引文格式:谢亚坤,陈铭臻,赵耀纪,等.顾及小目标特征的无人机影像人员检测方法[J].武汉大学学报(信息科学版),2025,50(6):1077-1087.DOI:10.13203/j.whugis20240368

Citation:XIE Yakun,CHEN Mingzhen,ZHAO Yaoji,et al.A Person Detection Method for UAV Imagery Considering Small Object Features[J]. Geomatics and Information Science of Wuhan University, 2025, 50(6): 1077-1087. DOI: 10.13203/j. whugis20240368

顾及小目标特征的无人机影像人员检测方法

谢亚坤[1] 陈铭臻[1] 赵耀纪[1] 李雨霏[1] 涂佳星[1]
朱 军[1] 朱 庆[1]

1 西南交通大学地球科学与工程学院,四川 成都,611756

摘 要:利用无人机影像进行人员检测在应急救援与社会安全中具有重要意义,然而无人机影像中的人员目标较小且背景复杂,导致人员检测精度较低。为解决这一问题,提出了一种顾及小目标特征的无人机影像人员检测网络。首先,在特征提取阶段使用空间深度转换卷积代替下采样层,减少特征提取过程中的小目标特征丢失;然后,设计选择性特征融合模块,以深层特征权重引导获取浅层特征图中的重点区域,减轻低层特征中的背景噪声干扰;最后,建立上下文感知模块,综合多类型上下文信息,提高网络对于小目标人员的判别能力。为验证方法的有效性,在公开数据集上进行实验分析,结果表明,所提方法的平均准确率为68.9%,准确率为75.5%,召回率为67.7%,相较于经典的目标检测算法分别提升了3.1%~29.5%、0.9%~8.6%、1.0%~57.9%,表明所提方法具有较高的精度。此外,通过对不同场景、不同天气下的无人机影像进行测试,进一步证明了所提方法的泛化性与适用性。

关键词:无人机影像;小目标;人员检测;选择性特征融合;上下文感知

中图分类号:P237 **文献标识码**:A **收稿日期**:2024-11-20

DOI:10.13203/j.whugis20240368 **文章编号**:1671-8860(2025)06-1077-11

A Person Detection Method for UAV Imagery Considering Small Object Features

XIE Yakun[1] *CHEN Mingzhen*[1] *ZHAO Yaoji*[1] *LI Yufei*[1] *TU Jiaxing*[1]
ZHU Jun[1] *ZHU Qing*[1]

1 Faculty of Geosciences and Engineering, Southwest Jiaotong University, Chengdu 611756, China

Abstract: **Objectives**: The rapid development of unmanned aerial vehicle (UAV) remote sensing technology is leading to an increased role for devices in the emergency search and rescue operations and the law enforcement tracking. The unique advantages of UAV include rapid response, wide-area coverage, and panoramic view. The utilization of UAV imagery for person detection is of paramount importance in the context of emergency rescue and social security. However, due to the small and complex background of person in UAV imagery, the high-precision person detection task still presents a significant challenge. **Methods**: To address the aforementioned issue, a person detection network for UAV imagery is proposed. First, the proposed network employs a spatial depth-transformed convolution in lieu of a down sampling layer in the feature extraction stage. This approach ensures the retention of all information in the channel dimensions, preventing information loss and facilitating the preservation of small object features throughout the feature extraction process. Then, a selective feature fusion module is designed to address the issue of noise interference. This module employs high-level features as weights to guide the crucial information in

基金项目:国家自然科学基金(42301473, 42271424, 42171397);中国博士后创新人才支持计划(BX20230299);中国博士后科学基金(2023M742884);四川省自然科学基金(24NSFSC2264);四川省重点研发项目(24ZDYF0633)。

第一作者:谢亚坤,博士,副教授,研究方向为灾害信息智能感知与数字孪生建模。yakunxie@163.com

通信作者:朱军,博士,教授。zhujun@swjtu.edu.cn

low-level features, thereby mitigating the impact of background noise on the latter. Furthermore, it facilitates a more comprehensive integration of the semantic information of high-level features and the detailed information of low-level features. This integration significantly enhances the capacity of network to distinguish between background and foreground information in UAV images. Finally, a context-aware module is designed to address the issue of limited feature information associated with small objects. This module integrates environmental information surrounding the object, including local features, contextual information, and global context, thus enhancing the contextual features of small objects and improving the final object detection accuracy. **Results**: To verify the effectiveness of the proposed method, case experiment analysis is conducted on public datasets. The proposed method achieves mean average precision (mAP) of 68.9%, with precision of 75.5% and recall of 67.7%. Compared to the conventional object detection algorithms, the proposed method shows improvements ranging from 3.1% to 29.5% in mAP, from 0.9% to 8.6% in precision, and from 1.0% to 57.9% in recall. The results indicate that the proposed method exhibits a high level of accuracy. Additionally, the generalization and applicability are confirmed through testing on UAV images in various scenarios and under different weather conditions. **Conclusions**: The proposed method can effectively solve the challenge of detecting small persons in UAV images and significantly improve the detection accuracy. It has a wide range of application prospects in the fields of emergency rescue and social security, and possesses good generalization ability for UAV image detection tasks in different environments.

Key words: UAV imagery; small object; person detection; selective feature fusion; context-aware

随着无人机遥感技术的快速发展,其凭借快速响应、广域覆盖和全景视角等独特优势,在紧急搜救和执法追踪等领域正发挥着日益重要的作用[1-4]。通过将无人机技术与先进的人员检测算法相结合,能够在复杂多变的环境中实时获取精确的感知数据。在自然灾害发生后的搜救行动中,无人机能够迅速覆盖整个灾区,及时发现并定位被困人员的位置[5-6]。同样,在执法行动中,无人机展现出高效追踪逃犯或失踪人员的能力,极大地增强了执法部门的监控与应对能力[7]。这不仅显著提升了任务执行的效率和准确性,还有效降低了人力成本和操作风险,因此,高效的无人机人员检测技术对于加强公共安全和提升应急救援能力具有重要意义。

人员检测由于其应用的广泛性已成为计算机视觉领域中的一个重要研究方向[8-10]。传统的人员检测方法如方向梯度直方图(histogram of oriented gradient, HOG)[11]和支持向量机(support vector machine, SVM)[12],首先通过手工设计特征描述算子提取图像特征,然后通过分类器检索图像内的人员目标。基于深度学习的人员检测算法可以分为两阶段和单阶段检测算法。两阶段检测算法,如Faster R-CNN[13]和Cascade R-CNN[14],先通过区域提议网络选取生成候选区域,再对其进行分类和回归得到检测结果。单阶段检测算法,如SSD[15]和YOLO[16],能够直接对目标进行定位,输出目标的类别检测信息。小目标由于占比面积小、分辨率低、可用特征少,其检测精度相比大中目标有所下降。因此,学者们提出了多种改进方案,文献[17-19]利用多尺度特征融合以保留更多小目标特征;文献[20-21]通过上下文学习来更好地理解小目标与周围环境的关系;文献[22-23]采用注意力机制以提高对小目标的关注;文献[24-25]通过超分辨率技术解决小目标特征不足的问题。与其他视角相比,无人机视角下的广阔视场提供了丰富的信息,但也伴随着更为复杂的背景和更多噪声的干扰。同时,无人机影像中的目标尺度变化大,且小目标的比例远高于自然场景图像。

针对上述问题,本文提出面向小目标的无人机影像人员检测方法,利用空间深度转换卷积(space-to-depth convolution, SPD-Conv)提高检测过程中对于小目标的保持能力,设计选择性特征融合模块,更好地融合高层特征的语义信息和低层特征的细节信息,提高对于背景的抗干扰能力,通过上下文感知模块综合局部、全局及周围上下文,增强网络对于小目标人员的特征提取能力,提升无人机影像小目标人员检测的准确性与泛化性。

1　方法

1.1　方法框架

顾及小目标特征的无人机影像人员检测方法框架如图 1 所示,包括特征提取、特征融合及检测头 3 个部分。首先,采用 CSPDarkNet[26] 进行特征提取,同时,使用 SPD-Conv[27] 代替下采样层,以减少小目标特征丢失;在特征融合部分,设计了一个选择性特征融合(selective feature fusion,SFF)模块,利用高层特征作为权重来引导低层特征中的重要信息,减轻低层特征中的背景噪声干扰,更好地融合高层特征的语义信息和低层特征的细节信息;在检测头部分,设计了一个上下文感知模块(context-aware module,CAM),综合目标周围的环境信息,丰富小目标的上下文特征,提高大视角下的人员检测精度。

如图 1 所示,输入单个无人机影像,首先提取多层特征,其中特征映射的通道数 $i=\{1,2,3,4\}$,每层对应的通道数为 $\{160,320,640,640\}$,$H_i=H/2^{i+1}$ 代表高度,$W_i=W/2^{i+1}$ 代表宽度;然后将多层特征自上而下通过 SFF 模块进行特征融合;最后将融合后的特征通过 CAM 捕获上下文信息,得到最终分类和定位的特征,生成检测框并进行后处理,得到最终输出。

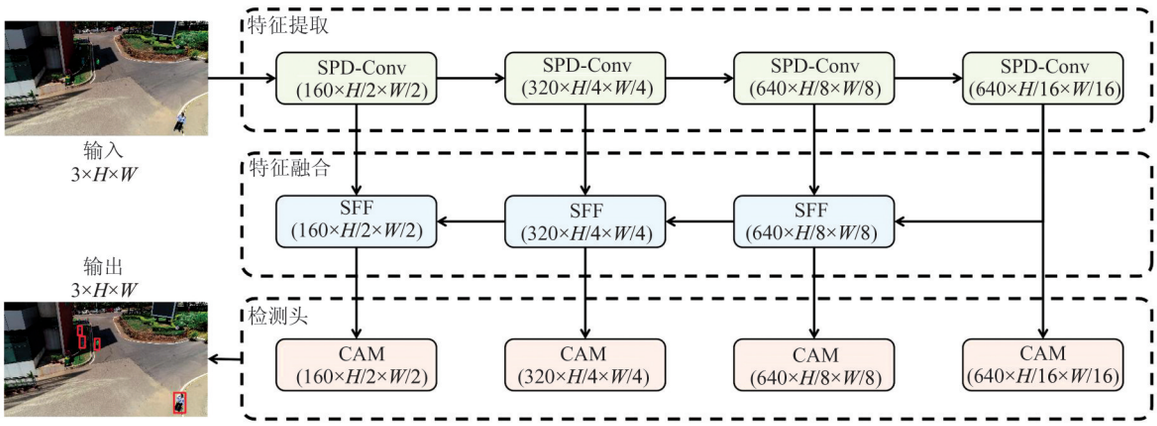


图 1　本文方法框架

Fig. 1　Framework of the Proposed Method

1.2　SPD-Conv

常用的图像特征提取过程多通过跨步卷积和池化操作,在目标大小适中时,跨步卷积和池化操作忽略的那部分像素信息不会对模型性能产生较大影响。然而,在小目标任务下,基于此种设计理念的卷积神经网络(convolutional neural network,CNN)会出现低细粒度信息丢失和模型学习特征不足的问题。为此,本文使用 SPD-Conv 替代跨步卷积与池化操作。SPD-Conv 由一个 SPD 层和一个非跨步卷积层构成,其中 SPD 层使用一种图像转换技术对特征图进行下采样,在通道维度中保留所有信息,确保无信息丢失;每个 SPD 层后紧跟一个无步长卷积操作,无步长卷积以逐元素卷积的方式在很大程度上保证不丢失图像特征。

1.3　选择性特征融合

网络结构中高层特征具有较大的感受野和丰富的语义信息,但会丢失几何细节信息,难以检测到小目标人员。相比之下,低层特征在很大程度上保留了小目标人员的细节信息,但缺乏语义信息,若直接利用低层特征进行检测,无法保证检测的精度。因此,需要将多个尺度的特征融合在一起,获得既有高级语义特征又有低层细节特征的特征图。然而,在特征融合的过程中,无人机影像的背景噪声特征也融合到特征金字塔中,影响人员检测的准确性。为了解决这一问题,本文设计了一个选择性特征融合模块,如图 2 所示。通过利用高层特征作为权重来引导低层特征中的重要信息,减轻低层特征中的背景噪声干扰,并更好地融合高层特征的语义信息和低层特征的细节信息,以提高小目标人员的检测能力。

1)空间信息聚合。输入高层特征 F_h 和低层特征 F_l,并利用转置卷积将高层特征的大小与低层特征统一。在统一大小后的高层特征上,分别使用最大池化操作和平均池化操作聚合空间信息,输出两个池化特征。最大池化操作选取每个池化窗口内的最大值,以强调图像中的显著特征,有助于提取和突出小目标人员的关键特征;平均池化操作通过计算每个池化窗口内特征值

的平均值来降低特征图的细节度,有助于减少背景噪声对小目标人员的干扰。空间信息聚合过程可表示为:

$$F_h' = \mathrm{Conv}^{\mathrm{T}}(F_h) \tag{1}$$

$$F_{\max} = \mathrm{MaxPool}(F_h') \tag{2}$$

$$F_{\mathrm{avg}} = \mathrm{AvgPool}(F_h') \tag{3}$$

式中,F_h'为经过转置卷积后的高层特征;$F_{\max}$为最大池化特征;F_{avg}为平均池化特征;$\mathrm{Conv}^{\mathrm{T}}$表示转置卷积操作,MaxPool、AvgPool分别表示最大池化操作、平均池化操作。

2)重点特征权重获取。将两个池化特征输入包含隐层的多层感知器(multilayer perceptron,MLP)中,得到两个通道注意力图。MLP能够学习输入特征的抽象表示,并且通过引入隐层增强了网络对特征之间复杂关系的建模能力。同时,将两个通道注意力图相加,融合两个池化特征的通道注意力信息,并经过Sigmoid激活操作,得到最终的通道注意力图M_c,使网络能够对不同通道的特征进行加权,以便更好地区分人员和背景。重点特征权重获取过程可表示为:

$$M_c = \sigma(\mathrm{MLP}(F_{\max}) + \mathrm{MLP}(F_{\mathrm{avg}})) \tag{4}$$

式中,σ表示Sigmoid激活。

3)选择性特征融合。将M_c与低层特征加权,调整特征图中各个通道的权重,从而提高网络对小目标人员的关注度并抑制背景噪声干扰,得到加权的低层特征。在此基础上,将加权的低层特征与统一大小后的高层特征相加融合,使其能够更好地融合高层语义和低层细节特征,实现特征的选择性融合,提高网络对于前景与背景信息的判别能力。选择性特征融合过程可表示为:

$$F_{\mathrm{fused}} = M_c \cdot F_l + F_h' \tag{5}$$

式中,F_{fused}为融合后的特征;·表示加权。

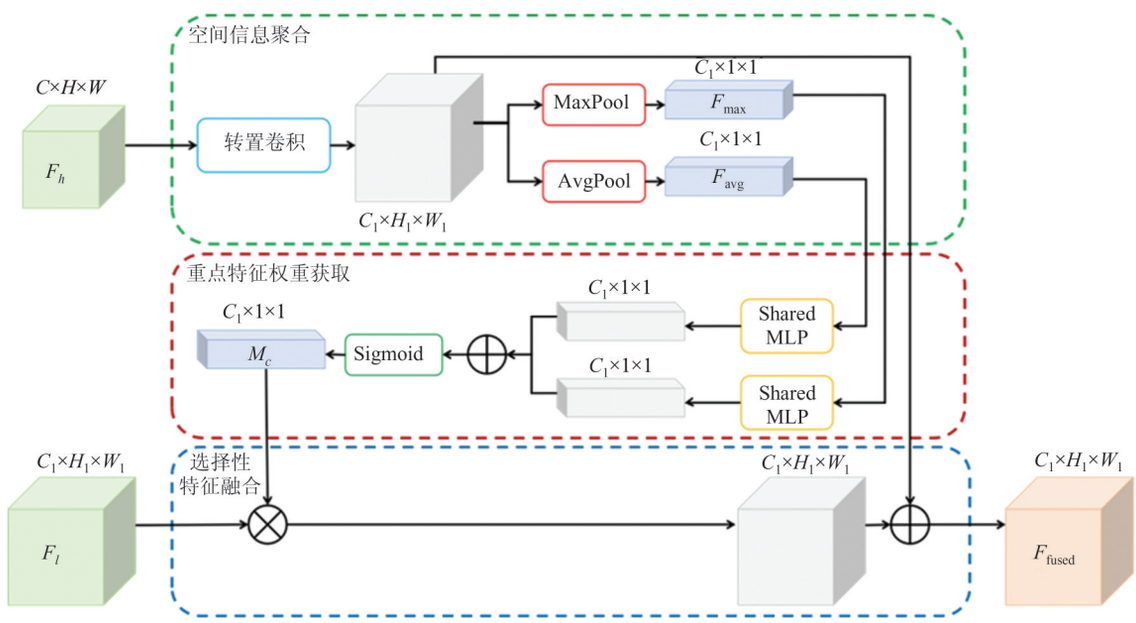


图2 选择性特征融合模块

Fig. 2 Selective Feature Fusion Module

1.4 上下文感知

在无人机影像中,小目标人员所占的像素较少,仅依靠小目标人员自身的特征难以实现准确检测。同时,小目标人员附近区域能够提供有用的上下文信息,有助于检测小目标人员。然而,现有的人员检测方法通常会忽略这些局部区域之外的上下文特征。因此,本文设计了一个上下文感知模块,如图3所示,综合小目标人员周围的环境信息,包括局部特征提取器f_{local}、周围上下文提取器f_{surround}、联合特征提取器f_{joint}和全局上下文提取器f_{global},用来丰富小目标的上下文特征,可有效提升模型对于小目标特征的提取能力。

1)联合特征获取。减少输入特征的通道数,通过3×3卷积以小感受野的遍历获取局部特征,同时,利用扩张卷积扩大感受野,有效地学习小目标人员的周围上下文,在此基础上,将局部特征与周围上下文特征进行融合以得到联合特征,连接操作将小目标人员局部特征和周围上下文特征结合起来,形成一个综合的特征表示,联合后的特征通过批量归一化(batch nor-

malization, BN)和参数化线性单元(parametric rectified linear unit, PReLU)等操作进行进一步的加工,以增强小目标人员的特征表示能力。联合特征获取可表示为:

$$F_{\text{joint}}=\text{PReLU}(\text{BN}(\text{Conv}_{3\times3}(\text{Conv}_{1\times1}(X))\oplus\text{DilatedConv}(\text{Conv}_{1\times1}(X)))) \tag{6}$$

式中,X为输入特征;F_{joint}为联合特征;$\text{Conv}_{n\times n}$表示$n\times n$卷积;DilatedConv表示扩张卷积;$\oplus$表示特征连接。

2)全局上下文特征提取。通过全局平均池化(global average pooling, GAP)来聚合整个特征图的全局信息,生成全局特征向量。该特征向量捕获了输入图像中每个通道的平均响应,随后通过MLP进一步处理。MLP能够学习特征间的复杂非线性关系,进一步细化得到最终的全局上下文特征。全局上下文特征提取可表示为:

$$F_{\text{global}}=\text{MLP}(\text{GAP}(X)) \tag{7}$$

式中,F_{global}为全局上下文特征。

联合特征与全局上下文特征在特征表达上具有互补性。联合特征通过整合局部特征与其周围的上下文信息,能够捕捉目标对象及其邻近环境的细节,适合用于小目标对象的局部特征表达。而全局上下文特征则通过GAP和MLP捕获图像的整体语义信息,反映出更全面的上下文关系。两者的结合使得模型既能精准关注局部细节,又具备对全局背景的理解,从而显著提升特征表达的完整性与判别力,为小目标对象检测提供更为精确的支持。

3)上下文信息融合。将提取的全局上下文与联合特征加权,这个操作相当于将全局上下文信息作为权重,在通道级别重新标定联合特征,强调有用的特征部分,抑制不重要的特征部分。在此基础上,将上下文信息与输入的特征相加融合,得到分类和回归预测的特征图。该特征图融合了局部特征、周围上下文特征以及全局上下文信息,帮助模型更准确地定位和识别小目标人员。上下文信息融合可表示为:

$$F_{\text{context}}=F_{\text{global}}\cdot F_{\text{joint}}+X \tag{8}$$

式中,F_{context}为上下文信息融合后的特征。

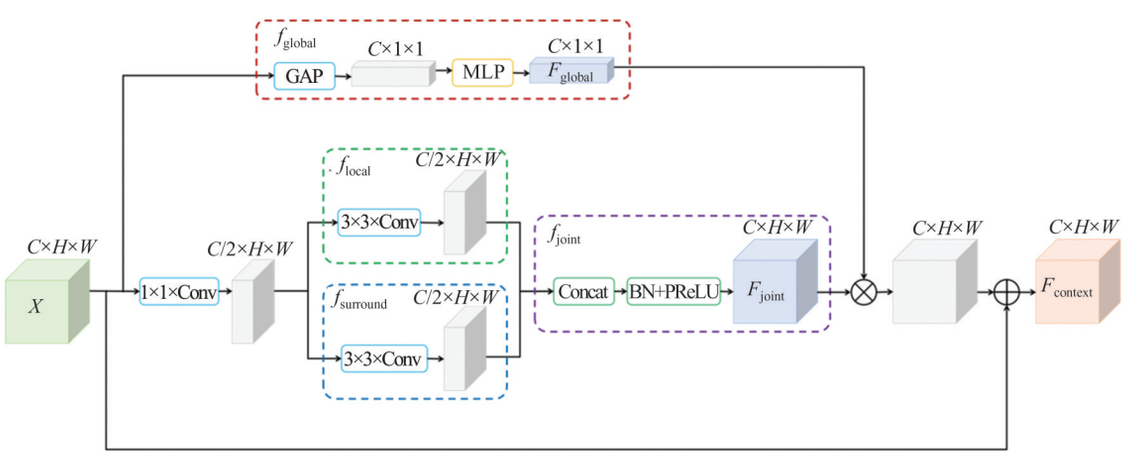


图3　上下文感知模块

Fig. 3　Context-Aware Module

2　实验案例分析

2.1　数据集构建

本文选择公开数据集Manipal-UAV[28]进行方法效果对比。该数据集包括从两架在不同高度、位置和天气条件下飞行的无人机收集的数据,可用于研究无人机影像人员检测,包含来自33个视频的13 462张图像,具有153 112个目标,其中约70%属于小目标(小目标定义为分辨率小于32×32像素的目标)。该数据集划分为训练集和测试集,其中训练集9 113张,验证集2 106张,测试集2 243张,共计13 462张图像。需要说明的是,训练集和测试集的图像不存在重叠情况。

为了测试本文方法的泛化性和适用性,笔者团队在西南交通大学校园内不同场景和时间采集了无人机影像。场景类型包括教学楼、图书馆、食堂、宿舍楼等人员密集的场景,确保本文方法对不同场景的适用性。此外,考虑到天气的影响,分别在不同时间段采集图像,获取明暗程度各不相同的图像,以确保本文方法对于各种光线的场景均具有较好的泛化性。

2.2 参数设置与评价标准

2.2.1 参数设置

本文及对比方法的实验在Windows 10环境下进行,CPU型号为Intel(R) Xeon(R) Platinum 8280 CPU @ 2.70 GHz,内存为512 GB,GPU为NVIDIA Quadro RTX 8000,显存为48 700 MB。网络采用PyTorch框架实现,并行计算加速采用CUDA 11.3和CUDNN 8.6框架。在模型训练阶段,训练轮数设置为100,训练批次为8,初始学习率为0.01,训练批量大小设置为8。模型通过随机梯度下降法进行优化,其动量和权重衰减系数分别设置为0.937和5×10^{-4}。

2.2.2 评价标准

为了准确客观地对人员检测算法性能进行评价,本文采用准确率P、召回率R、平均准确率(mean average precision,mAP)等通用的评价指标,其中P用于衡量算法对于正样本的查准能力,R用于衡量算法对于正样本的查全能力,mAP是以R为X轴、P为Y轴做P-R曲线,mAP为P-R曲线与坐标轴所围成的面积,理论上面积越大,模型的性能越好。计算公式如下:

$$P=\frac{N_{TP}}{N_{TP}+N_{FP}}\times 100\% \tag{9}$$

$$R=\frac{N_{TP}}{N_{TP}+N_{FN}}\times 100\% \tag{10}$$

$$P_{mAP}=\int_0^1 P(R)\mathrm{d}R \tag{11}$$

式中,TP(true positive)为真正例,N_{TP}指真实标签为人员且被正确预测为人员的数量,即模型正确检测人员数量;FP(false positive)为假正例,N_{FP}指真实为背景且被错误预测为人员的数量,即模型误检人员数量;FN(false negative)为假负例,N_{FN}指真实标签为人员且被错误预测为背景的数量,即模型漏检人员数量。

2.3 人员检测实验结果分析

2.3.1 模型精度对比分析

1)定量分析

为验证本文方法的有效性和先进性,将本文方法与近年来经典的人员检测方法在Manipal-UAV数据集上进行对比测试,结果如表1所示。测试数据包括从不同高度、位置和天气条件下飞行的两架无人机收集的数据,具有较大的挑战性。由对比结果可以看出,本文方法在Manipal-UAV数据集上的mAP为68.9%,P为75.5%,R为67.7%。相较于其他6种算法,本文方法在3个指标上都有较大的提升,分别提升了3.1%~29.5%、0.9%~8.6%、1.0%~57.9%,这充分表明了本文方法的先进性。

表1 人员检测算法精度对比结果/%

Table 1 Comparison Results of Accuracy of Person Detection Algorithms/%

方法	P	R	mAP
Faster R-CNN	66.9	66.4	59.9
Cascade R-CNN	68.8	66.7	62.8
RetinaNet	73.8	34.1	53.2
CornerNet	70.1	9.8	39.4
FCOS	74.6	46.8	64.5
YOLOv8x	74.5	59.9	65.8
本文方法	75.5	67.7	68.9

2)定性分析

为了更直观地对比本文方法和其他算法的检测结果,选择不同场景对实验结果进行进一步的定性分析,并选取了部分检测结果进行展示,如图4所示。由图4可以看出,Faster R-CNN算法对于场景一中距离最远的小目标人员能检测出来,但将车辆和人影误检为人员;对于场景二中距离最远的小目标人员存在漏检。Cascade R-CNN算法在场景一和场景二中对距离最远的小目标人员均存在漏检。RetinaNet算法在场景一和场景二中能够检测出距离最远的小目标人员,但存在误检的情况,分别将人影和公园设施误检为人员。CornerNet算法在场景一中对距离最远的小目标人员存在漏检,但在场景二中能够检测出距离最远的小目标人员,然而对于距离较近的两个小目标人员存在漏检。FCOS算法在场景一中能够检测出距离最远的小目标人员,但将人影误检为人员;在场景二中对距离最远的小目标人员存在漏检,同时将公园设施误检为人员。YOLOv8x算法在场景一中对距离最远的小目标人员存在漏检,但在场景二中能够检测出距离最远的小目标人员,然而对于距离较近的一个小目标人员存在漏检。相比之下,本文方法能够检测出场景一和场景二中的所有小目标人员,且不存在误检情况。总的来说,本文方法相比其他主流的检测算法具有更好的性能。

2.3.2 模型泛化性分析

为测试本文方法的泛化性和在复杂场景中的适用性,笔者在西南交通大学校园内采集了无人机影像进行测试。测试场景包括教学楼、图书馆、食堂和宿舍楼,其中教学楼场景包含明暗程度各不相同的无人机影像。为了与现有方法进

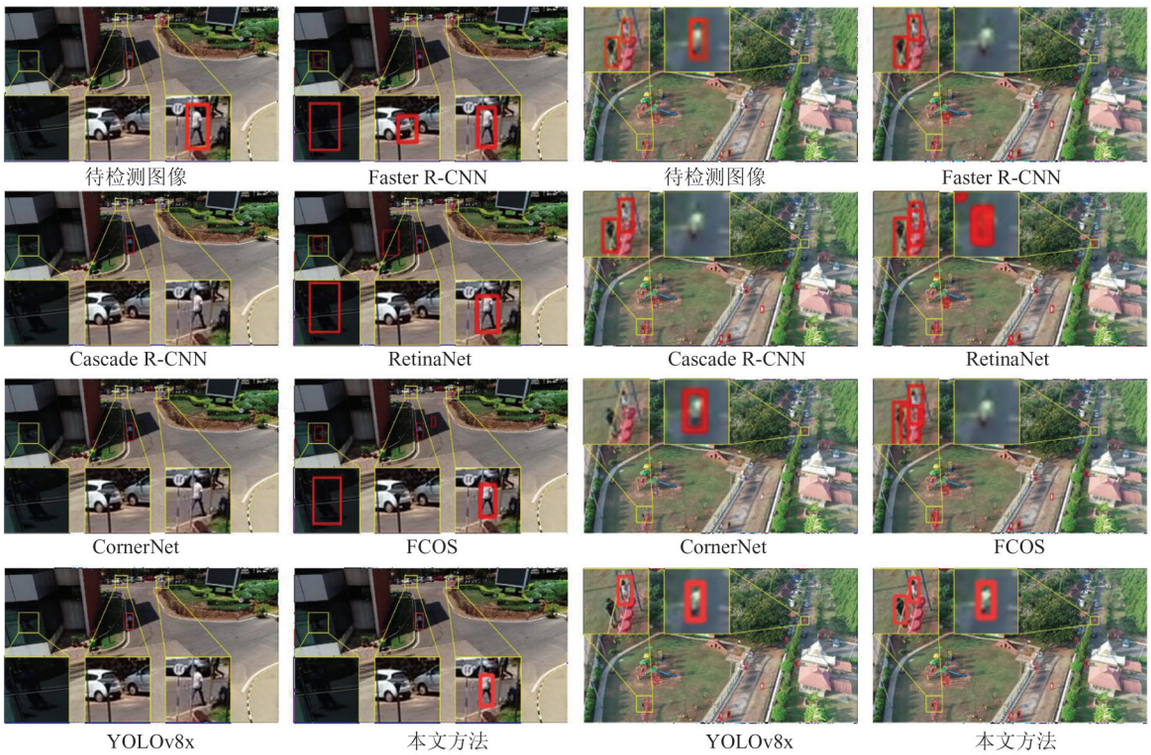


图 4　检测结果对比

Fig. 4　Comparison of Detection Results

行对比,本文选择了表现最好的 3 种算法进行对比分析。需要说明的是,所有算法皆没有通过任何的迁移学习,而是直接在不同场景和天气情况下进行人员检测实验。

图 5 展示了不同场景下的人员检测结果,其中场景一至场景四分别代表教学楼、图书馆、食堂和宿舍楼。场景一中阳光极强,有树木及建筑物产生的阴影,黑色衣服的人员易与阴影混淆,同时小型的自行车也对人员检测产生了干扰,Cascade R-CNN 算法和 YOLOv8x 算法存在漏检,Cascade R-CNN 算法和 FCOS 算法存在误检。场景二中存在极小的人员目标,且树干较多,远距离下干扰严重,Cascade R-CNN 算法和 FCOS 算法均存在漏检。场景三和场景四由于拍摄角度较大,导致对人员的特征表达能力更弱。场景三中将自行车误检为人员,存在小目标人员漏检;场景四中 Cascade R-CNN 算法和 YOLOv8x 算法将电动车误检为人员,YOLOv8x 算法还存在人员漏检情况,而本文方法能够准确地检测出无人机图像中的小目标人员信息。

考虑到场景和天气的复杂性,本文针对同一场景在不同明暗程度下的检测结果进行了分析,结果如图 6 所示。由图 6 可以看出,本文方法能够检测出明暗程度不同的场景中的小目标人员,尤其在第 3 组的较暗环境下,不存在误检和漏检的情况。而在第 1 组中,Cascade R-CNN 算法和 YOLOv8x 算法存在小目标人员漏检,Cascade R-CNN 算法和 FCOS 算法将自行车误检为人员;在第 2 组中,Cascade R-CNN 算法、FCOS 算法和 YOLOv8x 算法均存在小目标人员漏检;在第 3 组中,Cascade R-CNN 算法、FCOS 算法和 YOLOv8x 算法均将地砖误检为人员。这证明了本文方法在各种光线条件下的场景中均具有较强的泛化性。

2.3.3　网络模块性能测试

本文方法包括多种模块和复杂结构,包括 SPD-Conv、SFF 和 CAM。为充分展示每个模块的作用,通过详细的消融实验进行了模块性能测试,如表 2 所示。由表 2 可以看出,去除 SPD-Conv 模块后,mAP 下降了 1.7%,P 下降了 0.5%,R 下降了 3.8%,说明 SPD-Conv 模块可以减少小目标特征的丢失,提高网络的特征提取能力。去除 SFF 模块后,mAP 下降了 1.1%,P 下降了 0.2%,R 下降了 1.4%,证明 SFF 模块能够减轻低层特征中的背景噪声干扰,提高小目标人员的检测能力。去除 CAM 模块后,mAP 下降了 1.3%,P 下降了 0.6%,R 下降了 2.9%,表明 CAM 模块能够丰富小目标的上下文特征。



图5 不同场景对比分析

Fig. 5 Comparative Analysis of Different Scenarios

为了更直观地显示本文方法各模块的效果,将上述方法在Manipal-UAV数据集上的消融实验结果进行可视化对比,并选取了其中部分检测结果和热力图进行展示,如图7所示。由图7的检测结果可以看出,除了本文方法,其余模型对于图像1中距离较远的多个小目标人员和被屋顶遮挡的小目标人员都存在漏检,对于图像2中距离最远的小目标人员都存在漏检。去除SPD-Conv的模型、去除SFF的模型和去除CAM的模型对于图像2中距离最近的小目标人员输出的检测框没有将人员完整框住。改进前的模型和去除SPD-Conv的模型将交通锥误检为人员。本文方法相较于去除模块的模型,对于图像1和图像2中的小目标人员都能检测出来,且不存在误检的情况,而且输出的检测框更加紧凑。

除检测结果外,本文进一步分析了检测结果热力图,热力图中颜色的深浅程度反映算法对特定区域的关注度,颜色越深,表示该区域在算法中的权重越大,对最终检测结果的贡献越显著。由图7的热力图可以看出,在未加入SPD-Conv模块的情况下,小目标人员没有位于明显的深颜色区域内,表明去除SPD-Conv的模型对小目标人员的关注度不够;在未加入SFF模块的情况下,小目标人员位于明显的深颜色区域内,但是较多非人员也位于明显的深颜色区域内,表明去除SFF的模型受到背景噪声的干扰;在未加入CAM模块的情况下,小目标人员没有位于明显的深颜色区域内,表明去除CAM的模型对小目标人员的关注度不够;本文方法的热力图中,所有小目标人员均位于明显的深颜色区域内,且较少非人员位于明显的深颜色区域内,说明本文方法能够更精准地聚焦于小目标人员的特征信息,有效降低背景噪声的干扰。

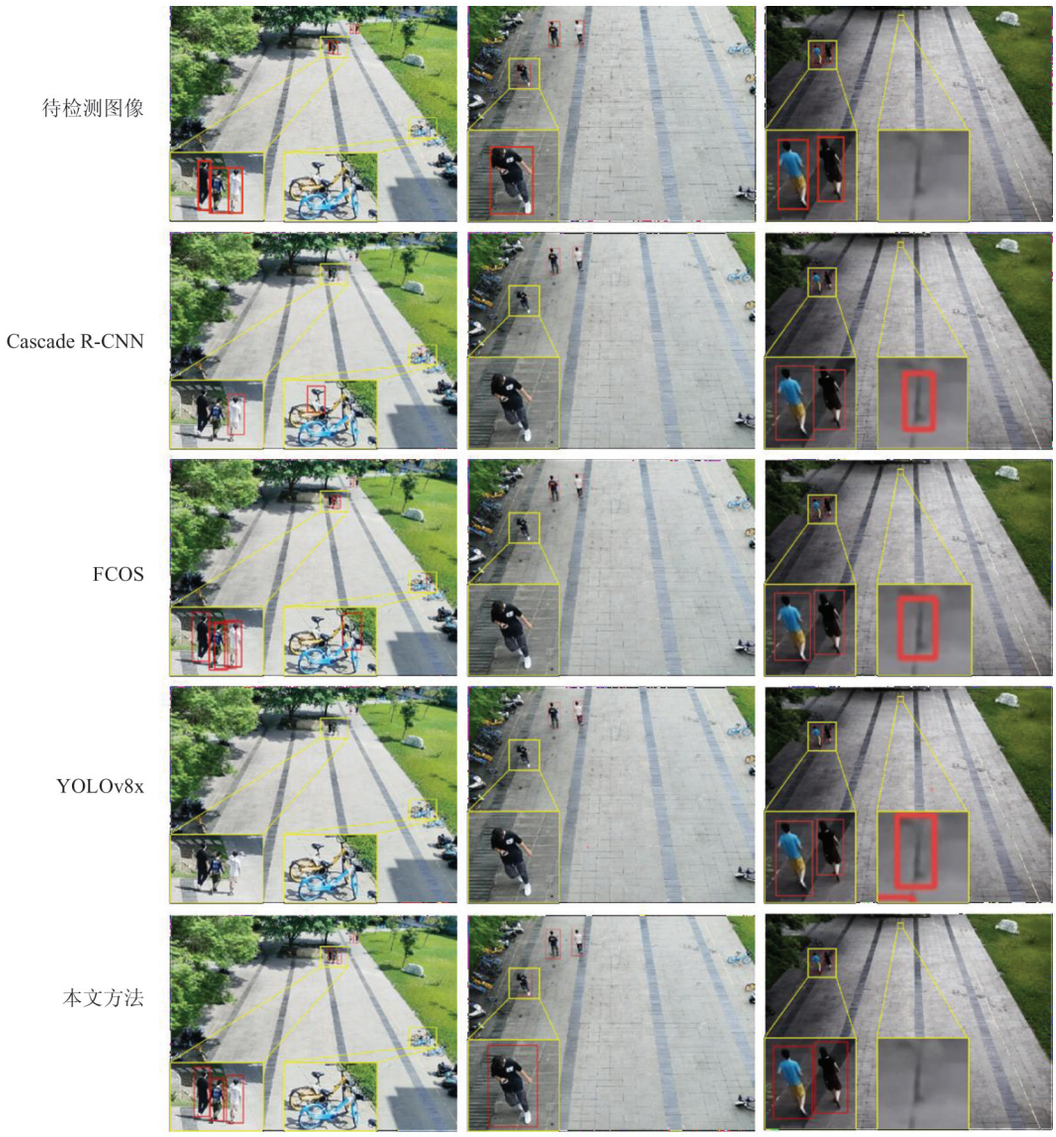


图 6　不同明暗程度对比分析

Fig. 6　Comparative Analysis of Different Light and Dark Levels

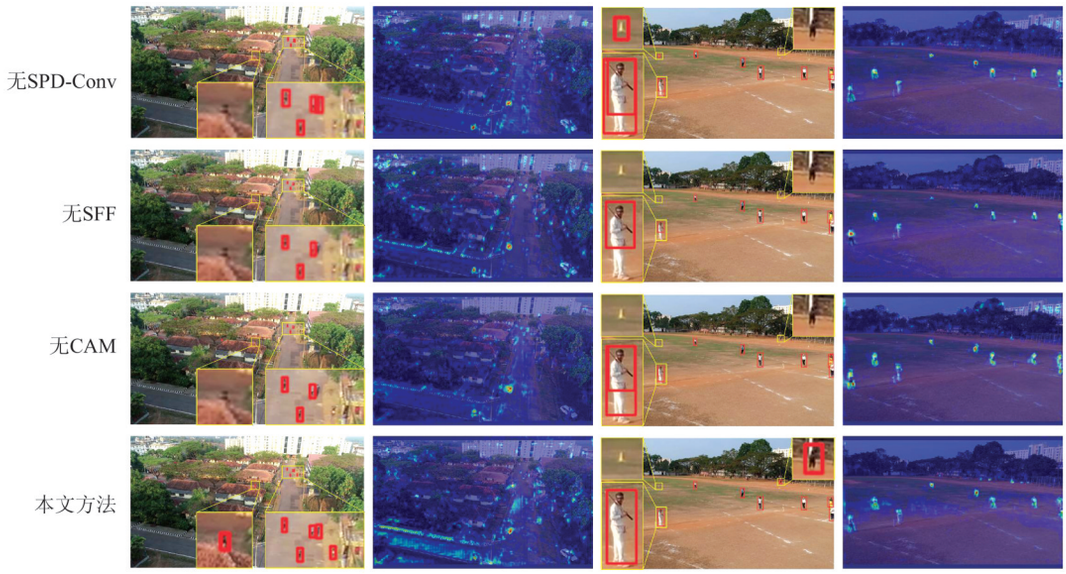
表 2　消融实验对比结果/%

Table 2　Comparison Results of Ablation Experiment/%

方法	*P*	*R*	mAP
无SPD-Conv	75.0	63.9	67.2
无SFF	75.3	66.3	67.8
无CAM	74.9	64.8	67.6
本文方法	75.5	67.7	68.9

3　结　语

本文提出了一种顾及小目标特征的无人机影像人员检测方法,旨在解决无人机影像中人员目标较小且背景复杂的问题。主要结论如下:(1)针对背景噪声干扰的问题,建立了SFF模块,利用高层特征作为权重来引导低层特征中的重要信息,减轻低层特征中的背景噪声干扰,更好地融合高层特征的语义信息和低层特征的细节信息,显著提高网络对于无人机影像背景与前景信息的判别能力;(2)针对小目标特征信息少的问题,设计了CAM模块,综合目标周围的环境信息,包括局部特征、周围上下文和全局上下文,丰富小目标的上下文特征,进一步提升了方法对于无人机大视角及小目标人员的特征提取能力。将本文方法与6种深度学习方法进行对比分析,结果表明,mAP提高了3.1%～29.5%,*P*提高了0.9%～8.6%,*R*提高了1.0%～57.9%,明显高于现有方法精度。此外,从方法泛化性、消融实验等方面对本文方法进行了详细分析,进一步验证了本文方法缓解了无人机影像人员检测精度低



(a) 图像1检测结果 (b) 图像1热力图 (c) 图像2检测结果 (d) 图像2热力图

图7 消融实验检测结果和热力图

Fig. 7 Detection Results and Heat Maps of Ablation Experiments

的问题。然而,现实场景复杂多样,尤其对于人员密集等特殊场景的人员高效检测仍面临巨大挑战,下一步将致力于算法的精度提升;同时,将探索光学影像与多类型传感器相结合,提升无人机影像人员检测在夜间及云雨雾等场景中的应用效果,以适应更复杂的应用场景。

参 考 文 献

[1] TANG G Y, NI J J, ZHAO Y H, et al. A Survey of Object Detection for UAVs Based on Deep Learning[J]. *Remote Sensing*, 2024, 16(1): 149.

[2] ZHAO B G, SONG R, ZHOU Y, et al. Heuristic Data-Driven Anchor Generation for UAV-Based Maritime Rescue Image Object Detection[J]. *Heliyon*, 2024, 10(10): e30485.

[3] 邵振峰, 程涛, 李德仁, 等. 空地协同移动智能服务平台研发及典型应用[J]. 武汉大学学报(信息科学版), 2023, 48(5): 749-755.
SHAO Zhenfeng, CHENG Tao, LI Deren, et al. Development and Applications of Aerial-Ground Collaborative Mobile Intelligent Service Platform [J]. *Geomatics and Information Science of Wuhan University*, 2023, 48(5): 749-755.

[4] 王欣, 方成勇, 唐小川, 等. 泸定Ms 6.8地震诱发滑坡应急评价研究[J]. 武汉大学学报(信息科学版), 2023, 48(1): 25-35.
WANG Xin, FANG Chengyong, TANG Xiaochuan, et al. Research on Emergency Evaluation of Landslides Induced by the Luding Ms 6.8 Earthquake [J]. *Geomatics and Information Science of Wuhan University*, 2023, 48(1): 25-35.

[5] XU J H, FAN X T, JIAN H D, et al. YoloOW: A Spatial Scale Adaptive Real-Time Object Detection Neural Network for Open Water Search and Rescue from UAV Aerial Imagery [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 5623115.

[6] 董秀军, 邓博, 袁飞云, 等. 航空遥感在地质灾害领域的应用: 现状与展望[J]. 武汉大学学报(信息科学版), 2023, 48(12): 1897-1913.
DONG Xiujun, DENG Bo, YUAN Feiyun, et al. Application of Aerial Remote Sensing in Geological Hazards: Current Situation and Prospects[J]. *Geomatics and Information Science of Wuhan University*, 2023, 48(12): 1897-1913.

[7] 蒋小明. 基于深度神经网络的无人机行人检测方法[D]. 西安: 西安电子科技大学, 2022.
JIANG Xiaoming. Pedestrian Detection Method of UAV Based on Deep Neural Network[D]. Xi'an: Xidian University, 2022.

[8] 邵香迎, 郭颖, 王友伟. AF-RetinaNet: 一种基于自适应融合与特征细化的微小行人检测算法[J]. 控制与决策, 2024, 39(3): 939-946.
SHAO Xiangying, GUO Ying, WANG Youwei. AF-RetinaNet: A Tiny Person Detection Algorithm Based on Adaptive Fusion and Feature Refinement [J]. *Control and Decision*, 2024, 39(3): 939-946.

[9] GABSTEIGER J, MAIWALD T, KURIN T, et al. Deep Learning-Based Person Detection on a Moving Robot [C]//IEEE Topical Conference on

Wireless Sensors and Sensor Networks (WiSNeT), San Antonio, USA, 2024.

[10] YANG Y, HAN X M, WANG K R, et al. NRPerson: A Non-Registered Multi-modal Benchmark for Tiny Person Detection and Localization[J]. *Electronics*, 2024, 13(9): 1697.

[11] DALAL N, TRIGGS B. Histograms of Oriented Gradients for Human Detection[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, USA, 2005.

[12] MAJI S, BERG A C, MALIK J. Classification Using Intersection Kernel Support Vector Machines Is Efficient[C]//IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, USA, 2008.

[13] REN S Q, HE K M, 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*, 2017, 39(6): 1137-1149.

[14] CAI Z W, VASCONCELOS N. Cascade R-CNN: Delving into High Quality Object Detection[C]//IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018.

[15] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector[C]//Computer Vision — ECCV Workshop, Amsterdam, The Netherlands, 2016.

[16] REDMON J, DIVVALA S, GIRSHICK R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 2016.

[17] ZHANG H Y, CHEN E Y. Bi-AFN++CA: Bi-Directional Adaptive Fusion Network Combining Context Augmentation for Small Object Detection[J]. *Applied Intelligence*, 2024, 54(1): 614-628.

[18] YANG H B, QIU S. A Novel Dynamic Contextual Feature Fusion Model for Small Object Detection in Satellite Remote-Sensing Images[J]. *Information*, 2024, 15(4): 230.

[19] 黄丽, 龚志鹏, 刘凡凡, 等. 图像跨尺度特征融合与数据增强的公交客流检测模型[J]. 武汉大学学报(信息科学版), 2024, 49(5): 700-708.
HUANG Li, GONG Zhipeng, LIU Fanfan, et al. Bus Passenger Flow Detection Model Based on Image Cross-Scale Feature Fusion and Data Augmentation[J]. *Geomatics and Information Science of Wuhan University*, 2024, 49(5): 700-708.

[20] HE X, ZHENG X L, HAO X Y, et al. Improving Small Object Detection via Context-Aware and Feature-Enhanced Plug-and-Play Modules[J]. *Journal of Real - Time Image Processing*, 2024, 21(2): 44.

[21] LI Z, WANG Y C, ZHANG Y X, et al. Context Feature Integration and Balanced Sampling Strategy for Small Weak Object Detection in Remote Sensing Imagery[J]. *IEEE Geoscience and Remote Sensing Letters*, 2024, 21: 6009105.

[22] CHE M L, WU Z H, WANG X W, et al. Designing Lightweight Small Object Detection Models Using Attention and Context[J]. *Multimedia Tools and Applications*, 2024, 83(4): 9523-9546.

[23] LI L, GAO S K, WU F F, et al. MBAN: Multi-branch Attention Network for Small Object Detection[J]. *PeerJ Computer Science*, 2024, 10: e1965.

[24] GAO Y X, WANG Y C, ZHANG Y X, et al. Feature Super-Resolution Fusion with Cross-Scale Distillation for Small-Object Detection in Optical Remote Sensing Images[J]. *IEEE Geoscience and Remote Sensing Letters*, 2024, 21: 6008105.

[25] HARIS M, SHAKHNAROVICH G, UKITA N. Task-Driven Super Resolution: Object Detection in Low-Resolution Images[C]//Neural Information Processing, Amsterdam, The Netherlands, 2021.

[26] TERVEN J, CÓRDOVA-ESPARZA D M, ROMERO-GONZÁLEZ J A. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS[J]. *Machine Learning and Knowledge Extraction*, 2023, 5(4): 1680-1716.

[27] SUNKARA R, LUO T. No more Strided Convolutions or Pooling: A New CNN Building Block for Low-Resolution Images and Small Objects[C]//Machine Learning and Knowledge Discovery in Databases, Zurich, Switzerland, 2023.

[28] AKSHATHA K R, KARUNAKAR A K, SATISH SHENOY B, et al. Manipal-UAV Person Detection Dataset: A Step Towards Benchmarking Dataset and Algorithms for Small Object Detection[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2023, 195: 77-89.