1 YOLO-SCAT 网络模型

目标检测网络 YOLO 通常采用传统 CNN(convolutional neural networks)架构,其中 3×3 卷积层占了模型参数和计算量的大部分,传统卷积运算成本昂贵,易引入冗余特征,因此从减少传统卷积层提取冗余特征出发,可以增强目标检测网络对无人机小目标的特征提取能力。选择 YOLOv7 网络中卷积层堆叠较深的结构,引入 SCConv 进行模块重构,从而充分发挥 SCConv 模块可减少空间和通道维度冗余特征的特性。

如图 1 所示, YOLO-SCAT 网络模型由输入端(input)、主干特征提取网络(backbone)、颈部网络(neck)和检测头(head)4个部分构成。原始的 ELAN 模块利用分割梯度流思想,通过多个卷积层,聚合了不同深度的特征提取信息,避免输入特征或梯度信息因网络深度增加而消失的现象,但这也带来了特征之间的空间和通道冗余。ELAN-SC 模块引入空间和通道重建卷积模块(spatial and channel reconstruction convolution, SCConv),通过减少空间和通道特征冗余,降低计算成本,增强模型的特征信息学习能力。在主干特征提取网络中将最后一个 ELAN 模块替换为 ELAN-SC 模块,通过对输入特征进行分割融合重构,减少输入到颈部网络的冗余特征,接着在 ELAN-SC 模块后面引入

CBAM注意力机制,有利于模型在空间和通道维度高效学习目标特征。MPC(max pooling convolution)模块通过最大池化下采样层和卷积标准化激活层,将提取的特征与进行两次卷积标准化激活层提取的特征相拼接,丰富感受野的特征信息。SPPCSPC模块则是利用空间金字塔池化结构(spatial pyramid pooling, SPP)进行改进,通过自底而上的上采样和自顶而下的下采样,从网络模型

的浅层提取高分辨率特征,用于捕捉细粒度语义,这些语义信息被用于检测小目标,网络深层阶段的低分辨率特征被用于检测大型目标^[14]。通过并行分支,卷积核大小分别为 5×5 、 9×9 、 13×13 的最大池化提取不同尺度感受野的特征,进一步与输入特征进行拼接融合,丰富特征信息。经过重参数化结构 Repconv^[15],检测头输出网络模型的预测结果(预测类别、预测框和置信度信息)。

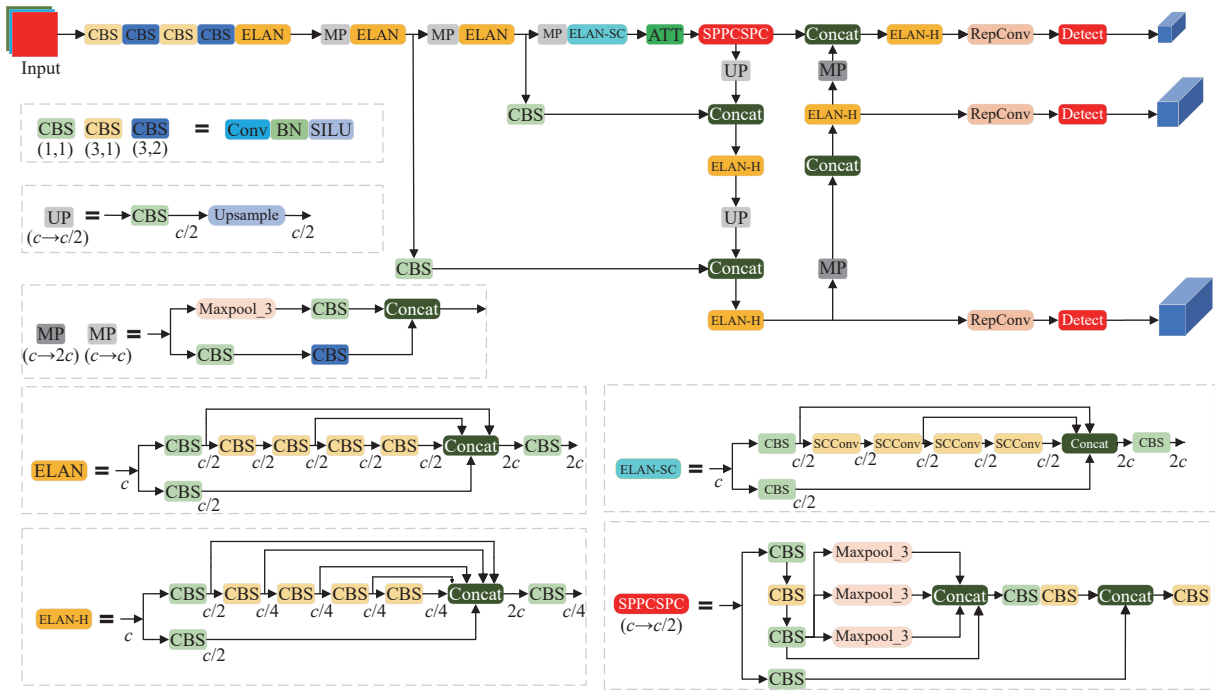


图1 YOLO-SCAT 算法模型图

Fig. 1 Diagram of YOLO-SCAT algorithm model

1.2 空间和通道重建卷积模块

SCConv(spatial and channel reconstruction convolution)是由空间重构单元(SRU)和通道重构单元(CRU)构成的。SRU利用权重分离冗余特征并进行重构,以抑制空间维度上的冗余,并增强特征的代表能力。CRU则采用分割、变换和融合策略,以减少通道维度的冗余,同时降低计算成本和存储需求。SCConv将SRU和CRU按顺序组合,可以替代标准卷积应用于各种主干特征提取网络中,从而减轻计算负担,提高模型在目标检测任务上的性能。将SCConv模块嵌入到不同架构中,能够有效地消除特征冗余,减少模型参数量,并提升特征表达能力。

1.2.1 空间重构单元

如图2所示,空间重构单元(SRU)分为分离和重构两部分。分离的目的是将信息丰富的特征图

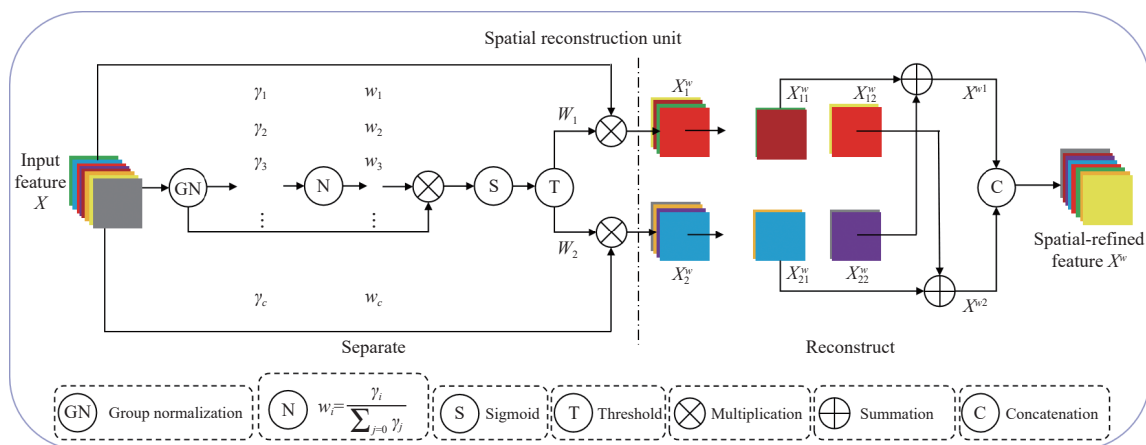
与空间内容对应的信息较少的特征图区别开。利用组归一化(GN)层中的比例因子来评估不同特征图的信息内容。

1) 定义一个中间特征映射 $X \in R^{N \times C \times H \times W}$,其中 N 为批次, C 为通道, H 和 W 为空间高度和宽度,通过减去平均值 μ 并除以标准差 σ 来标准化输入特征 X ,如式(1):

$$X_{\text{out}} = \text{GN}(X) = \gamma \frac{X - \mu}{\sqrt{\sigma^2 + \varepsilon}} + \beta \quad (1)$$

式中: μ 和 σ 是 X 的均值和标准差; ε 是用于保证稳定性的常数; γ 和 β 是可训练的参数变量。

2) 利用GN层中的可训练参数 $\gamma \in R^C$ 来测量每个批次和通道的空间像素方差。 γ 代表像素之间的变化, γ 越大,代表像素变化越大,空间信息越丰富。归一化相关权值 $W_\gamma \in R^C$ 表示不同特征映射的重要性,如式(2):



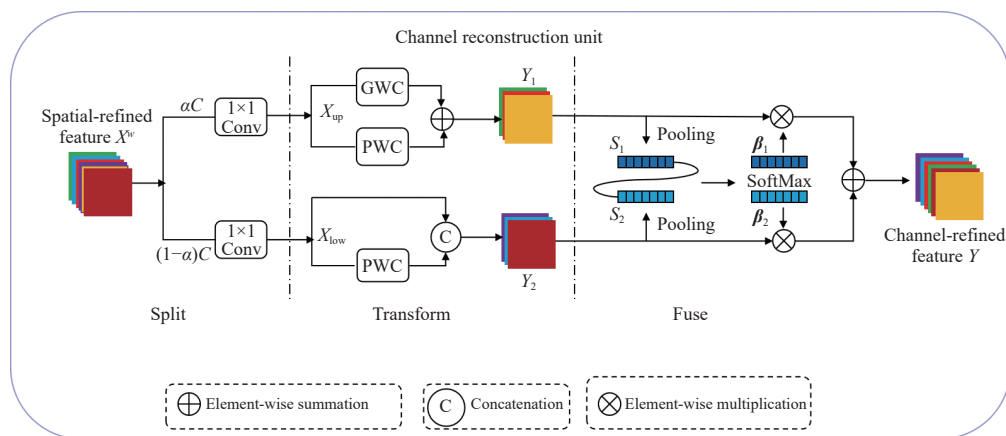


图3 CRU 单元模型图

Fig. 3 Diagram of CRU unit model

和最大池化特征向量。这 2 个池化特征被输入到由多层感知器(MLP)和一个隐藏层组成的共享网络。在该共享网络中,先将特征通道数压缩为原来的 $1/\gamma$,再扩张回原通道数,最后通过元素求和的方式合并输出特征向量。具体的计算过程如式(8):

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) = \sigma(W_1(W_0(F_{\text{avg}}^c)) + W_1(W_0(F_{\text{max}}^c))) \quad (8)$$

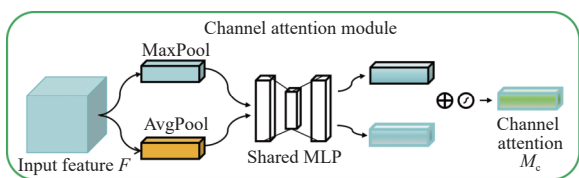


图4 CAM 模块模型图

Fig. 4 Diagram of CAM module model

式中 σ 代表 Sigmoid 函数,两池化特征向量共享 MLP 的权重 $W_0 \in R^{C/\gamma \times C}$, $W_1 \in R^{C \times C/\gamma}$ 。

1.3.2 空间注意力模块

如图 5 所示,空间注意力模块(spatial attention module, SAM)首先对输入特征进行平均池化和最大池化处理,将池化操作得到的特征图在通道维度进行拼接后,生成有效的特征图,最后经过卷积层生成空间注意力图,该图对空间位置特征的重要性进行了编码。通过两个池化操作来聚合特征图的通道信息,生成表示跨通道的平均池化特征和最大池化特征。然后,将两个特征图进行拼接后通过标准卷积层进行卷积,生成二维空间注意力图。具体的计算过程如下:

$$M_s(F) = \sigma(f^{7 \times 7}([\text{AvgPool}(F); \text{MaxPool}(F)])) = \sigma(f^{7 \times 7}([F_{\text{avg}}^s; F_{\text{max}}^s])) \quad (9)$$

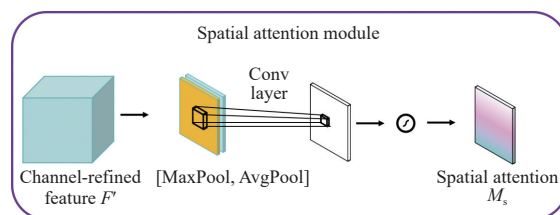


图5 SAM 模块模型图

Fig. 5 Diagram of SAM module model

1.3.3 CBAM 注意力机制

如图 6 所示, CBAM 注意力机制结合了 CAM 和 SAM 的输出,用于加权调整输入特征图。它同时采用全局平均池化和全局最大池化策略,生成最终的注意力增强特征。CBAM 的优势在于同时从通道和空间两个维度上学习到不同维度特征。对于通道注意力,它融合最大池化特征和平均池化特征,增强每个通道的特征表达。对于空间注意力,它分别提取最大池化特征和平均池化特征并进行拼接处理,强调图像重要位置信息。CBAM 通过强调有效特征与抑制无效特征的方式帮助网络模型学习,提高网络模型的代表能力。这使得网络模型更加灵活,能够更好地适应不同的数据分布和任务。

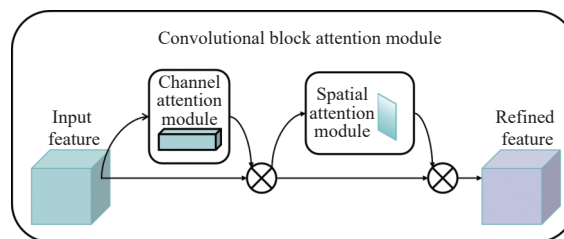


图6 CBAM 模块模型图

Fig. 6 Diagram of CBAM module model

2 数据集及实验评估指标介绍

2.1 数据集

实验选用多光谱探测仪采集无人机可见光视频数据,该设备的可见光采集图像传感器为 1/3 Type CMOS Sensor,采集图像分辨率为 1 920×1 080 像素/30 Hz。视频中无人机种类包含伟力 X1S 无人机、大疆 MINI3 Pro、世翼六旋翼无人机,采集场景包含晴天、多云、傍晚,在建筑楼、树林等特殊复杂环境下进行拍摄,对 12 段视频每隔 2 帧进行提取图片,并剔除纯背景和放大后人眼仍难以识别的图片,得到 500 张包含一个或多个不同远近尺度无人机小目标的图像,使用 labeling 进行数据集的标注,通过编写 Python 程序将所有筛选出的图片按照训练集、测试集、验证集 8 : 1 : 1 的比例进行划分,最终得到 400 张训练集图片、50 张验证集图片、50 张测试集图片。

2.2 实验评估指标

本文以精确率 (Precision, P), 召回率 (Recall, R), F_1 分数 (F_1 -Score, F_1), 平均精度 (average precision, P_{AP}), 均值平均精度 mAP(mean average precision, P_{mAP}) 作为评估模型指标。精确率、召回率、 F_1 分数的计算公式分别为

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \quad (10)$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad (11)$$

$$F_1 = \frac{2 \times P \times R}{P + R} \quad (12)$$

式中: N_{TP} 表示预测为正样本的真实正样本数量; N_{FP} 表示预测为正样本的真实负样本数量; N_{FN} 表示预测为负样本的正样本数量。

AP 和 mAP 表示单个类别和全部类别的平均精度,由于本文研究的检测目标只有单一类别,AP

和 mAP 在数值和意义上相同。实验选用 mAP@0.5 和 mAP@0.5:0.95 作为评估指标,分别表示真实框与预测框交并比 (IOU) 阈值为 0.5 时的 mAP 值和在不同的交并比阈值 (从 0.5 到 0.95, 步长为 0.05) 上的平均 mAP。计算公式为

$$P_{AP} = \int_0^1 P(R) dR \quad (13)$$

$$P_{mAP} = \frac{1}{C} \sum_{i=1}^C P_{AP_i} \quad (14)$$

$$mAP@0.5:0.95 = \frac{1}{9} \sum_{IOU=0.5}^{0.95} P_{mAP_{IOU}} \quad (15)$$

式中: P 表示精确率; R 表示召回率; C 表示目标类别总数; P_{AP_i} 表示第 i 类目标 AP 值; $P_{mAP_{IOU}}$ 表示该交并比阈值下的 mAP 值。

2.3 实验环境及参数设置

本实验使用 Ubuntu 18.04 操作系统的服务器平台,搭载 Intel Xeon Gold 5120 CPU @ 2.2 GHz 和 NVIDIA GeForce RTX 2080ti。所采用的深度学习算法基于 Python 3.10 编程语言,利用 PyTorch 1.12 框架构建。实验中,模型的训练设置为 300 轮, batch size 为 12, 图片大小为 640×640 像素。优化器选择随机梯度下降法 (stochastic gradient descent, SGD), 初始学习率为 0.01, 动量设为 0.937, 衰减系数为 0.0005。为提高模型的鲁棒性,实验在训练前采用 Mosaic 数据增强方法对数据集图像进行处理。

3 实验及结果分析

3.1 实验结果与分析

为体现 YOLO-SCAT 对于小目标无人机检测的优越性能,将改进后的 YOLO-SC、YOLO-AT、YOLO-SCAT 与 YOLOv7 原版进行对比,如表 1 所示。

表 1 YOLO-SC、YOLO-AT、YOLO-SCAT 和 YOLOv7 在无人机数据集整体性能对比

Table 1 Comparison of overall performance between YOLO-SC, YOLO-AT, YOLO-SCAT and YOLOv7 on unmanned aerial vehicle datasets

Model	$P/\%$	$R/\%$	mAP@0.5/%	mAP@0.5:0.95/%	$F_1/\%$	FPS	Gflops
YOLOv7	93.9	86.1	90.4	50.2	89.8	80.6	103.2
YOLO-SC	95.4	86.1	92.2	51.0	90.5	77.5	103.1
YOLO-AT(CBAM)	93.8	88.9	93.1	51.6	91.3	80.0	105.3
YOLO-SCAT	94.4	94.4	94.7	52.9	94.4	76.9	103.8

表 1 展示了在 UAV 数据集上的实验结果, YOLOv7 表现出卓越的目标检测性能。改进后的

YOLO-SCAT 在可见光条件下对无人机小目标的检测性能进一步提升。YOLO-SC 为 YOLOv7 引

入 ELAN-SC 模块进行改进后的目标检测网络,对比 YOLOv7 的 P (精确率)提升 1.5%, R (召回率)保持不变, $mAP@0.5$ 、 $mAP@0.5:0.95$ 以及 F_1 分数分别有所提升,仍有改进空间。YOLO-AT 为 YOLOv7 引入注意力机制 CBAM 后的网络结构,相比于 YOLOv7, P 降低 0.1%, R 提升 2.8%, $mAP@0.5$ 、 $mAP@0.5:0.95$ 、 F_1 均有所提升。YOLO-SCAT 的 P 、 R 、 $mAP@0.5$ 、 $mAP@0.5:0.95$ 以及 F_1 分数分别为 94.4%、94.4%、94.7%、52.9%、94.4%。相较于 YOLOv7,改进后的算法分别提升了 0.5%、8.3%、4.3%、2.7%、4.6%,充分证实 YOLO-SCAT 在无人机微小目标可见光图像检测任务中的卓越性能。虽然引

入 SCConv 卷积和注意力机制产生了轻微的检测效率损失,略微增加了模型的检测推理时间,但 76.9 frame/s 的性能完全满足实时视频检测的需求,保障了一阶段目标检测算法实时检测目标的天然优势。

3.1.1 主干特征提取网络对比实验 (YOLO-SC)

为解决在执行模型缩放时深度模型的收敛性会逐渐恶化的问题, YOLOv7 中作者设计了 ELAN 结构,但其使用大量卷积模块,会出现特征冗余现象,所以 SCConv 的引入能够充分发挥 SCConv 减少特征冗余的作用,提出了几个引入 SCConv 的 ELAN 改进结构,如图 7 中 a、b、c、d 结构所示。

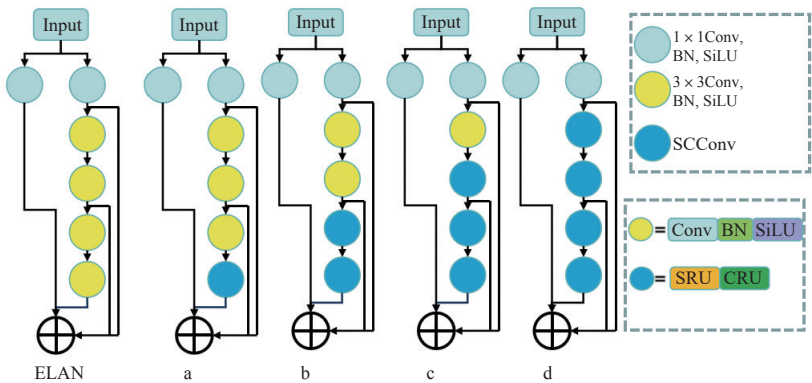


图 7 ELAN 模块改进示意图

Fig. 7 Schematic diagram of ELAN module improvement

从图 7 可见, 4 种结构中卷积数量和张量拼接方式与 ELAN 相同,但它们在第 2 分支经历的 4 次卷积各不相同。在 a、b、c、d 4 种结构的特征处理过程中,输入特征的尺寸保持不变。两个分支都首先经过 1×1 的普通卷积,第 1 分支用以保留原始信息,而第 2 分支则通过 4 次卷积操作来提取关键特征。这使得第 2 分支能够将原始信息、第 2 次和第 4 次卷积提取的特征进行整合,并与第 1 分支保留的原始信息进行 concat 拼接,以更有效地捕获目标特征。

在 a 结构中,前 3 次卷积使用 3×3 的普通卷积,第 4 次则采用 SCConv 模块。在 b 结构中,前两次卷积为普通 3×3 卷积,后两次使用 SCConv 模块。c 结构的第 1 次卷积为普通 3×3 卷积,其余 3 次采用 SCConv 模块。d 结构则全程使用 SCConv 模块进行卷积。

为说明引入 SCConv 的优势,在颈部网络和检测头的网络结构和损失函数等设置完全相同的情况下,在主干特征提取网络的 ELAN 结构中,对上述改进结构进行实验,结果如表 2 所示。

表 2 ELAN 模块改进对比实验结果

Table 2 Comparative experimental results of ELAN module improvement

Backbone	$P/\%$	$R/\%$	$mAP@0.5/\%$	$mAP@0.5:0.95/\%$	$F_1/\%$	FPS
YOLOv7	93.9	86.1	90.4	50.2	89.8	80.6
a	92.6	87.5	90.7	46.4	90.0	79.4
b	92.6	87.4	91.3	47.8	89.9	78.7
c	96.9	87.4	90.9	49.8	94.1	78.1
d	95.4	86.1	92.2	51.0	90.5	77.5

表2的实验结果表明,主干特征提取采用d结构后的网络在精确率、召回率、 F_1 分数、mAP性能指标上综合表现最好。上述改进的4种结构mAP@0.5数据均高于YOLOv7,其中c、d结构对比原版YOLOv7,精确率分别有所提高,在mAP@0.5:0.95(%)一项中d结构高于YOLOv7,证明在可见光图像的无人机小目标检测任务中,d结构的性能优于其他3种ELAN结构,将d结构命名为ELAN-SC。主干特征提取网络采用ELAN-SC模块,减少了特征的空间和通道冗余,有效增强网络对可见光无人机目标识别能力。

3.1.2 注意力机制对比实验

在YOLO-SC引入注意力机制之前,进行了原版YOLOv7直接引入注意力机制的实验,实验结果如表3所示,发现YOLOv7引入注意力机制后,精确率略微降低,召回率大幅提升,模型平均精度

得到显著提升,因此在ELAN-SC的网络模型基础上引入注意力机制进行模型的改进。

表3 原版YOLOv7引入注意力机制实验

Table 3 Introduction of attention mechanism experiment in original YOLOv7 %

Attention	P	R	mAP@0.5
Original	93.9	86.1	90.4
SE	92.9	90.3	93.0
ECA	85.7	91.7	92.4
EMA	91.8	93.0	93.0
CBAM	93.8	88.9	93.1

为验证引入注意力机制CBAM在UAV目标数据集上的检测性能的提升,在YOLO-SC算法模型下,对CBAM与SE、ECA、EMA注意力机制进行对比实验,实验结果见表4。

表4 注意力机制对比实验结果

Table 4 Comparative experimental results of attention mechanism %

Attention	P	R	mAP@0.5	mAP@0.5:0.95	F_1
SE	90.7	94.4	93.6	51.3	92.5
ECA	94.4	94.4	94.5	50.2	94.4
EMA	91.2	81.6	91.2	47.8	86.1
CBAM	94.4	94.4	94.7	52.9	94.4

表4中的实验结果表明YOLO-SC引入EMA注意力机制后,检测效果下降,而其他3种注意力机制均有所提升,其中CBAM在精确率、召回率、平均精度等指标均优于其他3种注意力机制,同时 F_1 分数与ECA持平。相对于直接引入注意力机制,CBAM与SCConv因为在空间和通道维度均进行了小目标特征提取整合的操作,反而更加契合,实验效果更加优异,SE、ECA注意力机制作为典型的通道注意力机制,对于空间维度特征没有充分利用,EMA注意力机制则更加注重空间维度特征学习,CBAM注意力机制的引入,充分利用SCConv模块通过减少空间和通道特征冗余,重构特征向量,对无人机小目标特征进行更加充分的小目标特征学习。

3.2 YOLO-SCAT 算法检测结果可视化

本文提出的YOLO-SCAT算法以YOLOv7为基础,相比于YOLOv7,对可见光无人机小目标检测效果有比较显著的提升,图8选取几类检测难度比较大的无人机场景对比两种算法的实际效果。

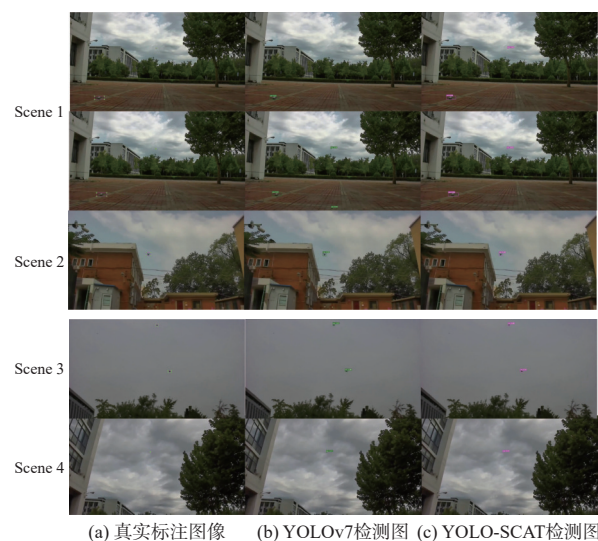


图8 算法检测结果图

Fig. 8 Diagrams of algorithm detection results

如图8所示,从左到右分别为真实标注图像、YOLOv7检测图和YOLO-SCAT检测图。在场景1、场景2中,检测难点分别为小目标云层背景、植

被环境和复杂建筑背景。在这些环境中, YOLOv7 算法表现出漏检、误检和检测平均精度不高的情况。相比之下, YOLO-SCAT 算法能够减少冗余特征, 有效地从背景中提取目标特征, 成功识别出无人机目标。对于场景 3 和场景 4, 检测难点分别为多云低光照环境和运动模糊环境。YOLOv7 算法由于无法准确识别目标特征而导致检测平均精度不高。然而, YOLO-SCAT 算法通过抑制背景信息干扰, 能有效提高检测平均精度。在复杂背景、低光照环境和运动模糊等造成特征模糊的干扰情况下, 面对不同型号无人机小目标, YOLO-SCAT 表现出更强的鲁棒性和更高的检测平均精度。

3.3 YOLO-SCAT 算法泛化性

为了验证 YOLO-SCAT 算法良好的泛化性, 选取西湖大学的公开 ARD-MAV 数据集进行验证, 数据集包含了复杂背景、非平面场景、目标遮挡、相机剧烈运动、运动模糊、小目标无人机等多种挑战场景, 图片分辨率为 1920×1080 像素, 目标平均尺寸仅为图像尺寸的 0.02%。通过对比原版 YOLOv7 与改进后的算法 YOLO-SCAT 的检测结果, 来体现 YOLO-SCAT 的算法优越性, 实验训练参数与 2.3 节保持一致, 结果如表 5 所示。

表 5 YOLO-SCAT 泛化性验证实验

Table 5 YOLO-SCAT generalization validation experiment

Model	P/%	R/%	mAP@0.5/%	F_1 /%	FPS
YOLOv7	92.3	72.7	90.9	81.3	80.6
YOLO-SCAT	93.9	93.9	95.0	93.9	76.9

如表 5 所示, YOLO-SCAT 在公开数据集 ARD-

MAV 上进行检测任务, 精确率、召回率、mAP@0.5 和 F_1 分数上均高于原版 YOLOv7, 分别提升了 1.3%、21.2%、4.1% 和 12.56%。网络模型具有良好的泛化能力, 该模型在其他数据集上也可以保持较好的准确性和可靠性, 上述实验充分证明了 YOLO-SCAT 具有优秀的泛化能力, 适用于复杂场景下可见光无人机小目标的检测任务。

3.4 YOLO-SCAT 算法对比其他常见目标检测算法

为进一步验证 YOLO-SCAT 算法的性能, 选取目前目标检测领域中几种主流的算法进行对比, 利用 2.1 节无人机小目标数据集进行实验, 在训练过程中, 为保证一致性, 除 Faster-RCNN 的 batchsize 设置为 6, 其余网络将 batchsize 统一设置为 12, epoch 设置为 300 轮。

如表 6 所示, 两阶段目标检测算法 Faster-RCNN^[16] 推理速度慢, 在精确率、召回率、mAP@0.5 和 F_1 分数指标也处于较低水平, 不适合进行小目标的实时检测任务。一阶段的目标检测算法基本符合实时检测的需求。SSD 网络^[17] 作为早期的一阶段检测算法, 召回率较低, 检测性能较低。YOLOX^[10] 作为无锚框的检测算法, 相对于基于锚框检测的 YOLOv5^[18]、YOLOv7、YOLO-SCAT 目标检测算法, 虽然精确率相差不大, 但召回率相对较低。本文提出的 YOLO-SCAT 目标检测算法相较于目标检测中的主流网络, 在精确率、召回率、mAP@0.5 和 F_1 分数进行提升, 可以更好地面对各种复杂场景造成的干扰, 同时因受益于 YOLOv7 良好的实时检测性能, 改进后的算法的实时检测性能也能满足反无人机系统中的需求。

表 6 YOLO-SCAT 与主流目标检测算法的对比实验结果

Table 6 Comparative experimental results of YOLO-SCAT and mainstream target detection algorithms

Model	P/%	R/%	mAP@0.5/%	F_1 /%	FPS
Faster-RCNN	28.8	73.9	54.4	41.0	18.2
SSD	86.9	67.8	70.8	76.1	43.8
YOLOX	86.7	66.1	70.9	75.0	40.7
YOLOv5	84.0	86.4	86.1	85.1	40.0
YOLOv7	93.9	86.1	90.4	89.8	80.6
Ours	94.4	94.4	94.7	94.4	76.9

4 结论

本文针对微小低空飞行无人机目标可见光检测场景存在现有检测算法平均精度不高、误检、漏检频繁等问题, 提出了 YOLO-SCAT 网络模型。首先通过在高效层聚合模块 ELAN 引入 SCConv 模块, 减少由在空间和通道中冗余特征提取产生的

计算成本, 提高无人机小目标的特征提取效率, 然后在 YOLO-SC 模型基础上, 引入注意力机制 CBAM, 进一步在空间和通道信息中提取有效特征信息, 提升了网络模型的平均精度, 最后实验结果证明, 在可见光图像无人机小目标数据集中, 相比于目前一些主流的目标检测算法模型, YOLO-SCAT 对

无人机小目标的检测具有一定的优势,实时检测性能符合实际要求,同时在多云低光照环境、植被环境、复杂建筑环境下减少了误检、漏检问题的发生,提高了检测算法的平均精度。本文未来将从扩充数据集、小目标检测等领域对无人机目标检测任务开展研究和实验,从增加恶劣极端天气下无人机的可见光图像和增加光电探测无人机的检测范围入手,进一步提高微小无人机目标检测系统的实用性。

参考文献:

- [1] 罗俊海,王芝燕. 无人机探测与对抗技术发展及应用综述[J]. 控制与决策, 2022, 37(3): 530-544.
LUO Junhai, WANG Zhiyan. A review of development and application of UAV detection and counter technology[J]. Control and Decision, 2022, 37(3): 530-544.
- [2] 亓贺. 光电图像制导系统中目标跟踪关键技术研究[D]. 北京: 北京理工大学, 2016: 5-8.
QI He. Research on key technologies of target tracking in optoelectronic image guidance systems[D]. Beijing: Beijing Institute of Technology, 2016: 5-8.
- [3] ZOU Z X, SHI Z W, GUO Y H, et al. Object detection in 20 years: a survey[J]. *Proceedings of IEEE*, 2023, 111(3): 257-276.
- [4] 王晓. 视频中无人机的实时检测与跟踪算法研究[D]. 哈尔滨: 哈尔滨工程大学, 2019: 25-47.
WANG Xiao. Research on real time detection and tracking algorithms for drones in videos[D]. Harbin: Harbin Engineering University, 2019: 25-47.
- [5] HOU X D, ZHANG L Q. Saliency detection: a spectral residual approach[C]//IEEE Conference on Computer Vision & Pattern Recognition. Minneapolis. New York: IEEE, 2007: 1-8.
- [6] 邵盼瑜. 基于视觉的无人机入侵检测与跟踪系统设计与实现[D]. 杭州: 浙江大学, 2018: 22-26.
SHAO Panyu. Design and implementation of a visual based unmanned aerial vehicle intrusion detection and tracking system[D]. Hangzhou: Zhejiang University, 2018: 22-26.
- [7] 陶磊,洪韬,钞旭. 基于 YOLOv3 的无人机识别与定位追踪[J]. 工程科学学报, 2020, 42(4): 463-468.
TAO Lei, HONG Tao, CHAO Xu. Drone recognition and localization tracking based on YOLOv3[J]. Chinese Journal of Engineering, 2020, 42(4): 463-468.
- [8] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL]. [2024-01-28]. <https://pjreddie.com/media/files/papers/YOLOv3.pdf>.
- [9] 薛珊,王亚博,吕琼莹,等. 基于 YOLOX-drone 的反无人机系统抗遮挡目标检测算法[J]. 工程科学学报, 2023, 45(9): 1539-1549.
XUE Shan, WANG Yabo, LYU Qiongying, et al. An anti obstruction target detection algorithm for anti UAV systems based on YOLOX drone[J]. Chinese Journal of Engineering, 2023, 45(9): 1539-1549.
- [10] GE Z, LIU S T, WANG F, et al. YOLOX: exceeding YOLO series in 2021[EB/OL]. (2021-08-06)[2024-01-28]. <https://arxiv.org/abs/2107.08430>.
- [11] WANG C Y, BOCHKOVSKIY A, LIAO M H Y. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, BC, Canada: IEEE, 2023: 7464-7475.
- [12] LI J F, WEN Y, HE L H. SCConv: spatial and channel reconstruction convolution for feature redundancy[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, BC, Canada: IEEE, 2023: 6153-6162.
- [13] SANGHYUN W, JONGCHAN P, JOON Y L, et al. Cbam: convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany: Springer, 2018: 3-19.
- [14] CHEN Y M, YUAN X B, WU R Q, et al. YOLO-MS: rethinking multi-scale representation learning for real-time object detection[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, BC, Canada: IEEE, 2023: 2132-2143.
- [15] DING X H, ZHANG X Y, MA N N, et al. RepVGG: making VGG-style ConvNets great again[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA: IEEE, 2021: 13728-13737.
- [16] 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.
- [17] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Proceedings of the European Conference on Computer Vision(ECCV). Amsterdam: Springer, 2016: 21-37.
- [18] ULTRALYTICS. YOLOV5[EB/OL]. [2023-8-2]. <https://github.com/ultralytics/yolov5>.