计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
[doi: 10.15888/j.cnki.csa.010072] [CSTR: 32024.14.csa.010072]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

UAV-YOLO: 基于 YOLOv7 的红外小目标检测①

苏　健, 张莉莉

(南京信息工程大学 软件学院, 南京 210044)
通信作者: 苏　健, E-mail: sj890718@gmail.com

摘　要: 在无人机红外小目标检测领域, 复杂地面背景与目标尺寸微小的特性, 常使检测模型出现漏检或误检问题. 针对这一情况, 本研究以 YOLOv7 框架为基础, 设计出一种轻量级的高精度红外小目标检测算法, 命名为 UAV-YOLO (unmanned aerial vehicle-you only look once). 首先, 考虑到检测目标多为小目标, 对 YOLOv7 基础网络进行了轻量化改造, 将 ELAN、ELAN-W 和 CARAFE 模块中的 1×1 卷积以及颈部网络的 1×1 卷积替换为 GSConv, 同时去除检测效率不高的 P5 检测头, 并新增专门用于检测小目标的 P2 检测头. 不仅提升了检测效率, 还大幅减少了参数量, 实现了模型的轻量化. 其次, 在骨干网络中融入了改进后的金字塔池化模块 SPPFCSPC. 该模块的加入, 能够有效扩展模型的感知范围, 从而提高对红外小目标的检测精度. 然后, 把内容感知特征重组 (CARAFE) 架构集成到 YOLOv7 中. 这一架构可以更好地保留和优化小目标的特征表示. 同时, 在检测头前引入 CA (coordinate attention) 注意力机制模块, 该模块能够精准定位小目标, 让检测头更容易聚焦于关键区域. 最后, 采用归一化高斯 Wasserstein 距离 (NWD) 损失来替代 CIoU. 这种替代降低了模型对位置偏差的敏感性, 进一步提升了检测效率. 实验数据显示, 与原始模型相比, 该模型的 *mAP* 达到了 95.7%, 较 YOLOv7 基础模型提升了 5.2%; 而参数量仅为 12.0M, 下降了 67.7%. 这些优化改进在保证高精度的同时, 大幅减少了参数量, 充分验证了基于 YOLOv7 的红外小目标检测模型的实用性, 使检测性能得到了显著提升.

关键词: 目标检测; 红外小目标; 轻量化模型; HIT-UAV 数据集; 无人机目标检测

引用格式: 苏健,张莉莉.UAV-YOLO: 基于 YOLOv7 的红外小目标检测.计算机系统应用. http://www.c-s-a.org.cn/1003-3254/10072.html

UAV-YOLO: Infrared Small Target Detection Based on YOLOv7

SU Jian, ZHANG Li-Li

(School of Software, Nanjing University of Information Science & Technology, Nanjing 210044, China)

Abstract: In the field of infrared small target detection for unmanned aerial vehicle (UAV), complex ground backgrounds and the inherently small size of targets often result in missed detections or false alarms in detection models. To address this problem, this study proposes a lightweight and high-precision infrared small target detection algorithm based on the YOLOv7 framework, termed unmanned aerial vehicle-you only look once (UAV-YOLO). First, given that the targets of interest are predominantly small, the 1×1 convolutions in the ELAN, ELAN-W, and CARAFE modules of the YOLOv7 backbone network, as well as those in the neck network, are replaced with GSConv. At the same time, the P5 detection head, which provides limited efficiency in infrared small target detection, is removed, and a newly added P2 detection head is designed specifically for small targets. These modifications not only enhance detection efficiency but also substantially reduce the number of parameters, thus achieving a lightweight model. Second, an improved SPPFCSPC pyramid pooling module is integrated into the backbone network. The inclusion of this module effectively expands the receptive field of the model, thus improving the detection accuracy of infrared small targets. Subsequently, the content-

① 收稿时间: 2025-07-16; 修改时间: 2025-08-13, 2025-08-22; 采用时间: 2025-09-09; csa 在线出版时间: 2025-12-19

aware reassembly of feature (CARAFE) module is incorporated into YOLOv7. This module enhances the preservation and optimization of feature representations for small targets. Furthermore, a coordinate attention (CA) mechanism is introduced before the detection head, enabling more precise localization of small targets and allowing the detection head to focus on key regions. Finally, the normalized Wasserstein distance (NWD) loss is adopted to replace the CIoU loss. This replacement reduces the sensitivity of the model to positional deviations and further improves detection performance. Experimental results demonstrate that compared with the baseline YOLOv7 model, the proposed UAV-YOLO achieves an mAP of 95.7%, representing a 5.2% improvement, while the number of parameters is reduced to 12.0M, a decrease of 67.7%. These improvements ensure high accuracy while significantly reducing the number of parameters, thus verifying the effectiveness and practicality of the proposed infrared small target detection model for UAV applications.

Key words: target detection; infrared small target; lightweight model; HIT-UAV dataset; unmanned aerial vehicle target detection

1 介绍

红外目标检测是早期预警、军事监视、红外搜索与跟踪 (IRST)、遥感和医学成像等许多实际应用中的重大问题之一, 具有不受环境等因素干扰的优势. 相较于可见光, 红外成像在面对雾、烟等大气障碍物时所受影响极小, 能够在各种复杂的天气条件下为人们提供清晰的图像, 然而, 目标的尺寸非常小, 为 2×2–9×9 像素, 小于整个图像的 0.15%[1]. 由于信号杂波比 (SCR) 较低, 小目标会被掩埋. 此外, 由于成像距离大和大气特性, 小目标的尺寸和纹理信息无法获得, 这使得检测过程变得相当困难. 近年来, 许多研究人员对红外目标检测表现出了浓厚的兴趣, 并提出了很多建议.

目标检测的核心就在于进行有效的目标分离与干扰 (背景及噪声), 这依赖于对图像多维度特征的判别性提取. 其中, 空间域 (如纹理、形状) 和频率域 (如频谱分布) 特征可从单幅图像中获取; 而刻画目标动态变化的时域特征, 则必须通过分析连续多帧图像间的时序关系才能有效建立. 目前红外小目标检测常用的单帧方法有滤波法、局部对比度法、频域变换法等. 当背景复杂或目标 SNR (信噪比) 较低时, 往往会导致更多的误报, 需要更多的参数调整.

在目标检测技术领域, 深度学习方法凭借其独特优势获得了广泛关注, 检测效率与识别准确度均实现了质的飞跃. 当前, 像 Faster R-CNN[2]、YOLO[3]、SSD[4] 这些典型的目标检测网络, 在不同类型的数据集测试中都呈现出优异的性能表现. 不过, 当面对背景干扰较为复杂的应用场景时, 红外弱小目标的检测效果往往会受到明显制约. 这就需要针对红外小目标自身特征匮乏、有效信息不足且背景环境复杂的特性展开更为深入的探索与研究.

应用深度学习解决红外弱小目标检测问题, 已成为该领域一个显著增长的研究趋势. Ju 等人[5]提出了一种图像滤波模块来获取置信图, 旨在增强红外小目标的响应并抑制背景的响应. Huang 等人[6]使用多个精心设计的局部相似度金字塔模块来提高红外小目标多尺度特征的捕获能力. Zhang 等人[7]提出了一种在红外跟踪中从 RGB 数据生成合成 TIR 数据的方法. Wu 等人[8]进行了创新性改变, 将弱小目标识别问题重构为分布模式分类任务, 构建了基于全卷积机制的分类网络框架. Ding 等人[9]则通过融合管道滤波技术与自适应感受野机制, 开发出多尺度形变卷积 SSD 架构, 显著提升了序列图像中的检测鲁棒性. 目的是解决红外与可见光图像的域差异问题. Zhang 等人[10]引入经典图像增强算子, 实现红外特征向 RGB 空间的渐进式对齐. 而 Chen 等人[11]另辟蹊径, 建立了基于高维张量建模与分解 (TCD) 的红外弱小目标检测理论体系. 该方法引入三阶张量构建和分解, 充分利用邻域空间信息, 同时设计了不连续的多尺度窗口以实现更稳健的检测性能. 在主流目标检测网络上进一步提高检测精度但增加训练难度[5].

针对无人机红外影像中小目标检测精度与召回率不足的问题, 本文设计并实现了一种改进型 UAV-YOLO 网络模型. 本文的主要贡献如下.

(1) 采用 GSConv 轻量化卷积替代原有网络结构 ELAN 模块的 1×1 卷积和颈部网络中的 1×1 卷积, 在保留跨层特征融合能力的基础上, 大幅削减了模型的

参数量和计算需求.

(2) SPPFCSPC (spatial pyramid pooling with feature concatenate spatial convolution) 模块通过采用更快速的池化替代方式以及对整体结构在计算层面的优化, 显著降低计算量, 使得模型能够在较短时间内完成对红外小目标的特征提取、处理任务.

(3) 利用内容感知特征重组 (CARAFE) 上采样聚合大感受野内的上下文信息, 丰富小目标的特征表示, 更好地捕捉目标与周围背景的关系, 突出小目标的特征表现.

(4) 在不同检测头前添加 CA (coordinate attention) 注意力机制, 不仅能够高效捕获通道间的信息交互, 还兼顾了特征空间中的位置信息. 它能够融合深浅层特征, 捕获小目标的低级结构和纹理特征, 提高红外小目标检测效率, 降低误报率.

(5) 对检测头进行重构, 设计了一种轻量级网络架构 UAV-YOLO, 提高了对红外小目标的检测能力.

(6) 将归一化高斯 Wasserstein 距离 (NWD) 引入红外小目标检测, 取代原有 CIoU 的计算损失, 降低微小目标检测中的假阳性率.

2 相关工作

2.1 YOLOv7

在深度学习领域, 目标检测是极为关键的研究方向, 往往作为计算机视觉系统的核心模块存在. 从车辆行人识别到自动驾驶场景应用, 再到医学影像分析等领域, 均离不开目标检测技术的支撑. 本研究选取 YOLOv7 作为基础检测框架, 该算法在性能表现上处于深度学习方向的前沿水平. 其核心工作体现在: 通过对梯度传播路径进行深度剖析, 系统性地探索了适用于不同网络层的模型重参数化方法, 并以此为基础构建了结构化的重参数化网络架构. 设计了一种基于粗到细策略的新型标签分配机制[12].

YOLOv7[13]的算法模型的主要结构与其之前的系列 YOLOv5[14]比较相似. YOLOv7 算法的整体架构由输入层、骨干网络、颈部网络和检测头这 4 个关键部分构成. 其中, 骨干网络作为核心特征提取模块, 负责对输入数据进行初级特征的抽取, 最终输出包含不同尺度信息的特征图集合, 这些输出的特征层将作为后续网络处理的基础, 因而被称作“有效特征层”. 值得关注的是, YOLOv7 骨干网络采用的 E-ELAN 模块通过四分支的特殊堆叠设计, 构建了更为密集的残差连接结构, 这种架构不仅简化了网络优化过程, 还为通过加深网络层次提升检测精度提供了可能. 颈部网络在 YOLOv7 中承担着特征融合的重要功能, 该模块的设计充分考虑了深度学习目标检测任务中目标尺度变化大、图像噪声干扰等实际问题. 在继承原有 PANet 结构的基础上, YOLOv7 的颈部网络通过增加额外的下采样操作, 进一步强化了不同层级特征之间的交互与融合, 有效提升了网络对多尺度目标的适应性. 检测头部分则主要负责完成目标的分类与定位任务. 在 YOLOv7 中, 骨干网络与特征金字塔网络 (FPN) 输出的有效特征层共同作为检测头的输入. 检测头针对每个层级的特征图均配置独立的参数组, 通过将特征图映射为特定的检测点, 并与对应的预测框进行关联, 实现对目标位置和类别的判断. 具体实现上, YOLOv7 沿用了经典的解耦头设计, 通过 1×1 卷积分别完成分类和回归任务, 这种设计在保证检测精度的同时, 有效降低了计算复杂度.

为了预测与前一帧匹配的目标情况, 整个 YOLOv7 网络执行以下任务: 输入图像处理、特征提取、功能改进和预测[15].

但是, 这种逐层传递式的结构也会导致各层之间信息传递的损失. 另外, 由于 YOLOv7 模型检测框的结构是固定的, 所以无论是对于小目标还是尺寸变化较大的多目标检测都存在一定的缺陷. 其自带的 SPPCSPC 模块无论是在处理信息的速度还是准确性上也存在一些不足, 因此, 我们设计了一种带有改进的 SPPFCSPC 模块的新网络模型结构, 如第 3.1 节所述; 除此之外, 针对红外小目标的特性与检测要求, 本文对基础网络结构进行了相关内容的改进, 具体优化的内容如第 3 节所述.

2.2 红外小目标检测

红外小目标检测是计算机视觉领域中的一个重要任务, 特别是在军事侦察、安全防范、智能监控等领域发挥着不可或缺的作用. 该技术聚焦于红外成像场景下微小目标的识别与定位, 这类目标在图像中往往仅占少数像素, 加之其尺寸微小、热辐射信号微弱的特性, 导致检测过程面临特征提取困难、背景干扰复杂等多重技术挑战, 目标的尺寸非常小, 为 2×2–9×9 像素, 小于整个图像的 0.15%[16]. 红外小目标检测的核心问题包括目标增强、背景抑制和特征提取. 在红外图像中, 目标物体往往呈现为微小的像素点, 加之背景

环境的复杂性, 使得准确检测这些小目标变得颇具挑战. 更进一步, 由于较大的成像距离和大气条件的影响, 目标的细节和纹理特征难以捕捉, 这进一步增加了检测工作的难度, 使得检测过程变得相当困难[17]. 例如, 红外图像缺乏物理信息, 对比度较低, 小物体很容易被淹没在复杂的背景中.

为攻克无人机航空红外小目标检测难题, 研究者们提出了多种方法. 比如说背景抑制法就是通过构建背景模型降低背景噪声、凸显目标信号, 常用算法包含帧差分法、背景差分法、高阶统计法; 基于滤波的方法涵盖空间域滤波 (例如高通滤波、边缘增强滤波) 与时间域滤波 (如卡尔曼滤波、粒子滤波); 基于模型的方法则是通过搭建目标物理或数学模型开展检测, 像基于光流、热辐射特性的目标检测; 以卷积神经网络 (CNN) 为代表的深度学习技术, 在目标检测领域成果显著, 其具备强大的特征提取能力, 对复杂环境适应度高且检测精度出色; 多传感器融合方法将红外成像与可见光、雷达等其他传感器信息融合, 综合各传感器优势, 降低虚警与漏检率. 随着技术的持续进步, 新的检测方法不断涌现, 例如 HCF-Net[18]这一深度学习方法, 借助并行化斑块感知注意力 (PPA)、维度感知选择性融合 (DASI)、多稀释通道细化器 (MDCR) 等实用模块, 该方法大幅优化了红外弱小目标的检测效能.

以上方法中, 先检测后跟踪方法不需要图像序列以及目标速度和形状的信息. 因此, 在每一帧数据处检测到目标的鲁棒性不强. 跟踪前检测方法的缺点是在低信噪比的情况下表现不佳[19].

2.3 注意力机制

近年来, 注意力机制在深度学习中的应用越来越多, 其中在图像检测、语音处理、自然语言等方面表现优异[20–23]. 在计算机视觉领域, 注意力机制也产生了重大影响, 特别是在目标识别方面. 其中, 挤压激励网络 (SENet)[22]和卷积块注意力模块 (convolutional block attention module, CBAM)[23]显示了最传统的性能. SENet 作为经典的注意力机制模型, 其核心在于引入“挤压-激励 (squeeze-and-excitation, SE)”模块, 以此提升模型对特征的表征能力. 该模块通过学习每个特征通道对应的权重系数, 实现对通道重要性的动态调整, 使得网络能够自适应地聚焦关键特征. 尽管 SENet 额外引入的参数较少, 在模型训练与工程部署上具备轻量化优势, 但由于需要对各通道权重进行独立计算, 不可避免地会带来额外的计算开销, 在一定程度上增加了模型的运算复杂度.

卷积神经网络 (CNN) 可集成卷积块注意力模块 (CBAM), 该机制能有效提升模型的表征学习与泛化性能. CBAM 的核心优势在于其融合了通道域与空间域的双重注意力策略. 该策略引导网络对输入特征图进行动态校准, 强化关键信息区域的权重, 同时弱化无关或冗余特征的干扰, 从而显著优化特征提取过程. 以空间为中心丰富特征空间, 只考虑局部区域的数据, 无法判断长程依赖, 缺乏各级空间信息的明显缺点[24].

在本文中, 我们在 YOLOv7 的 head 模块中添加了 CA 注意力机制, 它不仅可以通过融合深层和浅层特征, 捕获小目标的低级结构和纹理特征, 提高红外小目标的检测概率, 还能够增强目标特征, 同时抑制背景干扰, 提升特征提取能力, 降低复杂背景下的误报率, 有效提高了检测的准确性和鲁棒性.

3 方法与技术

3.1 UAV-YOLO 网络结构

本文基于 YOLOv7 提出一种改进的网络结构 UAV-YOLO, 并利用 UAV-YOLO 完成无人机红外小目标检测. 如图 1 所示. 考虑到轻量化需求, 将 ELAN 模块以及颈部网络中的 1×1 卷积替换成轻量化卷积 GSConv, 并在颈部网络将 SPPCSPC 替换成 SPPFCSPC 模块来加快特征处理速度, 最大程度保留小目标特征; 采用具有特征重组的上采样 CARAFE 模块, 使模型更准确定位和识别红外小目标, 生成自适应卷积核, 提高检测精度; 在各个检测头前引入 CA 注意力机制不仅可以高效地捕获通道间的信息交互, 还兼顾了特征空间中的位置信息; 在 head 部分对检测头进行了优化重构, 删除了检测大目标的 P5 检测头, 添加了主要用来检测小目标的 P2 检测头, 利用得到的多尺度特征图通过 3 个不同尺度的检测头 (P2, P3, P4) 得到预测结果. 最后将 CIoU 替换为更适合小目标的归一化高斯 Wasserstein 距离 (NWD), 通过度量分布差异提升小目标检测鲁棒性, 并降低了对小目标位置偏差的敏感性. 该策略在减少计算开销的同时, 显著提高了红外小目标的检测效率与定位准确性.

3.2 轻量化改造

针对无人机目标检测场景中终端算力有限 (如嵌入式芯片算力、内存约束) 及续航要求高的特点, UAV-

YOLO 从底层逻辑出发进行轻量化改造, 通过针对性优化降低计算复杂度, 使模型在保持高精度检测效果的同时尽量保持低算力和计算复杂度, 满足检测场景和目标的多维度需求.

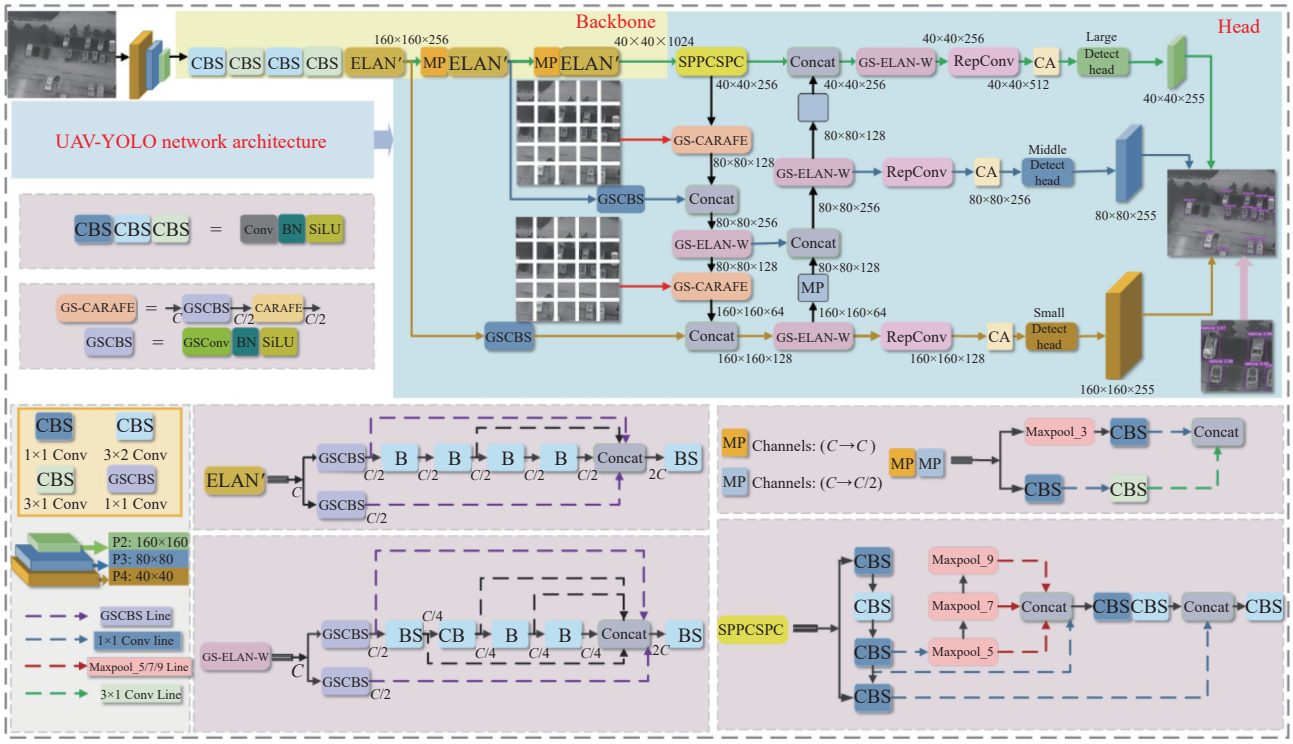


图 1 UAV-YOLO 网络结构

首先, 传统 1×1 卷积的参数量计算方式为 $C_1 \times C_2$ (其中 C_1 为输入通道数, C_2 为输出通道数), 其通过全局特征映射实现通道维度调整, 但参数量随通道数增长呈线性增加. 为此, UAV-YOLO 中 GSConv 采用“标准卷积+深度可分离卷积+通道混洗”的混合结构, 将输入通道分为两组: 一组通过标准卷积完成下采样 (参数量 $C_1 \times (C_2/2)$), 另一组通过深度可分离卷积 (DWConv) 进行特征融合 (参数量 $(C_2/2) \times (C_2/2) \times 1 \times 1$), 因 DWConv 的卷积核尺寸为 1×1, 最终通过通道混洗重组特征. 将参数量从传统 1×1 卷积的 $C_1 \times C_2$ 降至 $(C_1 \times C_2/2) + (C_2^2/4)$, 在保留特征融合能力的同时减少冗余计算, 具体的方法介绍在第 3.3 节. 在检测头设计上, 通过重构检测头对红外小目标的特性进行更好的适配, 不仅能大大提升检测精度, 还能大幅去除无效计算, 实现轻量化的效果. 具体的检测头优化细节在第 3.4 节.

改进后的模型仅替换 ELAN 模块和颈部网络的 1×1 卷积为 GSConv, 保留 ELAN 的跨层连接和 CARAFE 的内容感知上采样, 参数量降低的同时且未损失特征融合能力; 除此之外, 对检测头的优化调整也使得网络结构实现了大幅度的参数量减少, 同时还能更好的保留小目标细节, 避免精度损失. 而针对其他的添加模块也没有对整个网络模型的计算复杂度有影响, 在轻量化改造后, UAV-YOLO 的参数量仅为原始 YOLOv7 模型的 32.3%, 可适配无人机终端的内存限制, 且满足实时推理需求. 同时, 计算量的减少直接降低了终端功耗, 延长了无人机续航时间, 使其更适应复杂场景下的长时间任务. 更具体的实验效果可以通过第 4.5 节的消融实验数据体现.

3.3 GSConv

为优化深度可分离卷积 (depthwise separable convolution, DSC) 的输出性能, 使其尽可能逼近标准卷积 (standard convolution, SC) 的效果, 从而在轻量级模型中实现更高的精度并满足实时检测需求, Li 等人[25] 设计了一种新的轻量卷积操作: GSConv. GSConv 的架构设计融合了标准卷积、深度可分离卷积与通道混洗技术, 其具体结构如图 2 所示.

在运算流程上, 该模块首先运用常规卷积完成下采样操作, 进而借助深度可分离卷积 (DWConv) 对标准卷

积 (SC) 与深度可分离卷积 (DSC) 的输出特征进行融合, 最终通过通道混洗实现特征通道间的高效重组. 值得注意的是, 该模块中的“Conv”部分整合了卷积层、批归一化处理层 (batch normalization) 及激活函数层构成, 而“DWConv”则专指深度可分离卷积运算过程. 相较于标准卷积, GSConv 显著减少了参数量和计算复杂度.

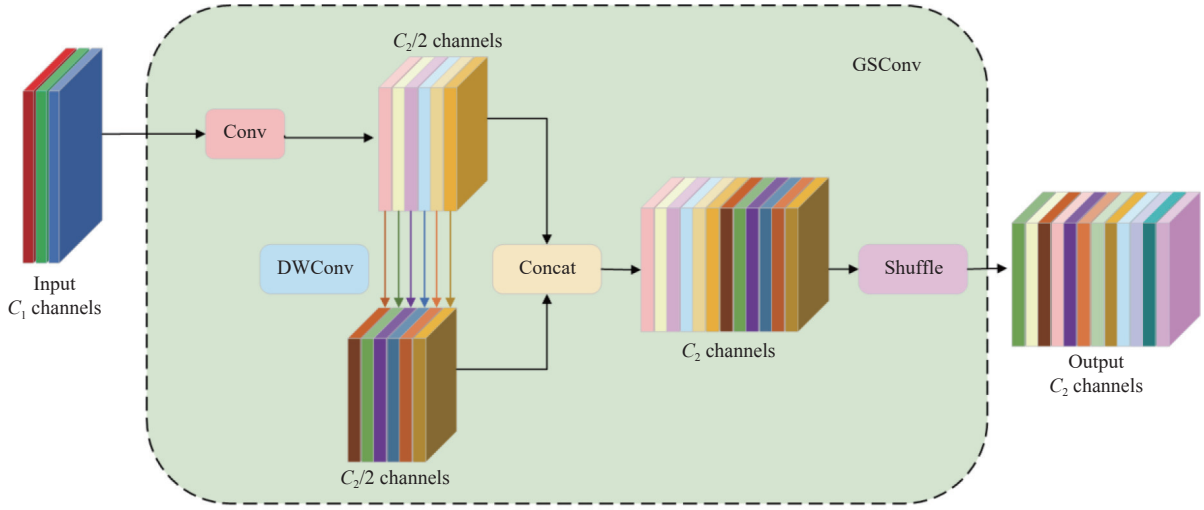


图 2 GSConv 卷积结构

针对无人机红外弱小目标检测对模型轻量化的具体要求, 本研究引入轻量级 GSConv 卷积模块实施网络优化. GSConv 充分结合标准卷积的全局特征提取优势与深度可分离卷积的高效计算特性, 在保证检测精度的基础上, 有效简化了模型架构, 大幅提升了实时检测性能. 为规避因全卷积替换引发的性能衰减问题, 研究采用选择性替换策略, 仅将 ELAN 模块及颈部网络中的 1×1 卷积层替换为 GSConv 模块. 该方案不仅削减了冗余计算量, 还强化了注意力机制对核心特征的捕捉能力, 进而实现检测精度的提升. 实验结果显示, 该优化设计成功兼顾轻量化需求, 显著增强了无人机红外小目标检测的准确性与实时处理能力.

3.4 P2

对在无人机图像中实现高精度红外小物体检测是计算机视觉研究中普遍存在的一个关键难题, 特别是在物体检测任务中[26]. 尽管 YOLOv7 在多场景目标检测任务中表现卓越, 但在红外弱小目标检测方面仍存在性能瓶颈. 原模型配备 P3、P4、P5 这 3 级检测头, 分别对应小、中、大尺度目标检测. 由于红外小目标自身信噪比低, 特征信号微弱且语义信息不明显, 在多层卷积运算过程中极易出现特征衰减或丢失. 本研究在 YOLOv7 框架基础上, 将 P5 检测头替换成 P2 小目标检测头, 以此强化特征融合效果, 有效保留小目标语义信息. P2 检测头输出特征图分辨率达 160×160 像素, 凭借更高的空间分辨率, 能够更精准地捕捉小目标细节特征, 为小目标定位提供更丰富的空间信息, 从而显著提升检测准确率. 此外, P2 检测头的加入完善了模型的多尺度特征交互机制, 促进了不同分辨率特征层间的信息流通, 实现了多尺度特征的深度融合, 使模型在处理不同尺寸目标时具备更强的适应性和检测能力.

P2 检测头输出的高分辨率特征图, 通过跨感受野的信息融合机制, 为模型捕捉红外弱小目标提供了关键技术支撑, 有效优化了模型对小目标的召回率与定位准确度. 该机制同步提升了模型对不同尺度目标的鲁棒性表现, 显著增强了 YOLO 框架在复杂多尺度检测场景下的综合效能. 它是 YOLO 系列模型中用于提升小目标检测性能的一个重要组件, 通过在网络的较浅层增加检测能力, 有效地提高了模型对小目标检测的精确度和鲁棒性.

UAV-YOLO 使用轻量级网络架构的 3 个尺度, 调整网络输出特征图的尺度至 40×40、80×80 及 160×160 像素规格. 优化特征提取模块后, 特征融合部分也需要相应优化. 融合生成的特征映射尺度为原图的 1/4、1/8 与 1/16. 这种较大尺度的特征保留策略, 显著强化了浅层网络的高分辨率线索与初级语义表征的贡献度, 有

效缓解了因深度下采样引发的红外弱小目标热信号特征退化现象, 具体的完整网络结构见图 1.

3.5 CARAFE

在深度学习领域, 特征上采样是目标检测、实例分割以及场景解析等密集预测任务所涉及的卷积网络架构中的核心算子. 就目标检测任务而言, 内容感知特征重组 (content-aware reassembly of feature, CARAFE) 发挥着关键作用. 在 YOLOv7 模型中引入 CARAFE 算子, 能够高效整合来自广泛感受野的信息, 显著提升模型在各类相关任务中的性能表现[27].

在 YOLOv7 模型体系的优化过程中, CARAFE 模块的引入显著提升了目标检测性能. 该模块以承载高阶语义信息的下采样特征图为输入, 通过子像素卷积操作, 将特征图分辨率提升至任务需求水平. 相较于传统卷积, CARAFE 采用内容感知的特征重构策略, 通过学习权重参数, 对相邻特征进行加权聚合, 动态融合局部上下文信息. 这一过程基于输入特征的局部内容, 自适应生成卷积核系数, 对目标像素邻域内的特征进行选择性聚合, 确保在特征上采样时既能保留图像细节, 又能维持语义信息的完整性. 每个输出像素均对应一组由输入内容决定的多核权重, 最终通过特征重排形成结构化的空间块, 为后续检测任务提供高质量的特征表征.

CARAFE 上采样算子具有以下几个特点.

(1) CARAFE 具有较大的感受野. 传统方法通常使用最近邻插值和双线性插值, 这将模型的感受野限制为 1×1 或 2×2. 然而, CARAFE 在上采样过程中考虑了整个特征图, 这扩大了模型的感受野, 从而更好地保留图像细节和边缘信息, 同时减少上采样过程中出现的锯齿状边缘和模糊效果.

(2) CARAFE 通过动态生成适合不同目标的自适应内核来支持实例特定的内容感知处理. 但最近邻插值和双线性插值依赖于固定的上采样规则, 无法根据内容调整核的大小.

(3) CARAFE 引入的计算开销很小, 并且对现有网络模型具有良好的适应性[28]. CARAFE 可以无缝集成到深度神经网络的任何位置. 与最近邻插值或双线性插值相比, CARAFE 更容易与其他网络层 (例如卷积层或池化层) 结合.

CARAFE 模块的结构如图 3 所示.

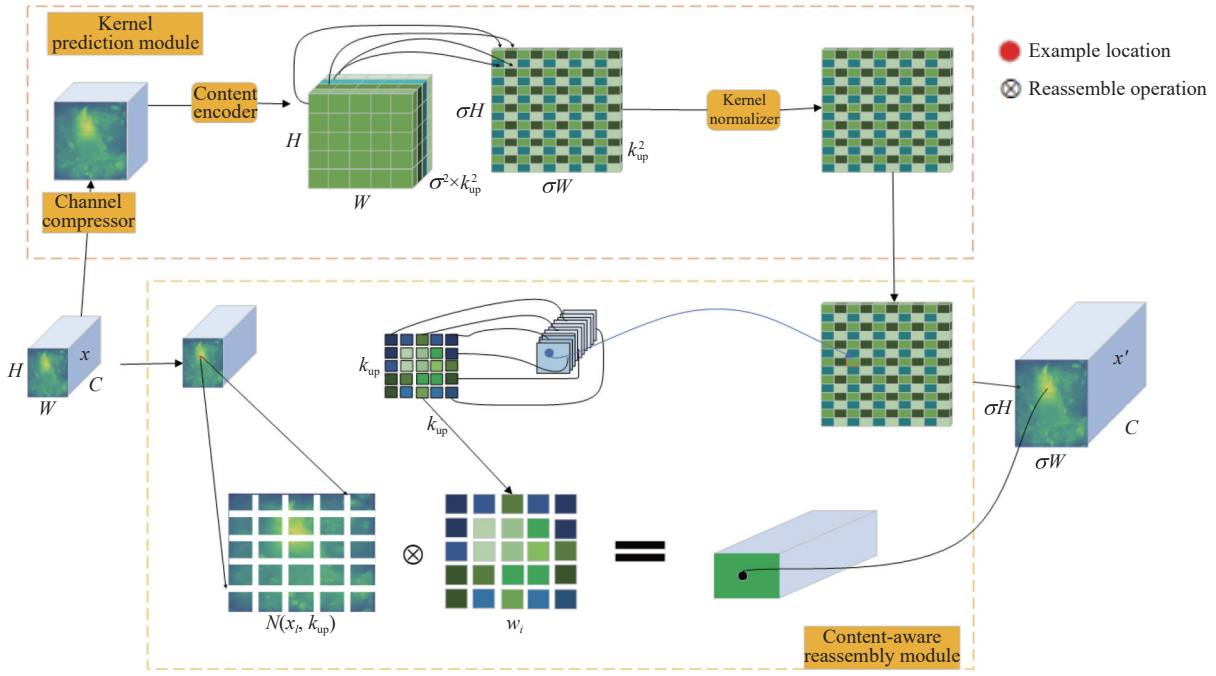


图 3 CARAFE 模块结构

3.6 SPPFCSPC

在目标检测领域, SPPCSPC (spatial pyramid pooling, cross stage partial channel) 模块是 YOLOv7 的核心部分, 通过不同大小的最大池化操作来扩展感知领域, 以处理不同分辨率的图像. SPPCSPC 包含了两部分: 空间金字塔池化 (spatial pyramid pooling, SPP) 和跨阶段

部分通道连接 (cross stage partial channel, CSPC). SPP 结构通过分层池化不同空间子区域的特征, 生成尺度统一的特征表示, 有效压缩了后续网络层的计算需求. CSPC 结构将特征流拆分为卷积分支与恒等映射分支, 前者进行特征变换, 后者提供跨层连接. 这种通道分流机制显著减少了需进行卷积处理的通道数量, 协同优化了模型的计算效率. 通过 SPPCSPC 模块的优化, YOLOv7 模型可以在保持较高精度的情况下, 减少一半的计算量. 如图 4 所示, 在 SPPCSPC 模块的第 1 个分支中采用大小为 5、9、13 和 1 的最大池化层来辨别目标尺寸. 然而, 它有一个显著的缺陷: 与通道号不对齐的目标可能会被忽略, 导致特征信息丢失并影响检测精度, 特别是对于红外小目标的检测. 此外, 将特征图划分为 4 个分支进行池化操作会降低 YOLOv7 算法的处理速度.

本文提出的 SPPFCSPC 对此做了不同的改进. SPPFCSPC 包含了两个子模块: 空间金字塔池化 (SPP) 和全连接空间金字塔卷积 (fully connected spatial pyramid convolution, FCSPC). SPP 模块对输入特征图进行多尺度空间金字塔池化, 得到不同尺度的特征图, 捕获不同尺寸的目标和场景信息. FCSPC 模块对 SPP 模块输出的多个不同尺度的卷积核所对应的不同尺度的特征图进行卷积运算, 并将输入的特征图划分为块[29]. YOLOv7 中的 SPPFCSPC 模块通过集成 4 个池通道改进了以前的 SPPF 结构. 这种集成允许将池化结果反馈至最终层和下一个池化层, 从而增强复杂环境中的特征保留和小目标检测. 连续池化过程还提高了处理效率和模型的整体速度. 除此之外, SPPFCSPC 模型的参数数量比 SPPCSPC 更少, 在保持高性能的同时还能更高效. SPPFCSPC 模型如图 5 所示.

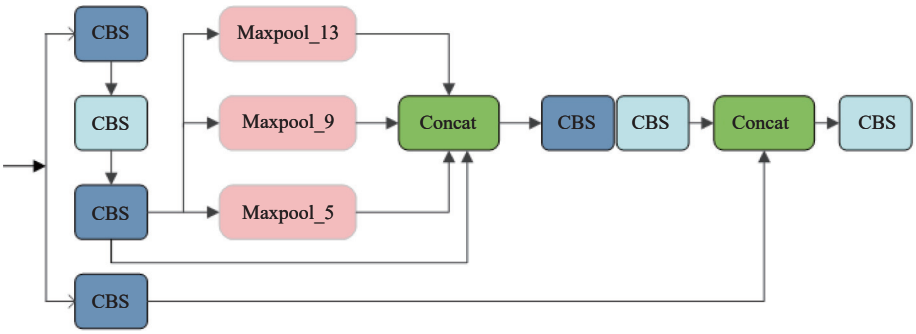


图 4 SPPCSPC 模块结构

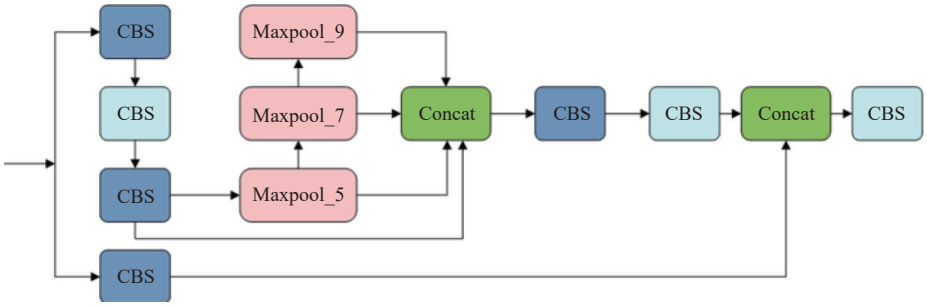


图 5 SPPFCSPC 模块结构

3.7 CA 注意力机制

现有注意力机制研究中, SE 模块因其显著的计算密集性面临应用瓶颈. 相比之下, CBAM 虽整合了通道与空间维度, 但其核心注意力机制仍过度偏重空间特征关联, 导致整体性能提升受限. 为了克服这些局限性, 本文引入协调注意力机制 CA (coordinate attention) 以实现高效的移动网络设计来增强模型效果. 这种 CA 模块不仅能够高效地捕获通道间的信息交互, 还兼顾了特征空间中的位置信息. 通过精确编码通道相关性和长距离依赖性的位置信息, CA 模块有效地解决了 SE 和 CBAM 模块所面临的挑战. 这不仅显著降低了漏检和误检率, 还提升了目标定位和识别的准确性[30].

传统通道注意力依赖全局池化压缩空间信息为通道标量, 导致关键空间结构严重损失, 极大制约了弱小

目标的识别性能. 为此, 本文引入坐标注意力机制 (CA), 将全局池化解耦为沿正交空间维度的独立一维聚合, 分别通过宽度与高度方向的定向特征编码, 保留位置敏感的上下文信息. CA 模块如图 6 所示.

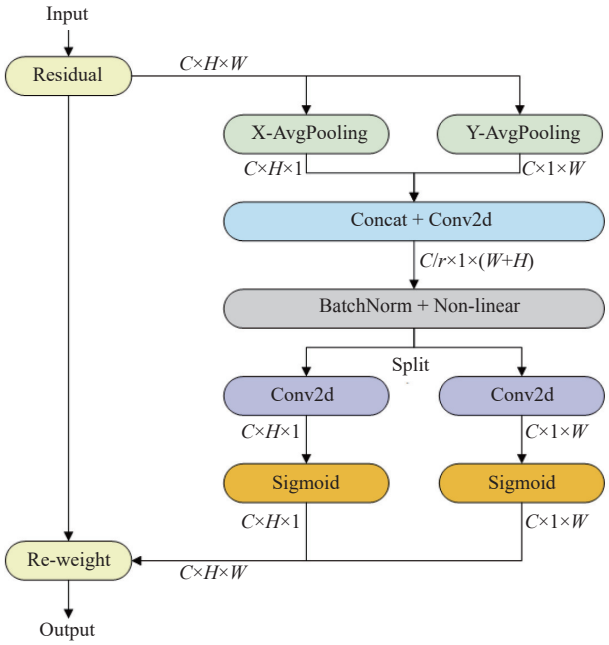


图 6 CA 注意力机制结构

为了赋予注意力模块位置感知的长程建模能力, 将全局池化替换为式 (1) 的正交方向特征压缩: 沿水平轴采用 (H, 1) 核进行通道响应聚合, 垂直轴则通过 (1, W) 核独立执行特征编码, 实现输入 X 的空间结构解耦表示. 因此, 第 c 个通道在高度 h 处和宽度 w 处的输出分别可以表示为:

$$z_c^h(h)=\frac{1}{W}\sum_{0\in i<W}x_c(h,i) \tag{1}$$

$$z_c^w(w)=\frac{1}{H}\sum_{0\notin j<H}x_c(j,w) \tag{2}$$

然后, 将两个空间定向的特征图拼接在一起, 使用大小为 1×1 的卷积核缩小尺寸, 并通过批处理和 Sigmoid 函数进行非线性化:

$$f=\delta\left(F_1\left[z^h,z^w\right]\right) \tag{3}$$

特征图 f 的大小计算为 $C/r\times1\times(W+H)$, 其中 r 为压缩因子. 然后, 通过 11 个卷积核对 f 进行上维化, 以匹配沿垂直和水平轴在两个方向上被分解成张量后的通道总数. 采用 Sigmoid 激活函数处理后, 分别产生高度方向和宽度方向的注意力权重. 其过程如图 6 所示.

$$g^h=\sigma\left(F_h\left(f^h\right)\right) \tag{4}$$

$$g^w=\sigma\left(F_h\left(f^w\right)\right) \tag{5}$$

最后, 将得到的注意力权重与原始特征图 $X(i,j)$ 相乘, 得到具有注意力权重的特征图 $Y_c(i,j)$:

$$Y_c(i,j)=X_c(i,j)\cdot g_c^h(i)\cdot g_c^w(i) \tag{6}$$

将两个特征图相乘得到的新特征图同时考虑了空间特征和通道特征, 可以更有效地提高算法模型对小目标检测的性能[31].

3.8 归一化高斯 Wasserstein 距离

在训练检测红外小目标的网络模型时, 损失函数起着至关重要的作用. 不同的损失函数会对模型的训练结果产生不同的影响[32]. YOLOv7 的损失函数如式 (7) 所示, 由 3 个部分组成: 置信损失L_{conf}、分类损失L_{cls}和定位损失L_{loc}[33]. L_{conf}和L_{cls}都是通过 BCE 结合 logits 损失函数计算的, L_{loc}是通过 CIoU (complete intersection over union)[34]计算的. 式 (8) 是带有 logits 损失函数的 BCE 计算公式, 式 (9) 是 CIoU 计算公式. 传统的损失函数在衡量预测框与真实框的匹配程度时, 多采用常规的 IoU 或基于距离的度量方式.

$$Loss=L_{\text{conf}}+L_{\text{cls}}+L_{\text{loc}} \tag{7}$$

$$Loss_{\text{BCE}}=-\frac{1}{N}\sum_{i=1}^{N}\left[y_i\log(p_i)+(1-y_i)\log(1-p_i)\right] \tag{8}$$

$$Loss_{\text{CIoU}}=1-I_{IoU}+\frac{p^2(b,b^{\text{gt}})}{d^2}+\alpha\nu \tag{9}$$

在式 (8) 中, y_i表示样本的二元标签, p_i表示样本的概率, d 是同时包含预测框和真实框的最小封闭区域的对角线欧氏距离. 在式 (9) 中, b 表示预测边界框的质心, b^{gt}表示地面真实边界框的质心, p 表示 b 和b^{gt}之间的欧几里德度量. 通过式 (10)–式 (12) 来计算α、ν、I_{IoU}, 参数含义如图 7 所示.

$$\alpha=\frac{\nu}{(1-IoU)+\nu} \tag{10}$$

$$\nu=\frac{4}{\pi^2}\left(\arctan\frac{W_{\text{gt}}}{H_{\text{gt}}}-\arctan\frac{W}{H}\right)^2 \tag{11}$$

$$I_{IoU}=1-\left|\frac{A\cap B}{A\cup B}\right| \tag{12}$$

其中, A 表示真实边界框的面积, B 表示预测边界框的面积. 在图 7 中, 绿色区域代表真实边界框, 黄色区域

代表预测边界框. H_{gt}和W_{gt}是地面实况边界框的高度和宽度. H 和 W 是预测边界框的高度和宽度. H_g和W_g是同时包围预测边界框和真实边界框的最小矩形的高度和宽度. (x_{gt},y_{gt})和(x,y)分别是真实边界框和预测边界框的质心. H_i和W_i是真实边界框和预测边界框之间重叠的高度和宽度.

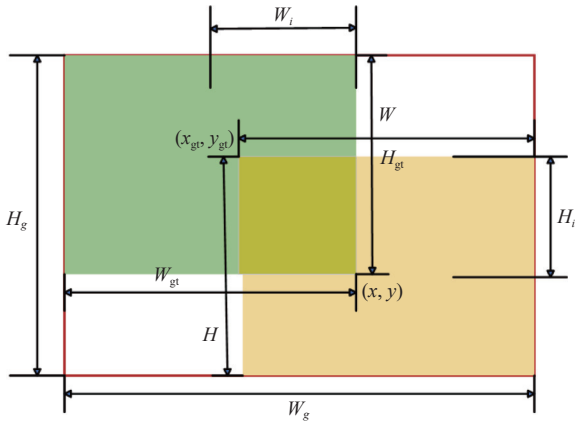


图 7 计算参数示意图

原始模型使用 CIoU 作为主要损失函数, 通过考虑重叠、中心距离和纵横比来提高精度. 然而, 多项研究分析表明, CIoU 对于红外小目标的定位不准确, 难以区分小目标与干扰物位置的差异, 收敛速度也较慢. 因此, 本研究引入归一化高斯 Wasserstein 距离 (NWD) 用于替换原始的 CIoU 计算损失, 以降低小目标位置偏差的灵敏度.

用于普通目标检测的 *IoU* 度量在用于红外小目标检测时表现出极高的灵敏度. 即使预测框和实际框之间的位置有轻微偏差, 也会导致 *IoU* 发生重大变化. *IoU* 度量对红外小目标的这种敏感性导致了训练过程中正负样本之间的高度相似, 给网络的有效收敛带来了挑战. 此外, 在极端情况下, 红外小目标可能只占用图像中的一个或几个像素. 因此, 基础真值与任何预测边界框之间的 *IoU* 低于最小阈值, 从而导致图像内的零阳性样本. 因此, 为了更准确地评估红外小目标检测模型, 需要替代评价指标.

IoU 度量实际上是样本之间的相似度计算, 对目标的尺寸变化较为敏感, 不适用于红外小目标, 因此引入 NWD 作为新的度量. Wasserstein 距离可以度量最小重叠或没有重叠的分布之间的相似性, 并且对不同尺度的对象不敏感. 因此, 该方法可以解决红外小目标训练过程中正负样本的相似性问题以及正样本稀疏问题. 具体而言, 边界框被建模为二维高斯分布, 如式 (13) 所示:

$$f(X\mid\mu,\Sigma)=\frac{\exp\left(-\frac{1}{2}(X-\mu)^{\mathrm{T}}\Sigma^{-1}(X-\mu)\right)}{2\pi|\Sigma|^{\frac{1}{2}}} \tag{13}$$

其中, X、μ 和Σ分别表示高斯分布的坐标(x,y)、均值向量和协方差矩阵.

当式 (14) 成立时,

$$(X-\mu)^{\mathrm{T}}\Sigma^{-1}(X-\mu)=1 \tag{14}$$

水平边界框$R=(c_x,c_y,\omega,h)$可以使用$N(\mu,\Sigma)$建模为二维高斯分布, 如式 (15) 所示:

$$\mu=\begin{bmatrix}c_x\\c_y\end{bmatrix},\ \Sigma=\begin{bmatrix}\frac{\omega^2}{4}&0\\0&\frac{h^2}{4}\end{bmatrix} \tag{15}$$

其中, (c_x,c_y)、ω 和 h 分别表示中心坐标、宽度和高度. 两个二维高斯分布$\mu_1=N(m_1,\Sigma_1)$和$\mu_2=N(m_2,\Sigma_2)$之间的二维 Wasserstein 距离定义为式 (16):

$$W_2^2(\mu_1,\mu_2)=\|m_1-m_2\|_2^2+\mathrm{tr}\left(\Sigma_1+\Sigma_2-2\left(\Sigma_2^{\frac{1}{2}}\Sigma_1\Sigma_2^{\frac{1}{2}}\right)^{\frac{1}{2}}\right) \tag{16}$$

式 (16) 可以简化为:

$$W_2^2(\mu_1,\mu_2)=\|m_1-m_2\|_2^2+\left\|\Sigma_1^{\frac{1}{2}}-\Sigma_2^{\frac{1}{2}}\right\|_F^2 \tag{17}$$

其中, $\|\cdot\|_F$为 Frobenius 范数. 由边界框$A=(cx_a,cy_a,w_a,h_a)$和$B=(cx_b,cy_b,w_b,h_b)$建模的高斯分布N_a和N_b之间的距离可以简化为式 (18):

$$W_2^2(N_a,N_b)=\left(\left[cx_a,cy_a,\frac{\omega_a}{2},\frac{h_a}{2}\right]^{\mathrm{T}},\left[cx_b,cy_b,\frac{\omega_b}{2},\frac{h_b}{2}\right]^{\mathrm{T}}\right)_2^2 \tag{18}$$

将其指数归一化到 0–1 的范围, 得到归一化高斯 Wasserstein 距离[35], 如式 (19) 所示:

$$NWD(N_a,N_b)=\exp\left(-\frac{\sqrt{W_2^2(N_a,N_b)}}{C}\right) \tag{19}$$

其中, C 是与数据集相关的常数.

4 实验

4.1 数据集

本文利用 UAV-YOLO 方法评估其在无人机红外小目标检测的效果. 由于 HIT-UAV 数据集[36]的公开可用性和全面的覆盖范围, 因此选择 HIT-UAV 数据集进

行对比实验. 该数据集对该领域做出了显著贡献, 因为它是第 1 个高空无人机红外数据集. 该数据集是专为无人机目标检测应用而设计的综合高空红外热成像数据集 (如图 8 所示).

HIT-UAV 数据集由 2 898 张分辨率为 640×512 的红外热图像组成, 包含 5 个类别 24 899 个标签, 即人、汽车、自行车、其他车辆和 DontCare. 这些图像是从无人机在学校、停车场、道路和操场等各种场景下捕获的 43 470 帧海量图像中提取而来, 并在不同的光照条件下 (包括白天和夜晚) 进行拍摄. 通过这种方式, 数据集不仅丰富了场景的多样性, 还确保了在不同光照条件下的广泛适用性, 尤其是夜视环境. 其真实背景使得模型在训练时能够更好地适应现实情况, 提高泛化能力. 使用 HIT-UAV 数据集可以提供多样化、真实、具有挑战性和可扩展的数据, 从而帮助模型更好地学习和适应红外目标检测任务.

图 8 HIT-UAV 数据集

为了能够使数据集更加干净有序, 有利于模型的训练和收敛, 提高模型的检测效率. 我们在该数据集的基础之上做了简单修改, 构建了简化数据集, 排除了 DontCare 类别, 并将汽车和其他车辆类别合并为一个称为车辆的类别. 这产生了 3 个类别, 即: 人、车辆和自行车, 总共 2 866 张图像, 并将其用于消融实验研究和对比实验研究. 在本文中, 数据集按照 7:2:1 的比例分为训练集、测试集和验证集这 3 部分. 这产生了包含 2 006 个图像的训练集、包含 574 个图像的测试集和包含 286 个图像的验证集. 该数据集的重要意义在于它对无人机红外目标检测与识别的研究和实现做出了贡献, 使其具有很强的学术研究和实际应用意义.

4.1.1 数据增强策略

在数据预处理阶段, 针对 HIT-UAV 数据集中的红外图像, 引入多尺度随机裁剪与缩放增强. 结合模型新增的 P2 检测头对高分辨率特征的需求, 将原始图像按 0.5–1.2 倍随机缩放后裁剪至 640×640 尺寸, 既保留小目标的细节特征, 又模拟无人机拍摄时目标距离变化导致的尺度差异, 与模型中 CARAFE 模块的内容感知上采样能力形成互补, CARAFE 通过动态生成卷积核优化特征重组, 而多尺度裁剪可增加模型对不同尺度目标特征的适应性.

同时, 引入随机高斯噪声与对比度扰动, 模拟红外成像中大气干扰与热噪声影响, 配合模型中 CA 注意力机制的背景抑制能力, 增强模型对低信噪比目标的特征捕捉能力. CA 模块通过编码空间位置信息聚焦目标区域, 而噪声扰动可迫使模型学习更鲁棒的目标本质特征 (如热辐射轮廓), 减少对背景噪声的依赖. 此外, 考虑到红外图像缺乏纹理信息的特点, 采用基于像素混叠的 CutMix 增强, 将不同场景的红外图像按随机比例拼接, 并保持目标标注的完整性. 该方法与模型中 GSConv 的轻量化特征融合机制相契合, GSConv 通过分组卷积与

通道混洗实现高效特征整合, 而 CutMix 增强可促使模型在特征融合过程中更精准地区分目标与背景的边界, 进一步提升 SPPFCSPC 模块的多尺度特征提取效率. 通过该数据增强策略, 能够在不增加模型参数量的前提下, 丰富训练样本的多样性, 使 UAV-YOLO 的特征提取模块 (如改进的 SPPFCSPC、CARAFE) 学到更具判别性的红外弱小目标特征, 进一步降低漏检与误检率.

4.2 实验平台和参数配置

本研究基于 Windows 10 操作系统、Python 3.9 和 PyTorch 环境开展实验, 具体硬件配置详见表 1. 训练过程中, 采用 epoch 为 500, 设置 batch size 为 16.

表 1 配置实验平台

名称	相关配置
GPU	NVIDIA GeForce RTX 4090 D
CPU	Intel(R) Core(TM) i9-14900KF
CUDA	12.4
操作系统名称	Ubuntu
深度学习架构	PyTorch

4.3 实验评估标准

为了评估改进的 UAV-YOLO 网络模型对红外小目标的检测效率, 本研究使用了以下关键指标: 精确率 (P)、召回率 (R)、$F1$ 分数 ($F1$)、平均精度 (AP)、平均精度均值 (mAP)、FLOPs 和参数量. $F1$ 分数是精度和召回率的综合衡量标准, 而 mAP 是模型识别准确性的最终衡量标准. FLOPs 是衡量计算机硬件性能和算法复杂度的指标. AP 和 mAP 评价指标中, mAP 用于通过对所有类别的平均精度 (AP) 值进行平均来评估算法的整体有效性, 本实验采用 $mAP@0.5$. 其中, 精确率计算如式 (20) 所示, 召回率计算如式 (21) 所示, $F1$ 分数计算如式 (22) 所示:

$$P = \frac{TP}{TP + FP} \tag{20}$$

$$R = \frac{TP}{TP + FN} \tag{21}$$

$$F1 = 2 \times \frac{P \times R}{P + R} \tag{22}$$

其中, TP 表示真阳性实例, 即被模型正确检测到的红外小目标数据; FN 表示假阴性, 即被模型错误识别为非目标, 但实际是被识别目标的数据; FP 表示假阳性, 即被模型错误检测到的数据. $F1$ 的值范围为 0–1, 其中 1 代表模型的最佳输出, 0 代表模型的最差输出. AP 和 mAP 计算如式 (23) 和式 (24) 所示:

$$AP = \int_0^1 P(R)\,\mathrm{d}R \tag{23}$$

$$mAP = \frac{1}{c}\sum_{j=1}^{c} AP_j \tag{24}$$

AP 值是由不同精确率 (P) 和召回率 (R) 点组合形成的曲线下面积. c 表示类别数.

参数量表示深度神经网络中学习所需的参数的总数. 在深度学习的框架内, 每个单独的神经元都拥有一个权重, 这些权重是通过训练过程迭代确定的. 深度神经网络中参数的总数是所有这些权重的合并. 这包括互连输入和输出的权重, 以及与所有神经元相关的偏置项.

4.4 比较算法

为了评估 UAV-YOLO 在无人机场景下对红外小目标检测的可用性和有效性, 在相同条件下与各种先进算法进行了对比研究. 所有实验模型在没有预训练权重的情况下进行训练, 并重新开始训练. 模型的输入图像大小调整为 640×640, 批次大小设置为 16, 迭代次数为 500 epochs. 训练结果表明, 与原始 YOLOv7 相比, UAV-YOLO 的 mAP 增加了 5.2%, 达到了 95.7% 的高精度, 同时参数量减少了 67.7%. UAV-YOLO 在检测红外条件下的小目标时, 模型复杂度显著降低, 精度在目前已有的模型中达到了最高的数值. 这些结果验证了在基于无人机场景下提出的红外小目标检测方法的有效性. 这项研究的结果如表 2 所示.

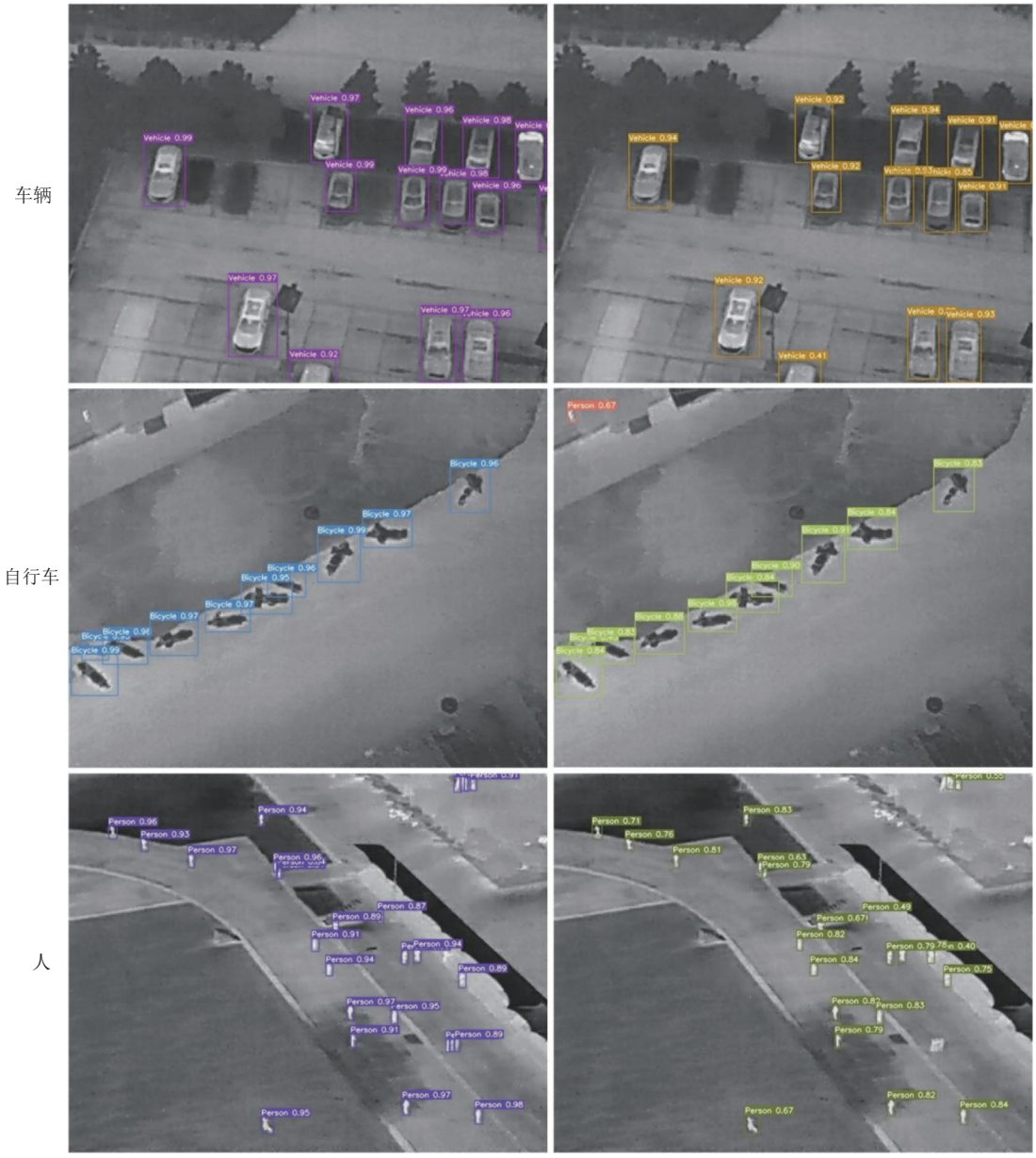
表 2 UAV-YOLO 和 YOLOv7 的性能比较

方法	参数量 (M)	FLOPs (G)	P (%)	R (%)	mAP (%)	$F1$ (%)
YOLOv7	37.2	105.1	88.1	85.3	90.5	86.2
UAV-YOLO	12.0	96.8	92.9	91.3	95.7	92.0

UAV-YOLO 与 YOLOv7 基础模型在红外场景下的可视化对比实验结果如图 9 所示. 其中, 第 1、2 列分别是 UAV-YOLO 和 YOLOv7 基础模型的检测结果. 模型的检测类别包括车辆、自行车和人. 从图 9 的检测结果可以看出, YOLOv7 模型对红外小目标的检测能力不够优秀, 甚至会出现误检的情况, 而 UAV-YOLO 相较于 YOLOv7 模型有很大的优势, 可以准确检测红外场景下各类小目标.

为进一步验证 UAV-YOLO 模型的检测性能, 本研究选取了 YOLO 系列中的多个代表性模型以及当前最新的红外目标检测算法, 在 HIT-UAV 无人机红外目标数据集上进行了系统的对比实验. 该数据集包含在 60–

130 m 高度范围内采集的图像数据, 具有背景复杂度高、目标尺度变化显著等特点. 特别值得注意的是, 数据集中超过 85% 的目标尺寸小于 32×32 像素, 这对检测算法提出了极大的挑战. 如表 3 所示的对比实验结果充分表明, UAV-YOLO 在模型性能与计算效率方面均展现出显著优势.



(a) UAV-YOLO (b) YOLOv7

图 9 UAV-YOLO 和 YOLOv7 结果对比图

具体而言, UAV-YOLO 在仅包含 12.0M 参数量的情况下, 实现了 95.7% 的 *mAP*, 这一性能指标显著优于所有的其他对比模型. 与 YOLOv5m 和 YOLOv5l 相比, UAV-YOLO 在参数量分别减少 52.2% 和 77.4% 的同时, 检测精度提升了 1.7% 和 1.8%. 与 YOLOv8s 相比, 在参数量相近的情况下, UAV-YOLO 的检测精度提升了 2.2%. 此外, 与 YOLOv7s 相比, UAV-YOLO 在参数量减少 67.1% 的同时, *mAP* 提升了 2.1%. 与当前最新的红外小目标检测算法相比, UAV-YOLO 在检测精度和模型效率方面也展现出明显的优势. 实验结果表明, UAV-YOLO 模型在复杂环境下的红外小目标检测任务中实现了检测精度与模型效率的最佳平衡. 该模型不仅显著提

升了检测精度, 同时保持了轻量级的网络结构, 这对于无人机平台的实时目标检测应用具有重要的实践意义.

表 3 在 HIT-UAV 数据集上的对比实验

方法	*mAP* (%)	参数量 (M)	AP_{Person} (%)	$AP_{Vehicle}$ (%)	$AP_{Bicycle}$ (%)	*F*1 (%)
YOLOv5s	93.7	7.0	92.8	97.1	91.0	91.1
YOLOv5m	94.0	25.1	92.7	97.9	91.4	90.3
YOLOv5l	93.9	53.1	92.5	98.1	91.1	91.2
YOLOv7s	93.6	36.5	92.1	97.6	91.2	84.5
YOLOv7	90.5	37.2	87.6	96.3	87.7	86.2
YOLOv7-Tiny	92.8	6.1	91.3	97.2	89.9	89.2
YOLOv8s	94.2	11.2	93.0	98.2	91.4	91.1
YOLOv8n	94.6	3.1	92.4	98.0	93.4	90.9
YOLO-ViT	94.5	17.3	91.3	98.1	90.6	90.6
AIMED-Net	89.6	34.2	88.7	92.5	87.6	87.4
ITD-YOLOv8	93.5	1.8	91.7	98.2	90.7	90.3
G-YOLO	91.4	**0.8**	89.7	97.2	87.4	87.1
Ours	**95.7**	12.0	**94.5**	**98.3**	**94.3**	**92.0**

注: 加粗数据表示最佳结果

4.5 消融研究与分析

为了验证本研究提出的 UAV-YOLO 模型的有效性, 本文以 YOLOv7 为基准模型, 在 HIT-UAV 数据集上进行了一系列消融实验, 以定量评估各改进模块对模型性能的贡献度. 实验采用 640×640 的输入分辨率, batch size 设置为 16, 训练 epochs 为 500, 所有对比实验均在相同实验环境下进行. 消融实验结果如表 4 所示, 充分证实了各改进模块的有效性.

表 4 UAV-YOLO 模型中各模块的消融实验

算法	P2	CA	CARAFE	SPPFCSPC	GSConv	NWD	*mAP* (%)	参数量 (M)	FLOPs (G)
Baseline	—	—	—	—	—	—	90.5	37.2	105.1
A	√	—	—	—	—	—	93.7	12.3	99.1
B	√	√	—	—	—	—	93.9	12.3	99.3
C	√	√	√	—	—	—	94.5	12.4	100.3
D	√	√	√	√	—	—	95.1	12.4	100.3
E	√	√	√	√	√	—	95.1	12.0	96.8
F	√	√	√	√	√	√	95.7	12.0	96.8

与原始 YOLOv7 相比, UAV-YOLO 通过架构优化实现了检测性能的显著提升. 首先, 针对无人机红外小目标检测任务的特点, 我们更换了更适合检测小目标的检测头, 使用 P2 检测头进行小目标检测, 该改进使模型参数量减少了 66.9%, 实现了轻量化需求的同时使得 *mAP* 提升了 3.2%, 显著提高了对小尺度目标的特征提取能力.

其次, 在 3 个检测头前引入的 CA 注意力机制模块, 能够有效捕获小目标的低级纹理特征和空间位置信息, 在不增加参数量的前提下, 使 *mAP* 进一步提升 0.2%. 同时, 为进一步优化特征融合过程, 本研究采用内容感知的 CARAFE 上采样算子替代传统上采样方法, 此改进使 *mAP* 增加了 0.6%. 在 neck 部分, 采用改进的 SPPFCSPC 结构, 通过连续池化操作显著提升了特征提取效率, 使 *mAP* 提升 0.6%. 除此之外, 在 backbone 与 neck 部分, 针对轻量化检测模型的需求, 对网络结构中 ELAN 模块的 1×1 卷积和颈部网络中的 1×1 卷积替换成了轻量化卷积 GSConv, 参数量降低了 3.2%, 与此同时 FLOPs 下降 3.5G.

最后, 我们将损失函数替换为更适合小目标的 NWD, 利用 NWD 降低对小目标位置偏差的敏感性, 显著提高红外小目标检测精度. 该策略显著提高了红外小目标的检测效率与定位准确性, 使检测精度进一步提升 0.6%. 综合以上改进, UAV-YOLO 在保持模型轻量化的同时, 实现了显著的性能提升, 其 *mAP* 达到 95.7%, 较基准模型提高了 5.2%. 实验结果表明, UAV-YOLO 在无人机红外小目标检测任务中展现出优异的性能, 且能更好地展现出模型的检测性能与轻量化之间的平衡关系, 为相关研究提供了更有效的解决方案.

4.6 *IoU* 对比实验

为了验证 NWD 损失函数在红外小目标检测中的有效性, 将其与 CIoU、WIoU、DIoU 和 SIoU 进行比

较. 如表 5 所示, NWD 的表现优于其他损失函数, 将整体精度提高到 95.7%, 在同等条件下, 与 CIoU、DIoU、WIoU 和 SIoU 相比分别提高了 0.6%、0.9%、0.5% 和 1.0%. 实验证明, 在 HIT-UAV 数据集中, NWD 的性能优于其他 *IoU* 损失函数. 这表明 NWD 在无人机红外目标检测任务中能够更有效地评估地面真实框与预测框之间的关系, 实现更佳的检测效果.

表 5 *IoU* 对比实验 (%)

算法	*P*	*R*	*mAP*
UAV-YOLO+CIoU	91.2	91.1	95.1
UAV-YOLO+DIoU	92.0	90.1	94.8
UAV-YOLO+WIoU	91.5	90.9	95.2
UAV-YOLO+SIoU	91.6	90.4	94.7
UAV-YOLO+NWD	92.9	91.3	95.7

4.7 实际应用场景分析与讨论

鉴于红外成像不受光照条件限制的特性, UAV-YOLO 在夜视监控、低能见度环境探测等场景中具有天然优势. 在 HIT-UAV 数据集的夜间子集中, 针对以行人、车辆为主的小目标, UAV-YOLO 的精确度达到了 95.7%, 较 YOLOv7 提升 5.2%. 其中, P2 检测头的高分辨率特征图结合 CA 注意力机制, 大幅降低了行人、车辆等夜间小目标的漏检率和误检率 (如表 6 所示); 除此之外, CARAFE 模块通过内容感知上采样保留细节, 最终通过多模块优化处理使得漏检率和误检率从 14.7% 和 11.7% 降低到 8.7% 和 7.1%, 可满足夜间安防、边境监控等需求.

表 6 误检率与漏检率对比 (%)

算法	漏检率	误检率
YOLOv7	14.7	11.9
UAV-YOLO	8.7	7.1

除此之外, 面对无人机视角下目标易被树木、建筑物遮挡的问题, UAV-YOLO 也进行了优化和改进. 改进的 SPPFCSPC 模块通过连续池化扩展感知范围, 能捕捉遮挡物周围的上下文信息辅助判识; 而 NWD 损失函数通过将边界框建模为二维高斯分布, 改变误差衡量方式, 降低了小目标检测中对微小目标及遮挡条件下的小目标的定位偏差的敏感度, 最终实现了定位精度的显著提升.

5 总结

在本文中, 针对复杂场景下的多尺度红外小目标检测问题, 提出一种无人机红外小目标检测模型 UAV-YOLO, 一个专门针对复杂无人机 (UAV) 航空图像中多尺度红外小目标检测的计算效率高的框架. 为了解决无人机成像中由于目标尺寸小而导致的特征提取和计算效率不足的问题, 所提网络引入了 6 个主要改进措施.

1) 通过 GSConv 对骨干网络进行结构优化, 取代 ELAN 和颈部模块中的标准 1×1 卷积, 在保留跨层特征融合能力的同时, 减少了参数量.

2) SPPFCSPC 模块采用更快速的池化替代方式以及对整体结构在计算层面的优化, 显著降低计算量, 加速实现红外小目标的特征提取、处理任务.

3) 在不同检测头之前引入 CA 注意力机制, 利用融合的深浅特征增强目标显著性, 抑制背景干扰, 从而在复杂环境中降低漏检率和误检率.

4) 基于上下文语义的内容感知特征重组 (CARAFE) 算子动态适应上采样核, 在不影响实时推理速度的情况下保留亚像素级结构细节.

5) 相较于添加一个检测小目标的检测头, 本文进行检测头重构的优化操作, 删除了检测大目标的 P5 检测头, 引入了检测小目标的 P2 检测头, 不仅提高检测效率, 还减少了模型的计算复杂度.

6) 采用归一化高斯 Wasserstein 距离 (NWD) 损失代替传统的 *IoU* 指标, 并将边界框建模为高斯分布, 解决了定位灵敏度问题, 并提高了检测网络的收敛稳定性.

在 HIT-UAV 数据集上进行评估, 与 YOLOv7 模型和其他现有检测模型相比, UAV-YOLO 在 *mAP* 上最高精度达到 95.7%, 在保持实时性的前提下, 比 YOLOv7 的 *mAP* 高出 5.2%, 参数量下降了 67.7%, 同时在现有检测模型中实现了最高检测精度. UAV-YOLO 模型是一种高精度红外小目标检测模型, 对复杂场景下的多尺度红外小目标具有良好的检测效果. 它可以检测多场景和多类别的小物体. 在复杂场景下的无人机多尺度红外目标检测任务中具有良好的应用前景.

参考文献

1 Mao X, Diao WH. Criterion to evaluate the quality of infrared small target images. Journal of Infrared, Millimeter, and Terahertz Waves, 2009, 30(1): 56–64. [doi: 10.1007/s10762-008-9410-5]

2 Ren SQ, He KM, Girshick R, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks.

IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149. [doi: 10.1109/TPAMI.2016.2577031]

3 Redmon J, Divvala S, Girshick R, *et al*. You only look once: Unified, real-time object detection. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 779–788.

4 Liu W, Anguelov D, Erhan D, *et al*. SSD: Single shot multibox detector. Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 21–37.

5 Ju MR, Luo JN, Liu GQ, *et al*. ISTDet: An efficient end-to-end neural network for infrared small target detection. Infrared Physics & Technology, 2021, 114: 103659.

6 Huang L, Dai SS, Huang T, *et al*. Infrared small target segmentation with multiscale feature representation. Infrared Physics & Technology, 2021, 116: 103755.

7 Zhang LC, Gonzalez-Garcia A, van de Weijer J, *et al*. Synthetic data generation for end-to-end thermal infrared tracking. IEEE Transactions on Image Processing, 2019, 28(4): 1837–1850. [doi: 10.1109/TIP.2018.2879249]

8 吴双忱, 左峥嵘. 基于深度卷积神经网络的红外小目标检测. 红外与毫米波学报, 2019, 38(3): 371–380.

9 Ding LH, Xu X, Cao Y, *et al*. Detection and tracking of infrared small target by jointly using SSD and pipeline filter. Digital Signal Processing, 2021, 110: 102949. [doi: 10.1016/j.dsp.2020.102949]

10 Zhang RZ, Zhang JL, Qi XP, *et al*. Infrared target detection and recognition in complex scene. Opto-electronic Engineering, 2020, 47(10): 200314.

11 Chen JY, Ding YH, Li Y, *et al*. Assessment of bypass patency using transcranial doppler sonography: Correlations with computerized tomography angiography findings in patients with moyamoya disease. Neurosurgical Review, 2023, 46(1): 64. [doi: 10.1007/s10143-023-01970-0]

12 Gong YQ, Yu XH, Ding Y, *et al*. Effective fusion factor in FPN for tiny object detection. Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2021. 1159–1167. [doi: 10.1109/WACV48630.2021.00120]

13 Wang CY, Bochkovskiy A, Liao HYM. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 7464–7475. [doi: 10.1109/CVPR52729.2023.00721]

14 Song QS, Li SB, Bai Q, *et al*. Object detection method for grasping robot based on improved YOLOv5. Micromachines, 2021, 12(11): 1273. [doi: 10.3390/mi12111273]

15 Li K, Wang YN, Hu ZM. Improved YOLOv7 for small object detection algorithm based on attention and dynamic convolution. Applied Sciences, 2023, 13(16): 9316. [doi: 10.3390/app13169316]

16 Gao CQ, Meng DY, Yang Y, *et al*. Infrared patch-image model for small target detection in a single image. IEEE Transactions on Image Processing, 2013, 22(12): 4996–5009. [doi: 10.1109/TIP.2013.2281420]

17 Fan YC, Qiu QL, Hou SH, *et al*. Application of improved YOLOv5 in aerial photographing infrared vehicle detection. Electronics, 2022, 11(15): 2344. [doi: 10.3390/electronics11152344]

18 Xu S, Chen Z, Xu WH, *et al*. HCF-Net: Hierarchical context fusion network for infrared small object detection. arXiv:2403.10778, 2024.

19 Rawat SS, Verma SK, Kumar Y. Review on recent development in infrared small target detection algorithms. Procedia Computer Science, 2020, 167: 2496–2505. [doi: 10.1016/j.procs.2020.03.302]

20 Zhou TF, Li JW, Wang SZ, *et al*. MATNet: Motion-attentive transition network for zero-shot video object segmentation. IEEE Transactions on Image Processing, 2020, 29: 8326–8338. [doi: 10.1109/TIP.2020.3013162]

21 Vaswani A, Shazeer N, Parmar N, *et al*. Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.

22 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 7132–7141. [doi: 10.1109/CVPR.2018.00745]

23 Woo S, Park J, Lee JY, *et al*. CBAM: Convolutional block attention module. Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich: Springer, 2018. 3–19.

24 Ma Q. YOLOv5-CBAM: A small object detection model based on YOLOv5 and CBAM. Proceedings of the 6th International Conference on Robotics, Intelligent Control and Artificial Intelligence (RICAI). Nanjing, 2024. 618–623. [doi: 10.1109/RICAI64321.2024.10911839]

25 Li H, Li J, Wei H, *et al*. Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicles. arXiv:2206.02424, 2024.

26 Zhang LH, Xiong N, Pan XH, *et al*. Improved object

detection method utilizing YOLOv7-tiny for unmanned aerial vehicle photographic imagery. Algorithms, 2023, 16(11): 520. [doi: 10.3390/a16110520]

27 Nouaze JC, Sikati J. YOLO-AppleScab: A deep learning approach for efficient and accurate apple scab detection in varied lighting conditions using CARAFE-enhanced YOLOv7. Biology and Life Sciences Forum, 2024, 30(1): 6.

28 Wang JQ, Chen K, Xu R, *et al*. CARAFE: Content-aware reassembly of features. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019. 3007–3016.

29 Li S, Tao T, Zhang Y, *et al*. YOLO v7-CS: A YOLO v7-based model for lightweight bayberry target detection count. Agronomy, 2023, 13(12): 2952. [doi: 10.3390/agronomy13122952]

30 Hou QB, Zhou DQ, Feng JS. Coordinate attention for efficient mobile network design. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 13708–13717.

31 Cao S, Xiao J, Meng X, *et al*. MFDF-YOLOv7: YOLOv7-based multiscale feature dynamic fusion small object detection algorithm. Proceedings of the 5th International Conference on Frontiers Technology of Information and Computer (ICFTIC). Qiangdao, 2023. 463–471. [doi: 10.1109/ICFTIC59930.2023.10454141]

32 Chen X, Yuan MJH, Yang Q, *et al*. Underwater-YCC: Underwater target detection optimization algorithm based on YOLOv7. Journal of Marine Science and Engineering, 2023, 11(5): 995. [doi: 10.3390/jmse11050995]

33 Shi QX, Zhong F, Li B, *et al*. Fast vehicle detection algorithm based on lightweight YOLO7-tiny. Proceedings of the 5th International Conference on Computer Vision and Data Mining (ICCVDM 2024). Changchun: SPIE, 2024. 1327208.

34 Zheng ZH, Wang P, Ren DW, *et al*. Enhancing geometric factors in model learning and inference for object detection and instance segmentation. IEEE Transactions on Cybernetics, 2022, 52(8): 8574–8586. [doi: 10.1109/TCYB.2021.3095305]

35 Wang J, Xu C, Yang W, *et al*. A normalized Gaussian Wasserstein distance for tiny object detection. arXiv:2110.13389, 2022.

36 Suo JS, Wang TY, Zhang XZ, *et al*. HIT-UAV: A high-altitude infrared thermal dataset for unmanned aerial vehicle-based object detection. Scientific Data, 2023, 10(1): 227. [doi: 10.1038/s41597-023-02066-6]

(校对责编: 张重毅)