

DOI:10.19651/j.cnki.emt.2211032

基于感受野增强与并行坐标注意力的 航拍小目标检测*

章曙光^{1,2} 邵政瑞¹ 鲍锐¹ 唐锐¹

(1.安徽建筑大学电子与信息工程学院 合肥 230601; 2.安徽建筑大学信息中心 合肥 230601)

摘要: 针对航拍图像小目标占比多、背景复杂、存在检测精度低的问题。提出一种基于感受野增强与并行坐标注意力的航拍小目标检测算法,设计一种感受野增强模块,使用不同大小的空洞卷积扩大感受野范围并融合有效通道注意力机制来提高网络的特征提取能力;改进特征融合结构来提高算法对小目标的检测能力;设计一种并行坐标注意力模块用以提高航拍密集小目标检测以及抗背景干扰能力。采用不同输入分辨率在 VisDrone 数据集上进行实验,实验结果表明,所提算法的 mAP0.5 相较于 YOLOv5 算法提高了 5.4%,mAP0.5:0.95 提高了 4.2%,在输入分辨率 1 536×1 536 下的 mAP0.5 可达 54.9%,可以实现较好的小目标检测效果。

关键词: 小目标检测;感受野增强;并行坐标注意力;航拍图像

中图分类号: TP391.4 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Aerial small object detection based on receptive field enhancement and parallel coordinate attention

Zhang Shuguang^{1,2} Shao Zhengrui¹ Bao Rui¹ Tang Rui¹

(1. School of Electronics and Information Engineering, Anhui Jianzhu University, Hefei 230601, China;

2. Information Network Center, Anhui Jianzhu University, Hefei 230601, China)

Abstract: This paper aims at the problems of small target proportion, complex background and low detection accuracy in aerial images. Proposes a small aerial target detection algorithm based on receptive field enhancement and parallel coordinate attention. A receptive field enhancement module is designed to expand the receptive field range by using cavity convolution of different sizes and integrate effective channel attention mechanism to improve the feature extraction ability of the network. The feature fusion structure is improved to improve the detection ability of small targets. A parallel coordinate attention module is designed to improve the ability of aerial photography dense small target detection and anti-background interference. Experiments are conducted on VisDrone dataset with different input resolutions. The experimental results show that mAP0.5 and mAP0.5:0.95 of the proposed algorithm are improved by 5.4% and 4.2% compared with YOLOv5 algorithm, and mAP0.5 can reach 54.9% at input resolution 1 536×1 536. It can achieve better effect of small target detection.

Keywords: small object detection;receptive field enhancement;parallel coordinate attention;aerial images

0 引言

随着航空遥感技术的发展,无人机因其具有高机动性,拥有广阔的拍摄视野,被广泛应用于交通监测^[1]、电力巡检^[2]、遥感测绘^[3]等多个领域。计算机视觉技术与无人机结合使得无人机的应用场景得到广泛拓展。

传统的目标检测算法通常使用 HOG、SIFT、Haar 等人工设计的特征进行目标检测,但检测精度低、效果不好。随着深度学习的发展,使用卷积神经网络直接提取图像中的特征信息并将其应用在目标检测中,极大地提升了目标检测精度。虽然各类目标检测算法已经取得了不错的检测成果,但将这些算法直接应用于检测航拍图像小目标仍存

收稿日期:2022-08-10

* 基金项目:安徽省高校协同创新项目(GXXT-2021-093)、安徽省重点研究与开发计划项目(202004a07020050)、基于国产智能芯片的宿舍安全管理平台应用开发服务(HYB20200190)项目资助

在许多不足。由此,一些学者更加关注航拍图像小目标检测问题,提出了许多有效的解决办法。DetNet^[4]通过保持深层网络高的空间分辨率和利用空洞卷积扩大感受野来提高小目标检测的性能。文献[5]提出的特征金字塔网络(FPN)和文献[6]中提出的(BiFPN)都采用了多尺度特征学习的思想,通过多尺度融合深层次和浅层次的特征来提高小目标的检测性能。文献[7]在YOLOv3^[8]网络的基础上,通过建立密集连接和添加CBAM^[9]注意力机制来增强局部和全局的特征信息,提高小尺寸目标的检测精度,文献[10]提出一种基于特征金字塔网络(FPN)的改进算法,在原有算法上增加并行分支,再融合两种不同上采样方法的特征信息来加强小目标的表达能力。文献[11]使用平均池化操作添加反卷积模块的额外缩放分支以形成特征金字塔,并融入上下文分析来提高小目标检测精度。

基于上述针对航拍图像目标难以检测的问题,本文选取YOLOv5^[12]作为模型基准,提出一种基于感受野增强与并行坐标注意力的航拍小目标检测算法。本文设计一种感受野增强模块(receptive field enhancement, RFE)来提高网络的特征提取能力,通过RFB^[13]模块扩大感受野的同时结合有效通道注意力机制^[14](effective channel attention, ECA)来提高网络对分支通道的关注度,使网络集中于检测目标的有效特征,降低无用特征的干扰。通过增加特征金

字塔层级覆盖更小的目标尺度的方式改进特征融合结构,设计一种并行注意力模块(parallel coordinate attention, PCA),加强网络模型对深层次小目标特征的挖掘能力,有效解决航拍密集小目标遮挡和漏检问题,进一步提高模型性能。

1 本文方法

本文将详细介绍基于感受野增强与并行坐标注意力的航拍小目标检测算法,网络结构如图1所示。为了解决特征信息丢失,提高小目标的表达能力,本文设计了一种由RFB模块和有效通道注意力机制模块(ECA)结合的感受野增强模块(RFE)。用RFE模块替换原始YOLOv5中SPP模块,RFE模块先利用空洞卷积在提升感受野的同时不损失信息的特点来扩大感受野,之后对各个分支不同感受野的不同通道按照重要程度分配不同权重使主干网络集中于提取小目标的有效特征。改进多尺度特征融合结构即将特征金字塔由原始的3层结构和连接3个尺度特征扩展为4层结构和连接4个尺度特征,将增加的特征层与主干网络中经过四倍下采样且位置信息丰富的特征层相拼接,以保留更多小目标特征的特征图,最后通过并行注意力模块,加强网络模型在特征融合阶段对密集小目标的关注度,进一步提高模型精度,下面介绍这3个改进的具体实现。

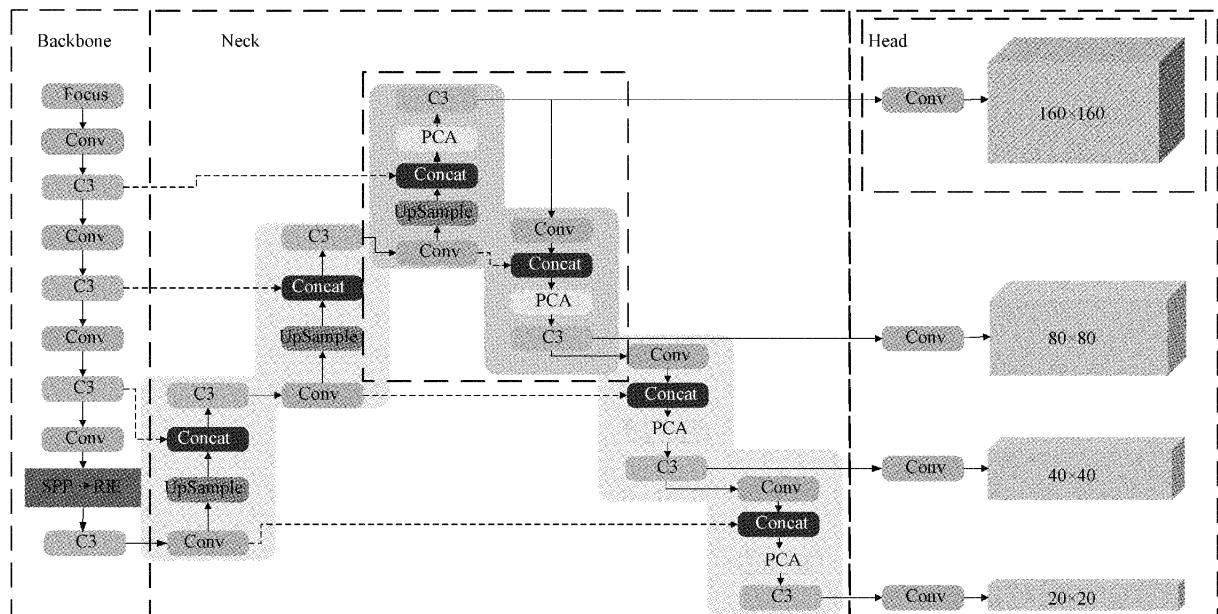


图1 基于感受野增强与并行坐标注意力的航拍小目标检测算法网络结构

1.1 感受野增强模块

在航拍目标检测中,需要高质量的细节特征来加强对航拍小目标特征的提取。原有SPP结构中使用池化核为 $5 \times 5, 9 \times 9, 13 \times 13$ 的最大池化虽然学习到不同尺度的特征,丰富了特征图的表达能力。但是池化操作会削弱特征图的信息,尤其是在解决小目标检测问题中,SPP结构中的池化操作为进一步削弱航拍图像小目标的特征信息,不

利于航拍图像小目标的检测。

本文结合RFB模块和ECA模块,设计了感受野增强模块(RFE)来提升特征提取效果,如图2所示。RFE保留了原始RFB模块中 $1 \times 1, 3 \times 3, 5 \times 5$ 的卷积核以及间隔跨度为1,3,5的3种空洞卷积,不同的是将有效通道注意力机制分别放在不同空洞卷积尺度的特征上,以便更多地关注特征语义信息,最后再对他们的输出进行拼接,完成融

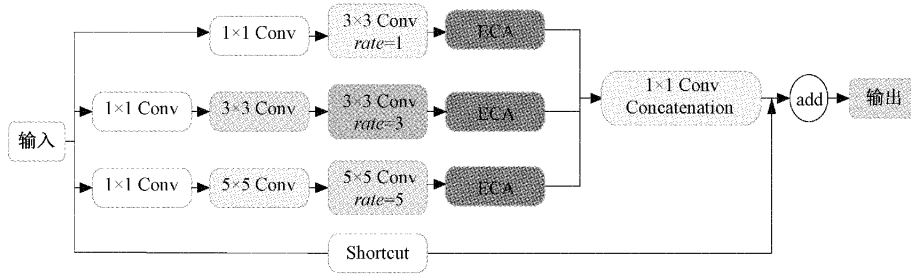


图 2 感受野增强模块

合不同特征的目的,增强网络特征提取的能力。

RFB 模块模拟人类视觉感受野,借鉴 Inception 网络^[15]以及空洞卷积思想,通过使用不同卷积核大小的卷积层来模拟感受野,不同大小的空洞卷积来模仿人眼视觉系统中感受野与偏心度的关联度,有效地增大了感受野,虽然空洞卷积解决标准卷积造成的内部数据结构丢失与空间层信息丢失的问题,随着网络层数的加深,含有目标特征的通道对目标语义特征表达更有益,不含目标特征的通道则会对语义特征进行干扰。因此本文在各分支加入有效通道注意力机制来增强小目标语义特征的表达。

ECA 提出了一种不降维的局部跨信道交互策略和自适应选择一维卷积核大小的方法。该方法只增加了少量的参数,却能获得明显的性能增益。因此,本文将 ECA 放在 RFB 模块的 3 个分支之后来减少各个分支特征提取过程中上下文信息的缺失。有效通道注意力模块示意图如图 3 所示。首先对输入图像数据进行一次全局平均池化来获得图像聚合的特征,再执行 $K=5$ 的 1 维卷积生成的通道权重。

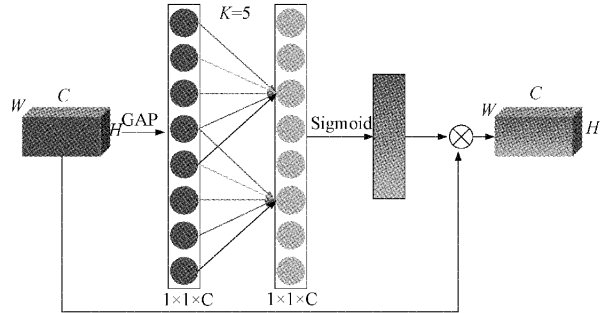


图 3 有效通道注意力网络结构

1.2 并行坐标注意力模块

注意力机制可以有效提高基于卷积神经网络任务的性能,能筛选出感兴趣的区域,让网络模型只关注对检测结果提升有益的部分。本文基于坐标注意力机制^[16]提出并行坐标注意力模块(PCA)。CA 注意力模块利用自适应平均池化方法增强了特征图中间区域的关注度,但航拍图像目标在复杂背景和小目标密集区域的信息也需要分配更多的注意力,故本文提出了一种结合自适应平均和自适应最大值池化的 PCA 模块,如图 4 所示。

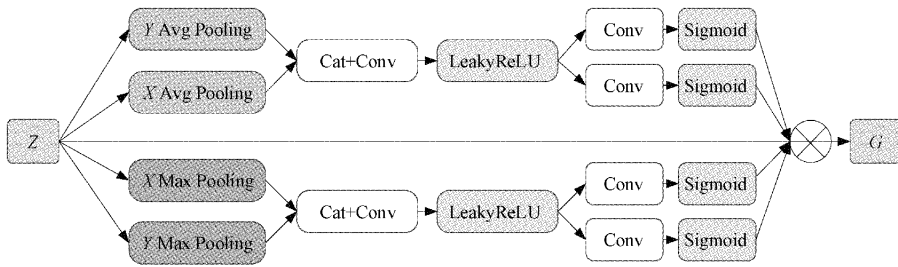


图 4 并行坐标注意力模块

PCA 注意力模块首先对输入 Z 进行两个并行的自适应池化操作并融合,然后将提取出来的包含坐标注意力的特征向量与原输入相乘,实现目标的定位与融合。并行的池化方法分为水平方向和垂直方向。在水平和垂直方向得到的特征图进行转置拼接和 1×1 卷积操、非线性激活操作。

以自适应最大值池化为例,首先对输入特征图的水平方向和垂直方向进行自适应最大值池化,使用两个空间上的池化核 $(H, 1)$ 、 $(1, W)$ 分别沿横坐标和纵坐标对每个通

道进行编码,得到一对大小为 $C \times 1 \times W$ 和 $C \times 1 \times H$ 的特征图,如式(1)、(2)所示。

$$g_c^h = \frac{1}{W} \left(\sum_{0 \leq i < W} X_c(h, i) \right) \tag{1}$$

$$g_c^w = \frac{1}{H} \left(\sum_{0 \leq j < H} X_c(j, w) \right) \tag{2}$$

式中: g_c^h 是高度为 h 的第 c 通道的输出, g_c^w 是宽度为 w 的第 c 通道的输出。然后将 g_c^h 和 g_c^w 进行合并操作,并通过一个 1×1 操作得到 F_{max} ,如式(3)所示。

$$F_{\max} = \zeta(\text{Conv}(\text{cat}(\text{max}g^h, \text{max}g^w))) \quad (3)$$

式中: ζ 代表 LeakyReLU 激活函数。Conv 代表 1×1 卷积, $\text{max}g^h$ 和 $\text{max}g^w$ 表示分别沿水平和垂直方向进行自适应最大值池化后得到的特征图, $\text{cat}()$ 表示拼接操作。

然后输入 F_{\max} 沿空间维度划分成两个独立的特征向量 F_{\max}^h, F_{\max}^w , 分别经过两个 1×1 卷积层将 F_{\max}^h 和 F_{\max}^w 转换成与输入 X 具有相同通道数量的特征向量, 如式(4)、(5)所示。

$$l^h = \sigma(\text{Conv}(F_{\max}^h)) \quad (4)$$

$$l^w = \sigma(\text{Conv}(F_{\max}^w)) \quad (5)$$

式中: σ 表示 Sigmoid 函数。输出的 l^h 和 l^w 为自适应最大值池化后的注意力权重。同理可得自适应平均池化后的输出 e^h, e^w 如式(6)、(7)所示。

$$e^h = \sigma(\text{Conv}(F_{\text{avg}}^h)) \quad (6)$$

$$e^w = \sigma(\text{Conv}(F_{\text{avg}}^w)) \quad (7)$$

最后, 经过 PCA 注意力模块的第 c 通道的输出 $O_c(i, j)$ 如式(8)所示。

$$O_c(i, j) = X(i, j) \times l_c^h(i) \times l_c^w(j) \times e_c^h(i) \times e_c^w(j) \quad (8)$$

PCA 通过两种不同的信息聚合方法充分利用捕获的位置信息, 以便准确找到复杂背景下被遮挡或密集情况下的小目标。本文将加入颈部特征融合阶段 C3 之前, 使得 PCA 能够充分获取不同尺度的特征融合信息, 通过对不同方向注意图的施加, 提升网络在复杂背景下对被遮挡和密集小目标的识别效果。

1.3 多尺度特征融合结构改进

在原始 YOLOv5 骨干网络中, 当输入图像尺寸为 640×640 时, 颈部特征融合网络将分别进行 8 倍、16 倍以及 32 倍下采样, 对应检测特征图的尺寸为 80×80 、 40×40 、 20×20 , 以检测小目标、中目标以及大目标。由于航拍图像拍摄场高度较高, 映射到图像中的各类目标也有不同大小的尺度, 尤其是包含多个小目标的密集图像。为了应对复杂密集的航拍场景, 本文改进原始 YOLOv5 算法的特征融合网络结构, 将原始三层特征金字塔结构扩展为四层特征金字塔结构, 改进部分在图 1 中用黑色虚线框出, 具体如图 5 所示。将从主干网络中提取的深层特征图在经过三层 FPN 网络中进行上采样之后继续进行一次上采样操作, 并与主干网络对应尺度浅层特征图进行 Concat 操作; 之后再与 FPN 网络融合后的特征图作为输入送入 PAN 网络, 利用 PAN 网络将 FPN 网络融合后的特征图再次自底向上融合并进行下采样操作。与原始特征融合结构相比, 改进后的特征融合结构通过加深网络的层数使得改进后的多尺度特征融合结构不仅可以学习到更多的深层语义特征信息还可以从浅特征层中学习较强的位置特征, 增强模型在密集目标下多尺度学习的能力, 提升检测小目标的检测效果。

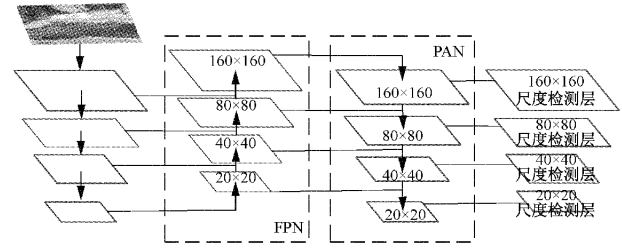


图5 多尺度特征融合结构改进图

2 实验结果与分析

本文实验所采用的数据集来源于 VisDrone 数据集。VisDrone 数据集由 288 个视频剪辑而成, 总计提供了由不同高度的无人机捕获的 10 209 幅静态图像, 其中包括 6 471 幅训练集图像、548 幅验证集图像及 3 190 幅测试集图像, 共计 260 万个目标实例样本。

本文实验的显卡是 Nvidia V100 GPU, 操作系统是 Windows10, 64 位。实验环境为 Pytorch1.7, Python3.8, CUDA10.2, 在相同超参数下进行训练、验证和测试, 其中初始学习率为 0.01, warmup epochs 为 3, 训练 150 个 epoch。选取 mAP0.5、mAP0.5:0.95、准确率 (Precision, P)、召回率 (Recall, R) 作为评价算法有效性的指标。

2.1 消融实验

为了验证本算法提出的各个模块的有效性, 本文进行消融实验评估不同模块在相同分辨率 (640×640) 条件下的检测性能, 具体结果如表 1 所示。表 1 第 1 行为原始 YOLOv5s 模型。第 2 行添加了感受野增强模块 (RFE), 从表 1 中可以看出添加 RFE 相较于基准算法在 mAP0.5 指标上提升了 1.1%, 在 mAP0.5:0.95 指标提升了 0.7%, 可以得出通过扩大感受野同时结合有效注意力机制可以提高网络的特征提取能力, 提高网络对小目标的关注度, 使网络集中于提取小目标的有效特征。第 3 行是进行多尺度特征融合结构改进 (multi-scale feature fusion structure, MFSS), 可见 mAP0.5 指标和 mAP0.5:0.95 指标增长最多, 分别增长了 3.2% 和 2.6%, 可以得出通过改进特征融合网络结构可以充分利用浅层特征信息, 减少了卷积级过程小目标区域特征的信息丢失。第 4 行实验中通过在颈部特征融合阶段增加 PCA 模块, mAP0.5 指标增长了 1.1%, mAP0.5:0.95 指标增加了 0.9%。可以得出在颈部嵌入不同尺度的特征信息, 以一种近似于坐标的形式决定图像中小目标的关注程度, 能够有效提高模型对于复杂背景下密集目标的敏感程度, 进而改善检测任务中目标密集、被遮挡对检测带来的负面影响。综上所述, 本文所提算法与原始 YOLOv5s 在 mAP0.5 指标上提升了 5.4%, 在 mAP0.5:0.95 指标上提升了 4.2%。

2.2 对比实验

表 2 给出了 YOLOv3、Scaled-YOLOv4、YOLOv5s 以及本文所提算法 (Ours) 在 VisDrone 数据集上的检测结

表 1 消融实验对比

模型	RFE	MFFS	PCA	P/%	R/%	mAP0.5/%	mAP0.5:0.95/%
YOLOv5s				45.3	34.1	32.6	16.8
YOLOv5s+RFE	✓			46.8	35.0	33.7	17.5
YOLOv5s+RFE+MFFS	✓	✓		50.1	38.5	36.9	20.1
YOLOv5s+RFE+MFFS+PCA	✓	✓	✓	50.9	39.2	38.0	21.0

表 2 实验结果对比

模型	mAP0.5/%	mAP0.5:0.95/%
YOLOV3	40.3	23.8
Scaled-YOLOv4	43.5	24.5
YOLOv5s	51.0	30.4
Ours	54.9	32.5

果。从表 2 中可以看出,改进后的算法相较于其他算法在精度上都有较高提升,本文算法在 mAP0.5 指标上比 Scaled-YOLOv4 提高了 11.4%,比 YOLOv5s 提高了 3.9%,综合来看,本文所提算法更适合航拍图像小目标的检测。

在 VisDrone 数据集中,航拍图像分辨率较高,而小目标分辨率低,大多数航拍图像分辨率在 $1\ 920 \times 1\ 280$ 、 $1\ 360 \times 765$,故本文折中选取分辨率为 $1\ 536$ 作为本文训练与测试时的图像分辨率。实验表明高分辨率图像可以保留更多的小目标特征,检测精度更高,表 3 是本文选取不同分辨率对原始算法和 Ours 精度的对比实验。

表 3 不同输入分辨率实验结果对比

模型	mAP0.5/%	mAP0.5:0.95/%
YOLOv5 ⁶⁴⁰	32.6	16.8
YOLOv5 ^{1\ 280}	50.9	29.6
YOLOv5 ^{1\ 536}	51.4	30.5
Ours ⁶⁴⁰	38.0	21.0
Ours ^{1\ 280}	51.6	30.1
Ours ^{1\ 536}	54.9	32.5

2.3 定性分析

为了充分展示本文提出的基于感受野增强与并行坐标注意力的航拍小目标检测算法效果,选取了不同场景下图片展示,改进后的算法与原算法的可视化对比效果图如图 6 所示。图 6(a)中白色箭头框出局部放大图为 YOLOv5 算法小目标漏检检测结果,图 6(b)中白色虚线箭头框出局部放大图是改进后的小目标漏检检测结果。可以从效果图中看出,原算法在夜景下存在漏检问题,但改进后的算法相较于原始算法可以准确检测夜景下难以检测的小目标,效果提升明显;图 6(c)中白色箭头框出局部放大图为 YOLOv5 密集小目标检测结果,图 6(d)中白色虚线箭头框出局部放大图为改进后密集小目标检测结

果,从图中可以改进后的算法可以检测到更多的密集小目标,有效降低了漏检率;图 6(e)中白色箭头框出局部放大图为 YOLOv5 小目标被遮挡检测结果,图 6(f)中白色虚线箭头框出局部放大图为改进后被遮挡小目标检测结果,从图中可以看出改进后的算法可以有效检测出被遮挡目标。

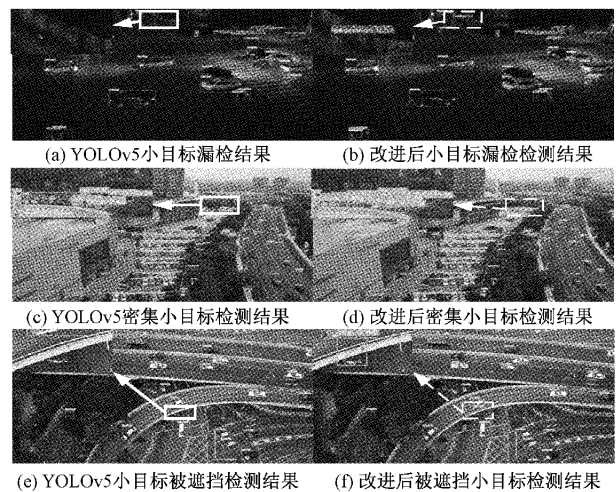


图 6 可视化结果及部分区域放大图对比

3 结 论

本文提出一种基于感受野增强与并行坐标注意力的航拍小目标检测算法。针对航拍图像小目标多导致特征提取困难问题,设计了感受野增强模块来加强特征提取功能,扩大感受野的同时对不同感受野通道之间的关系进行建模,有效地学习到不同通道对于关键信息的权重值,加强提取小目标信息的权重值。增加特征金字塔层数来改进多尺度特征融合结构,提升对小目标的检测能力。在特征融合网络中加入并行坐标注意力模块提升网络对密集小目标等情况的识别效果,改善因远景拍摄导致遮挡和密集情况对小目标的检测带来的精度损失。在 VisDrone 数据集上实验结果可知,本文提出的基于感受野增强与并行坐标注意力的航拍小目标检测算法相较于 YOLOv5 算法有较大的提升,能很好的应对航拍图像小目标多、背景复杂等问题。

参考文献

- [1] HUANG H, SAVKIN A V, HUANG C. Decentralised autonomous navigation of a UAV network for road traffic monitoring [J]. IEEE

- Transactions on Aerospace and Electronic Systems, 2021(99): 1-1.
- [2] 刘跃波, 马立新. 电力巡检无人机避障技术研究[J]. 电子测量技术, 2019, 42(13): 1-5.
- [3] 刘晓霞, 杨风暴, 卫红, 等. 一种农田作物无人机遥感影像阴影检测算法[J]. 电子测量技术, 2022, 45(2): 135-139.
- [4] LI Z, PENG C, YU G, et al. Detnet: Design backbone for object detection[C]. Proceedings of the European conference on computer vision (ECCV), 2018: 334-350.
- [5] 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, 2017: 2117-2125.
- [6] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781-10790.
- [7] WANG Z, ZHANG X, LI J, et al. A YOLO-based target detection model for offshore unmanned aerial vehicle data[J]. Sustainability, 2021, 13(23): 12980.
- [8] REDMON J, FARHADI A. Yolov3: An incremental improvement [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018.
- [9] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3-19.
- [10] 刘英杰, 杨风暴, 胡鹏. 基于 Cascade R-CNN 的并行特征金字塔网络无人机航拍图像目标检测算法[J]. 激光与光电子学进展, 2020, 57(20): 302-309.
- [11] LIANG X, ZHANG J, ZHOU L, et al. Small object detection in unmanned aerial vehicle images using feature fusion and scaling-based single shot detector with spatial context analysis[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 30(6): 1758-1770.
- [12] ULTRALYTICS. YOLOv5 [EB/OL]. (2020-06-03) [2021-04-15]. <https://github.com/ultralytics/yolov5>.
- [13] LIU S, HUANG D. Receptive field block net for accurate and fast object detection[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 385-400.
- [14] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2020: 11531-11539.
- [15] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1-9.
- [16] 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, 2021: 13713-13722.

作者简介

章曙光(通信作者), 教授, 主要研究方向为人工智能、目标检测与无线传感网。

E-mail: zsg@ahjzu.edu.cn

邵政瑞, 硕士研究生, 主要研究方向为人工智能与目标检测。

E-mail: szr@stu.ahjzu.edu.cn

鲍锐, 硕士研究生, 主要研究方向为人工智能与图像处理。

E-mail: 2466161079@qq.com

唐锐, 硕士研究生, 主要研究方向为人工智能与图像处理。

E-mail: 1661677179@qq.com