计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications [doi: 10.15888/j.cnki.csa.009471]

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

高阶深度可分离无人机图像小目标检测算法①

郭　伟, 王珠颖, 金海波

(辽宁工程技术大学 软件学院, 葫芦岛 125105)
通信作者: 王珠颖, E-mail: 18342989188@163.com

摘　要: 当前无人机图像中存在小目标数量众多、背景复杂的特点, 目标检测中易造成漏检误检率较高的问题, 针对这些问题, 提出一种高阶深度可分离无人机图像小目标检测算法. 首先, 结合 CSPNet 结构与 ConvMixer 网络, 深度可分离卷积核, 获取梯度结合信息, 并引入递归门控卷积 C3 模块, 提升模型的高阶空间交互能力, 增强网络对小目标的敏感度; 其次, 检测头采用两个头部进行解耦, 分别输出特征图分类和位置信息, 加快模型收敛速度; 最后, 使用边框损失函数 EIoU, 提高检测框精准度. 在 VisDrone2019 数据集上的实验结果表明, 该模型检测精度达到了 35.1%, 模型漏检率和误检率有明显下降, 能够有效地应用于无人机图像小目标检测任务. 在 DOTA 1.0 数据集和 HRSID 数据集上进行模型泛化能力测试, 实验结果表明, 该模型具有良好的鲁棒性.

关键词: 小目标检测; 递归门控卷积; 解耦头; 无人机图像; YOLOv5

引用格式: 郭伟,王珠颖,金海波.高阶深度可分离无人机图像小目标检测算法.计算机系统应用. http://www.c-s-a.org.cn/1003-3254/9471.html

Small Target Detection Algorithm for High-order Depth Separable UAV Images

GUO Wei, WANG Zhu-Ying, JIN Hai-Bo

(School of Software, Liaoning Technical University, Huludao 125105, China)

Abstract: At present, there are many small targets in UAV images and the background is complex, which makes it easy to cause a high error detection rate in target detection. To solve these problems, this study proposes a small target detection algorithm for high-order depth separable UAV images. Firstly, by combining the CSPNet structure and ConvMixer network, the study utilizes the deeply separable convolution kernel to obtain the gradient binding information and introduces a recursively gated convolution C3 module to improve the higher-order spatial interaction ability of the model and enhance the sensitivity of the network to small targets. Secondly, the detection head adopts two heads to decouple and respectively outputs the feature map classification and position information, accelerating the model convergence speed. Finally, the border loss function EIoU is leveraged to improve the accuracy of the detection frame. The experimental results on the VisDrone2019 data set show that the detection accuracy of the model reaches 35.1%, and the missing and false detection rates of the model are significantly reduced, which can be effectively applied to the small target detection task of UAV images. The model generalization ability is tested on the DOTA 1.0 dataset and the HRSID dataset, and the experimental results show that the model has good robustness.

Key words: small target detection; recursively gated convolution; decouple head; unmanned aerial vehicle (UAV) image; YOLOv5

① 基金项目: 国家自然科学基金 (62173171)
收稿时间: 2023-11-01; 修改时间: 2023-12-04; 采用时间: 2023-12-11; csa 在线出版时间: 2024-01-30

随着无人机 (unmanned aerial vehicle, UAV) 相关技术的不断深入研究, 其应用场景不断扩展, 由于无人机具有低成本、轻便快速的特点[1], 被广泛应用于工农业、抢险救灾、遥感图像和视频录制等实际行业领域, 它是卫星遥感的有力补充[2]. 无人机的理想摄影平台是低空航空拍摄, 利用高清传感器可以捕捉地面目标, 所拍摄的图像分辨率较高. 在云层不影响图像拍摄的低空飞行状态下, 无人机使用高清摄像头拍摄所获得的图像可以达到厘米级别的分辨率[3]. 但相对来说, 在大多数的无人机航拍图像中, 由于较高的拍摄高度, 无人机距离地面较远, 图像中的各个类别的待检测目标对象呈现出小目标的特点, 如人、摩托车、自行车等, 小目标往往特征单一, 像素比较低, 检测过程中容易出现漏检或者误检等不足之处. 背景复杂, 存在着目标遮挡, 光线不足、相似形状物体干扰等因素的干扰[4]. 因此提高无人机航拍图像中小目标的检测精度已成为目标检测中一个具有挑战性的研究方向.

传统的目标检测算法很难提取到足够好的特征来进行检测, 而深度学习的快速发展极大地推动了目标检测技术的前进, 在一定程度上改变了人类的生活方式, 减轻了人们的负担. 基于深度学习的目标检测算法, 可以根据其模型训练方式特点大致分为单阶段 (one-stage) 目标检测算法和二阶段 (two-stage) 目标检测算法两类. 二阶段目标检测算法的代表为 R-CNN[5]、Fast R-CNN[6]、Faster R-CNN[7]、R-FCN[8], 此类算法基于区域提取, 对候选区域进行位置精修和类别分类, 逐步优化, 虽然其检测精度略高, 但该类算法的检测流程相对复杂, 检测速度普遍较慢. 不适合应用于对实时性要求较高的场景中. 单阶段目标检测算法采用了回归思想的端到端的目标检测, 相比于二阶段目标检测算法, 单阶段目标检测算法起步较晚, 能够更好地吸收前者的优势, 同时克服不足之处. 其主要代表是 SSD (single shot multibox detector) 系列[9]、YOLO (you only look once) 系列[10–13]和 RetinaNet 系列[14], 此类检测算法精度略低, 但其检测速度很快, 广泛应用于工业界.

近些年来, 单阶段目标检测算法 YOLO 系列不断改进, 模型检测精度有了很大提升, 被广泛应用于无人机图像检测. 程江川等人[15]采用轻量化高效骨干网络对模型进行改进, 提高模型检测速度. 李利霞等人[16]提出浅层特征增强模块, 并设计多级特征融合模块, 动态调节各输出检测层的权重, 增强小目标特征信息. 韩俊等人[17]提出了 LSA_YOLO 模型, 构造多尺度特征提取模块, 设计自适应权重动态融合结构, 增强算法对密集小目标的检测能力; 李扬等人[18]以 YOLOX 为基线算法, 通过多尺度特征融合缩小网络感受野, 提高网络细节提取能力, 提升对小目标图像的敏感程度. 奉志强等人[19]提出了一种改进无人机实时密集小目标检测的算法, 提出一种空间-通道注意力模块, 在主干网络中引入 Transformer 模块, 提高复杂背景下密集小目标的特征提取能力.

上述算法存在检测能力较弱、模型参数量较大的缺点, 无法快速且准确的检测无人机小目标图像, 为了在实际场景中满足无人机图像小目标检测算法的要求, 精准且迅速地检测出对象目标信息, 选择 YOLOv5s 算法为基线算法进行改进, 提出 CHD-YOLOv5 模型. 首先在骨干网络部分, 融合 CSPNet 网络与 ConvMixer[20], 将递归门控卷积[21]与 C3 模块结合构建 C3HB 模块, 并且替换网络预测头部为精简的解耦头部, 损失函数选用 EIoU[22], 最后在 VisDrone2019 数据集上测试改进后的算法, 得到的实验结果数据表明, 改进后的算法对检测无人机小目标图像的精度值有效提高了, 明显优于原算法.

1 改进的 CHD-YOLOv5 模型

1.1 YOLOv5 网络结构

YOLOv5 检测算法集合了很多算法的优势, 是一个高效精准的目标检测框架, 该算法有良好的稳定性与适应性, 目前应用最为广泛. YOLOv5 算法共有 YOLOv5n、YOLOv5s、YOLOv5m、YOLOv5l 和 YOLOv5x 这 5 个版本, YOLOv5s 网络的模型文件只有 24M, 除 YOLOv5n 外, 其他 3 个版本都在 YOLOv5s 的基础上对网络不断进行加深与加宽. YOLOv5 算法在模型训练阶段使用数据增强方法 Mosaic, 通过随机缩放、随机裁剪、随机排布的方式对不同图像进行拼接, 能够有效提高算法对小目标的检测能力. 骨干网络部分, YOLOv5 引入了能够直接处理输入图片的 Focus 结构. 其结构最重要的作用是切片操作, 可以将 2×2×3 的图像切片为 1×1×12 的特征图. 网络输出部分: 延续 YOLO 系列的一贯做法, 与 YOLOv3、YOLOv4 类似, 采用的预测头部是耦合的 Head, 共有 3 个不同的输出头部, 进行多尺度预测. YOLOv5s 较好地平衡了速度与性能, 所以本文的基线模型选择 YOLOv5s 版本.

1.2 CHD-YOLOv5 模型

无人机图像存在尺寸较小、容易受环境影响、数量大且目标密集等问题, 为了快速且精准识别无人机小目标图像, 算法改进以 YOLOv5s 算法为基础. 基于 CSPNet 构建 CSP-ConvMixer 模块, 实现模型梯度信息的结合, 增强网络的学习能力, 提升算法性能; 通过引入递归门控卷积与 C3 模块结合, 明显提高了模型对小目标的高阶空间信息捕获能力, 提升模型的高阶交互能力; 把边框损失函数由 CIoU 替换为 EIoU, 对预测和真实框之间宽和高的预测结果进行惩罚, 提高收敛速度和边界框定位准确度, 增强模型检测效果; 解耦预测部分的分类任务与回归任务, 加快模型收敛速度. 本文将详细介绍本文所提出 CHD-YOLOv5 模型, 叙述 CHD-YOLOv5 模型中改进的网络结构以及针对小目标检测而新设计的模块, 最后将 CHD-YOLOv5 模型与其他主流模型进行对比, 有效表明本文所提出模型的高效性. 本文模型的网络结构图如图 1 所示.

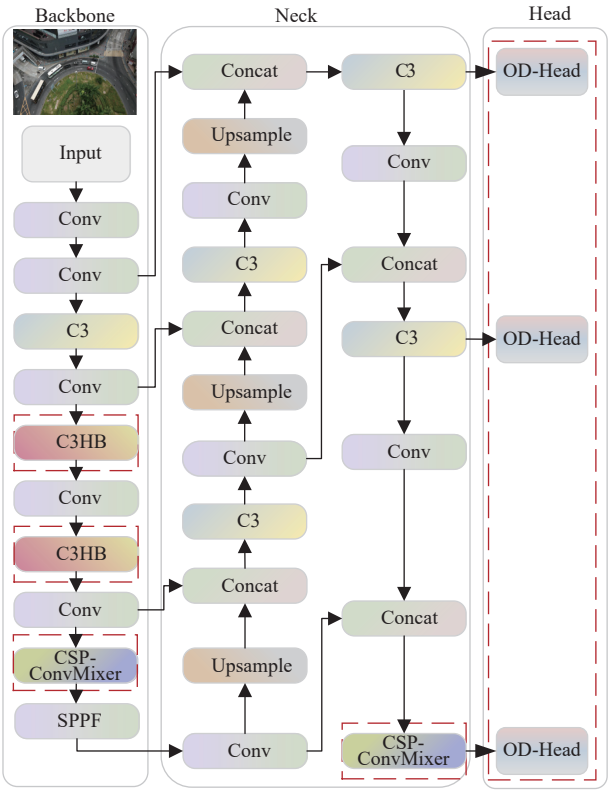


图 1 改进后的 CHD-YOLOv5 网络结构

1.2.1 CSP-ConvMixer 模块

Wang 等人[23]提出新型骨干网络 CSPNet (cross stage partial network) 结构. 该结构能够在降低网络计算量的同时, 获取更丰富的梯度融合信息, 保持甚至提高 CNN 的能力, 减少内存消耗. CSPNet 结构通过将浅层的特征图进行通道拆分, 一部分经由特征提取模块向后传播, 另一部分则采用跨阶段特征融合策略直接与特征提取模块的输出进行合并, 实现了梯度信息的结合, 增强网络的学习能力. CSPNet 结构可以与多种网络结构结合, 轻松应用于多数经典 CNN 模型, 在相同的实验环境下, 可以提高模型的学习能力, 提升算法性能.

ConvMixer 是一个只需要使用标准的卷积的 patch embedding 网络, 通过空间融合 (depthwise convolution) 和通道融合 (pointwise convolution) 来减少模型的参数. patch embedding 的主要功能是对原始输入图像 (h, w) 划分图像块. 首先指定每个图像块的 size 为 (*patch_size*, *patch_size*), 将每张图像划分出 (h/*patch_size*, w/*patch_size*) 个图像块, kernel size 为 q, stride 为 q, 计算公式为式 (1) 所示:

$$z_0 = BN(\sigma\{\mathrm{Conv}(X,q,q)\}) \tag{1}$$

在 ConvMixer 中, 特征提取层由深度卷积 (depthwise conv)、GELU 激活函数和逐点卷积 (pointwise conv) 这 3 部分组成. 混合分离空间和通道维度, 即先通过一个深度可分离卷积提取图片长宽的特征信息, 再通过一个逐点卷积融合通道信息, 同时在整个过程中, 保证大小和分辨率相同. 与 ResNet 等 CNN、Swin、AS-MLP 等金字塔结构的分类头一致, 每个卷积之后是一个激活函数和激活后的 BatchNorm, 计算公式如下所示:

$$z'_l = BN(\sigma\{\mathrm{ConvDepthwise}(z_l - 1)\}) + z_{l-1} \tag{2}$$

$$z_{l+1} = BN\left(\sigma\left\{\mathrm{ConvPointwise}\left(z'_l\right)\right\}\right) \tag{3}$$

Patch embedding 通过一次性进行下采样来减少内部分辨率, 有效增加感受野, 从而可以更容易的将远距离空间信息混合. 在 patch embedding 和多个 ConvMixer layer 之后, 经过全局池化获得特征向量, 再经过全连接层传递进行分类. 实验数据证明, 模型性能上, ConvMixer 可以与 ResNet 等经典模型相比较, 也明显优于 VIT 和 MLP-Mixer 等, 且越大的深度可分离卷积卷积核, 其模型的性能提升越强. CSPNet 结构应用于 ConvMixer 之后的网络结构如图 2 所示.

1.2.2 递归门控卷积 C3HB 模块

递归门控卷积模块由门控卷积和递归设计组成,

利用门控卷积和递归设计进行高阶空间交互作用, 具有执行高效、可扩展和平移等变的特点, 可以基于各种视觉 Transformer 和基于 CNN 的模型进行即插即的改进. 空间交互作用是在特征图采样过程中通过额外的或者自身的某种计算, 将空间的相互性加入到生成的特征中, 例如 Transformer 中就利用不同的变换生成 Q 和 K 进而进行空间位置相关性的计算. 高阶空间交互是通过增加通道宽度的设计而高效实现的具有有限复杂性的任意阶空间交互. 通道维度数计算公式如式 (4) 所示:

$$C_k = \frac{C}{2^{n-k}},\ k \in (0, n-1) \tag{4}$$

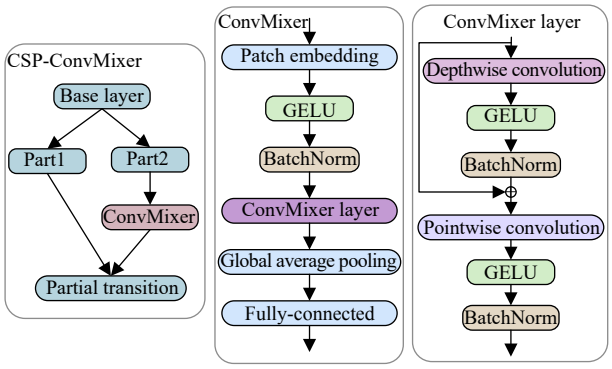


图 2 CSP-ConvMixer 模块

门控卷积首先利用两个卷积层来调整特征通道数, 深度可分离卷积的输出特征会沿着特征分为多块, 考虑空间信息交互, 前后两块的特征交互将进一步利用逐元素相乘的方式, 最终得到输出特征. 首先使用线性投影函数φ得到一组投影特征如式 (5) 所示:

$$[p_0^{H\times W\times C_0}, q_0^{H\times W\times C_0}, \cdots, q_{n-1}^{H\times W\times C_{n-1}}] = \varphi(x) \in R^{H\times W\times (C_0+\sum_{0\leqslant k\leqslant n-1} C_k)} \tag{5}$$

以递归的方式进行门控卷积如式 (6) 所示:

$$p_{k+1} = f_k(q_k) \odot n_k/\alpha\ (k = 0, 1, \cdots, n-1) \tag{6}$$

其中, f_k代表 depthwise 卷积操作, n_k代表递归操作中匹配的特征通道数. 递归设计通过不断进行元素的逐个相乘, 提高网络容以提升模型性能. 通过这种递归方式, 越靠后的特征其高阶信息保存越多, 使其高阶中的特征交互就会更多, 有助于提高视觉模型的表达能力, 有效表明高阶交互对处理信息的作用.

YOLOv5 算法特征提取部分的 CSP 结构集成残差组件, 但是由于残差模块无法充分学习特征信息, 缺乏全局信息提取能力, 无法有效检测出无人机小目标图像. 因此, 在 YOLOv5 算法中引入递归门控卷积. YOLOv5 算法主干网络部分的 C3 模块具有丰富的语义信息, 将最后两个 C3 模块与递归门控卷积相结合, 构建 C3HB (C3-HorBlock) 模块, 递归门控卷积 C3HB 模块的结构图如图 3 所示. 一般来说, 检测性能越强的模型, 其主干网络提取特征能力越强. 借鉴 Transformer 结构将 HorNet 中的递归门控卷积模块 g^{n}Conv 引入, 构建由空间混合层 (MLP) 和前馈网络 (FFN) 组成的 HorBlock. 替换 YOLOv5 中的 n 个残差网络为 n 个 HorBlock, 并引入递归门控卷积模块 g^{n}Conv 替换掉 C3 模块中的卷积模块, 有效增强了主干网络部分的特征信息提取能力. 该模块的计算量如式 (7) 所示:

$$FLOPs = HWC\left(2K^2 + 11/3\times C + 2\right) \tag{7}$$

该模块能够在无人机图像小目标检测中通过实现上下文的小目标信息交互而预测其具体位置, 提升模型的高阶交互能力, 它可以逐步通过相邻特征间的信息交互实现高阶特征目标获取, 提高网络对较小目标的敏感度. 在提升网络检测精度的同时, 既避免了过度增加参数量, 又保证了特征信息在送往特征增强网络之前不会丢失目标信息的语义, 有效提高了模型的高阶空间交互能力, 从而增强模型的检测能力.

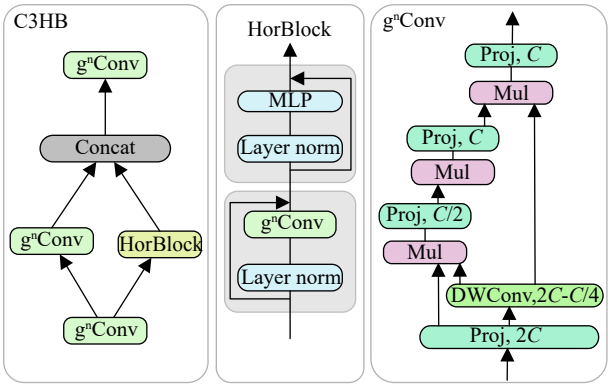


图 3 递归门控卷积 C3HB 模块

1.2.3 精简解耦检测头部 OD-Head

众所周知, 目标检测的分类任务和回归任务之间是相互矛盾的. 因此, 大多数一级和二级探测器应用解耦头部, 但是, 由于 YOLO 系列的主干和特征金字塔不断演化, 从 YOLOv3 发展到 YOLOv5, 原始模型的检测头部是通过分类和回归分支融合共享的方式来实现的, 检测头部仍然是耦合的, 其检测头部结构如图 4 所示.

YOLOX[24]的检测头分别采用两个不同的头解耦输出的特征图, 各自输出分类和位置, 独自解耦分类和回归两个分支, 在提升精度的同时, 有效加快网络收敛速度. 因此解耦头被广泛应用于目标检测算法中.

在 YOLOv5 的检测头部引入解耦头部的方法, 可以有效提高检测精度, 加速网络的收敛, 本文中提出的 YOLOv5 中引入的解耦头部也是基于 anchor 检测的方法, 但与 YOLOX 解耦头部所使用的检测方法有所不同. YOLOX 的解耦头部在提升模型精度的同时, 新增了多个额外的卷积层, 大大增加了延时和参数量, 本文提出了一个更加精简的解耦头部 OD-Head, 其检测头部结构如图 5 所示. OD-Head 的添加方式有所不同, 参数量和计算量有明显下降, 在维持精度的同时降低了延时, 缓解了解耦头部中 3×3 卷积带来的额外延时开销.

图 4 YOLOv3–YOLOv5 检测头部结构

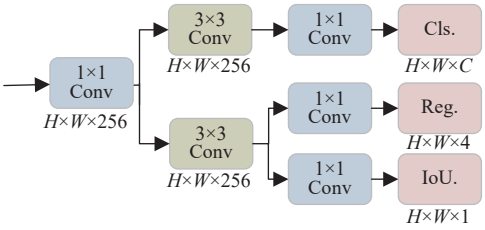


图 5 解耦检测头部 OD-Head 结构

1.2.4 EIoU 损失函数

目标检测任务的损失函数一般由分类损失函数 (classification loss) 和回归损失函数 (bounding box regression loss) 两部分构成, 是目标检测中的常见指标. *IoU_Loss* 主要是通过预测框和真实框的相交区域面积和合并区域面积的比值, *IoU* 的计算公式如式 (8) 所示:

$$IoU = \frac{A \bigcap B}{A \bigcup B} \tag{8}$$

其中, A 和 B 分别表示预测框和真实框, 即 *IoU* 值越高, 检测结果越好. 尺度不变性是 *IoU* 一个很好的特性, 也就是对尺度不敏感 (scale invariant), 在 regression 任务中, 判断 predict box 和 gt 的距离最直接的指标就是 *IoU*. 如果 *IoU* 值为 0, 即两个目标框没有任何重叠时, 梯度为 0, 则 *IoU* 无法反映两个目标的精准重合度, 不能优化. 如式 (9) 所示:

$$IoU\_Loss = 1 - IoU \tag{9}$$

IoU 的变体有 EIoU、SIoU、GIoU、WIoU 等, YOLOv5 算法的损失函数为 CIoU. CIoU 多用于训练, 该损失函数考虑了边界框回归的重叠面积、中心点距离和纵横比, 但当预测框的宽高满足一定条件时, CIoU 中此项的惩罚便失去了作用. EIoU 考虑了重叠面积、中心距离、长宽边长真实差, 解决了 CIoU 纵横比的模糊定义. EIoU 的计算公式如式 (10) 所示:

$$\begin{aligned} L_{\mathrm{EIoU}} &= L_{\mathrm{IoU}} + L_{\mathrm{dis}} + L_{\mathrm{asp}} \\ &= 1 - IoU + \frac{p^2(b,b^{\mathrm{gt}})}{c^2} + \frac{p^2(w,w^{\mathrm{gt}})}{c_w^2} + \frac{p^2(h,h^{\mathrm{gt}})}{c_h^2} \end{aligned} \tag{10}$$

其中, c 代表的是能够同时包含预测框和真实框的最小闭包区域的对角线距离, c_w 和 c_h 分别表示最小外接矩形的宽和高, $p^2(b,b^{\mathrm{gt}})$ 表示预测框与真实框之间中心点的欧氏距离, (w,h) 和 $(w^{\mathrm{gt}},h^{\mathrm{gt}})$ 分别为预测框和真实框的宽和高, EIoU 将真实框与预测框的宽高进行惩罚对比, 精准定位检测框, 提升模型的检测精度.

2 实验

本文实验环境为 Windows 11 操作系统, 搭载 10 vCPU Intel(R) Xeon(R) Gold 5218R CPU@2.10 GHz 处理器, 使用 RTX3090 (24 GB) GPU 进行训练推理, 采用 PyTorch 1.9.0 版本的深度学习框架, 加速库是 CUDA 11.1, 所使用的编程语言为 Python 3.8.

2.1 数据集

本文实验数据选取 VisDrone2019 无人机航拍数据集, 该数据集是由天津大学机器学习与数据挖掘实验室的 AI SKYEYE 团队所收集的, 图像采集于国内 14 个城市, 由不同类型的无人机在不同的天气情况、不同的场景和不同的光照条件下拍摄, 单张图像中往往包含多种目标信息、大量的小目标, 且目标存在不同程度的遮挡. VisDrone2019 数据集共有 8 629 张静态图片, 划分为 3 个部分, 训练数据集 6 471 张, 验证数据集 548 张, 和测试数据集 1 580 张. VisDrone2019 数据集图像种类包含 10 类, 分别为: 人、行人、自行车、汽车、卡车、面包车、三轮车、遮阳三轮车、公交车以及摩托车, 数据集共有 260 万个标注. VisDrone2019 数据集中的数据信息如图 6 所示. 由图 6(b) 可知, 该数据集中包含大量分布密集的小目标, 符合本文的研究问题.

2.2 评价指标与参数设置

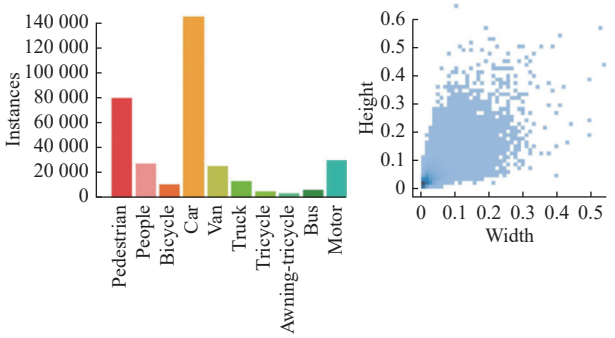
本文实验基于 YOLOv5s 模型进行改进, 初始学习率为 0.01, 训练轮数为 100 epochs, 优化器为 SGD, 数据增强采用 Mosaic, 输入图像的分辨率为 640×640, 实验选取精确率 (precision, P)、召回率 (recall, R)、整体平均精度均值 (mAP)、每秒传输帧数 (FPS) 以及网络的参数量 (Params) 为评价指标, 通过实验数据对比与分析对算法网络模型进行评估, 其中公式如下所示:

$$P=\frac{TP}{TP+FP}\times 100\% \tag{11}$$

$$R=\frac{TP}{TP+FN}\times 100\% \tag{12}$$

$$mAP=\frac{\sum\int_0^1 P(R)\mathrm{d}R}{n}\times 100\% \tag{13}$$

其中, 式 (11) 为精确度的计算公式, 式 (12) 为召回率的计算公式, 式 (13) 为整体平均精度均值的计算公式, TP、FP 和 FN 分别表示正确检测框、错误检框和漏检框的数量, n 表示所有需要检测的类别总数.



(a) 标签分布图 (b) 长宽分布图

图 6 VisDrone2019 数据信息图

2.3 模块对比实验分析

2.3.1 CSP-ConvMixer 模块对比分析

本文结合 CSPNet 和 ConvMixer 构建的 CSP-ConvMixer 模块, 实现了更加丰富的梯度组合, 有利于提高模型的学习能力, 仅使用标准卷积来实现混合步骤, 有效扩大感受野, 从而能够更容易的混合远距离的空间信息, 但较金字塔模型相比较, 提高了模型复杂度, 推理速度有所减缓, 实验数据如表 1 所列. 该模块较基线模型 mAP_{50} 值提升了 5.4%, mAP_{75} 值提升了 4.1%, 模型精度检测效果提升明显.

2.3.2 递归门控卷积 C3HB 模块对比分析

本文构建的递归门控卷积 C3HB 模块, 与具有相似的整体架构 Swin Transformer 和 ConvNeXtV2 相比, 实验数据如表 2 所列. 该模块具有空间高阶特征提取能力, 有良好的可扩展性, 能够更好地捕获小目标的细节信息, 增加了模型参数量, 减缓了网络检测速度, 但 mAP_{50} 和 mAP_{75} 均有明显提升, 优于 Swin Transformer 和在 ConvNeXt 基础上升级的 ConvNeXtV2, 证明了该模块对提高模型性能的有效性.

表 1 CSP-ConvMixer 模块的对比实验

模块	Params ($\times10^4$)	FLOPs ($\times10^6$)	mAP_{50} (%)	mAP_{75} (%)
CSP	7.03	15.8	27.5	13.5
CSP-ConvMixer	32.38	111.9	32.9	17.6

表 2 递归门控卷积 C3HB 模块的对比实验

模块	Params ($\times10^4$)	FLOPs ($\times10^6$)	mAP_{50} (%)	mAP_{75} (%)
C3	7.03	15.8	27.5	13.5
Swin Transformer	9.73	105.3	26.2	13.0
ConvNeXtV2	7.07	15.9	28.7	14.9
C3HB	44.13	98.2	29.7	15.8

2.3.3 解耦头部对比分析

本文提出的基于 anchor 检测的方法替换预测头部为 OD-Head 解耦头部, 平均精度均值提升了 0.8 个百分点, 有效提升网络收敛速度, 并提升模型检测精度; 为了表明本文所设计的 OD-Head 解耦头部更加具有高效性与实时性, 将 OD-Head 与 YOLOv5 的 coupled head 和 YOLOX[24]的 decoupled head 进行对比实验, 实验数据如表 3 所列. 与耦合头部相比较, 解耦头部的参数量和浮点运算数有所增加, 同时模型的精度也提高了. 经过优化的 OD-Head 较 YOLOX 的 decoupled head, 参数量和浮点运算数均有明显的大幅度下降, 平均精度均值却提升了 0.3 个百分点. 解耦头部的对比试验有效证明了本文提出的 OD-Head 解耦头部可以在提升模型检测精度的同时减少参数量, 加快网络的收敛速度.

表 3 检测头部的对比实验

模块	Params ($\times10^4$)	FLOPs ($\times10^6$)	mAP_{50} (%)	mAP_{75} (%)
Coupled head	7.03	15.8	27.5	13.5
Decoupled head	14.35	56.5	28.0	14.9
OD-Head	8.79	21.6	28.3	15.1

2.3.4 EIoU 损失函数对比分析

本文在 4 种不同损失函数的情况下进行对比实验, 以更进一步验证损失函数 EIoU 对无人机图像小目标检测的有效性, VisDrone2019 数据集在不同损失函数影响下的检测精度, 实验数据如表 4 所列. 引入 EIoU

损失函数后, 模型的 mAP_{50} 值提升了 1.0%. EIoU 引入目标框和预测框的长宽信息, 避免了出现中长宽比等比例变化惩罚失效的问题, 加快网络模型的收敛速度, 提高回归精度, 快速定位结果, 更适合于复杂背景下的无人机小目标检测任务.

表 4 损失函数的对比实验 (%)

损失函数	mAP_{50}	mAP_{75}
GIoU	27.4	14.3
CIoU	27.5	13.5
SIoU	27.9	14.8
WIoU	27.0	14.1
EIoU	28.5	15.1

2.4 消融实验分析

为了直观地体现各方法对网络性能的影响, 验证本文提出 4 个模块对算法改进的有效性, 即在 YOLOv5s 算法中增加 CSP-ConvMixer 模块, 主干网络引入递归门控卷积 C3HB 模块, 将检测头部 (coupled head) 替换为解耦头部 OD-Head 以及使用 EIoU 损失函数. 消融实验的进行选取完全相同的实验环境, 评估各个模块对模型性能的影响. 实验结果如表 5 所列. 引入 CSP-ConvMixer 模块, 平均精度均值提升了 3.2 个百分点, 验证了该模块的有效性; 将原 YOLOv5 算法中主干网络部分 C3 结构的残差网络替换为递归门控卷积 C3HB 模块, 平均精度均值提升了 2.2 个百分点, 有效地捕获了高阶空间信息交互能力, 逐步细化了局部信息, 增强了模型的特征信息提取能力和建模能力; 基于 anchor 检测的方法, 替换预测头部为 OD-Head 解耦头部, 平均精度均值提升了 0.8 个百分点, 有效提升网络收敛速度, 并提升模型检测精度; 引入 EIoU 损失函数, 模型的平均精度均值上升了 1.0 个百分点, EIoU 损失函数在 CIoU 损失函数的基础上, 分别计算预测框和真实框宽高的差异值, 解决了纵横比的模糊定义, 加快了预测框的收敛速度, 有效提升了小目标检测精度.

表 5 消融实验数据表

YOLOv5s	CSP-ConvMixer	C3HB	OD-Head	EIoU	Params ($\times10^6$)	FLOPs ($\times10^6$)	mAP_{50} (%)	mAP_{75} (%)
√	—	—	—	—	7.03	15.8	27.5	13.5
√	√	—	—	—	32.36	111.2	32.9	17.6
√	—	√	—	—	44.13	98.2	29.7	15.8
√	—	—	√	—	8.79	21.6	28.3	15.1
√	—	—	—	√	7.03	15.8	28.5	15.1
√	√	√	—	—	28.83	97.0	33.4	17.8
√	√	√	√	—	31.49	122.7	34.2	18.6
√	√	√	√	√	31.49	122.7	35.1	19.3

为了更加直观地展示出 CHD-YOLOv5 算法检测效果的优越性, 选取部分图像进行对比. 考虑到无人机图像易受光线变化影响、小目标密集等特点, 选取具有代表性的 4 种不同场景图片对传统的 YOLOv5 模型和改进后的 CHD-YOLOv5 模型进行检测效果对比. 图 7 为日常普通简单场景拍摄图片, 图 8 为大量目标有遮挡的场景, 图 9 为小目标较为密集的场景, 图 10 为光线昏暗的夜间场景. (a) 组图片为 VisDrone2019 数据集原始图片, (b) 组图片为 YOLOv5 算法检测效果图, (c) 组图片为改进后的 CHD-YOLOv5 算法检测效果图. 算法的效果图明显展示出在同一场景下, YOLOv5 算法容易忽略边缘区域的目标信息, 在有遮挡或小目标密集的位置, 以及光线昏暗的情况下, 出现漏检和误检现象, 而改进后的 CHD-YOLOv5 模型对小目标的漏检误检情况有所减弱, 且在光线不良的夜间场景下小目标检测效果也有所增强, 提升了模型检测精度. 因此, 改进后的算法检测效果更好, 检测到的目标信息更加细节, 具有良好的鲁棒性, 在无人机图像小目标的检测中更有优势.

(a) 原图 (b) YOLOv5s (c) CHD-YOLOv5

图 7 简单场景检测效果对比

(a) 原图 (b) YOLOv5s (c) CHD-YOLOv5

图 8 有遮挡场景检测效果对比

(a) 原图 (b) YOLOv5s (c) CHD-YOLOv5

图 9 密集场景检测效果对比

(a) 原图 (b) YOLOv5s (c) CHD-YOLOv5

图 10 夜间场景检测效果对比

2.5 模型对比实验分析

为了验证本文改进后算法对无人机图像小目标检测的有效性, 本文在 VisDrone2019 数据集上采用 YOLOv5s-6.2 作为基线模型, 与 YOLOv5s、YOLOv7-tiny、YOLOv8s、YOLOX 和最新的 YOLOv8s 进行对比试验, 各算法的实验数据如表 6 所列.

由于 VisDrone2019 数据集无人机图像分辨率较大, 小目标数量众多, 存在大量形似物, 目标特征极易弱化, 造成漏检错检情况, 检测精度普遍较低, 基线算法 mAP_{50} 值为 27.5%, mAP_{75} 值为 13.5%, 改进后的 CHD-YOLOv5 模型 mAP_{50} 值为 35.1%, mAP_{75} 值为 19.3%, 较基线算法 YOLOv5s 的 mAP_{50} 值提升了 7.6%, mAP_{75} 值提升了 5.8%; 与其他模型相比, 较 YOLOv7-tiny 的 mAP_{50} 值提升了 4.9%, mAP_{75} 值提升了 4.1%; 较 YOLOX 的 mAP_{50} 值提升了 7.4%, mAP_{75} 值提升了 5.2%; 较 YOLOv8s 的 mAP_{50} 值提升了 2.4%, mAP_{75} 值提升了 0.1%; 较 Light-RCNN 的 mAP_{50} 值提升了 12.5%, mAP_{75} 值提升了 7.0%; 较 Faster-RCNN 的 mAP_{50} 值提升了 9.2%, mAP_{75} 值提升了 5.6%; 较 BA-YOLOv5s 的 mAP_{50} 值提升了 1.7%. 相比于基线算法 YOLOv5s, CHD-YOLOv5 模型在个别检测精度较高的类别上, 精度值有略微下降, 但在检测精度较低的小目标数量较多的类别上, 精度值均有提升, 其中单类别精度值提升最大的高达 50.31%. 且 mAP_{50} 值和 mAP_{75} 值均有明显提升. 相比于其他模型, 虽然本文提出的 CHD-YOLOv5 模型的模型大小和参数量都有所增加, 但是模型性能有明显增强, 对无人机小目标图像的检测效果更好. 综合各项指标观察, 改进后模型综合性能更好, 有明显的优越性.

表 6 不同算法在 VisDrone2019 测试集上的对比分析 (%)

算法	Pedestrain	People	Bicycle	Car	Van	Truck	Tricycle	Awning-tricycle	Bus	Motor	mAP_{50}	mAP_{75}
YOLOv5s	**44.1**	**40.3**	7.83	72.3	13.3	28.4	15.3	4.57	5.99	**43.2**	27.5	13.5
YOLOv7-tiny	35.4	33.0	4.49	74.5	33.4	21.5	16.0	7.56	40.2	39.9	30.6	15.2
YOLOXs	15.3	11.7	12.3	50.5	**41.4**	30.5	**21.8**	16.7	39.7	22.6	27.7	14.1
YOLOv8s	28.2	15.1	11.1	72.7	38.1	**41.0**	7.8	17.8	**58.6**	30.3	33.1	19.2
Light-RCNN	19.3	20.6	7.2	52.8	28.7	18.3	12.6	8.6	38.5	19.5	22.6	12.3
Faster-RCNN	20.9	18.2	8.2	56.3	26.5	23.4	13.2	9.1	40.3	21.6	25.9	13.7
BA-YOLOv5s	32.5	18.8	11.9	74.8	35.6	39.0	17.6	18.2	57.6	28.5	33.4	—
CHD-YOLOv5	36.6	23.6	**12.4**	**76.8**	40.0	26.8	18.5	**18.4**	56.3	31.8	**35.1**	**19.3**

注: 加粗数据为最优值

2.6 泛化能力测试

为了证明 CHD-YOLOv5 模型的泛化能力, 另外选取 DOTA 1.0 数据集和 HRSID 数据集进行实验. CHD-YOLOv5 算法在 DOTA 1.0 数据集和 HRSID 数据集的检测结果集如表 7 所列.

2017 年 11 月 28 日, 武汉大学在 arXiv 上发布了 DOTA 数据集, 2018 年 6 月又在 IEEE 计算机视觉和模式识别会议 (CVPR) 上发布了 DOTA 数据集. DOTA 1.0 是用于航空图像中目标检测的大规模数据集, 共包含图片 2 806 张, 图片来源于 Google Earth 不同传感器和平台的不同尺度、方向和形状的物体, 图片的像素尺寸在 800×800 到 4000×4000 的范围内. DOTA 图像的目标类别 15 个常见类别, 数据集包含由任意四边形标记的 188, 282 个实例. HRSID 数据集于 2020 年 1 月由电子科技大学发布的, HRSID 数据集是用于船舶检测、语义分割和实例分割任务的高分辨率 SAR 图像数据集, 共包含 SAR 图像 5 604 张, 完全标注的 ship 实例 16 951 个. HRSID 数据集包括不同分辨率的 SAR 图像、极化、海况、海域和沿海港口. DOTA 1.0 数据集是航空影像图像, 尺度变化性较大, 小目标数量多且分布密集, HRSID 数据集是高分辨率舰船图像, 目标单一. 实验数据表明, CHD-YOLOv5 算法在 DOTA 1.0 数

据集和 HRSID 数据集上的检测效果均有提升, CHD-YOLOv5 模型具有良好的泛化能力.

表 7 CHD-YOLOv5 在 2 个数据集上的实验结果 (%)

Dataset	Class	YOLOv5s		CHD-YOLOv5	
		AP	mAP_{50}	*AP*	mAP_{50}
DOTA 1.0	Plane	66.3	69.4	68.8	71.6
	Ship	84.9		86.9	
	Storage tank	91.0		92.3	
	Baseball diamond	69.5		80.2	
	Tennis court	87.4		89.6	
	Basketball court	84.7		84.5	
	Ground track field	62.0		60.9	
	Harbor	46.4		49.6	
	Bridge	93.2		93.7	
	Large vehicle	63.2		62.7	
	Small vehicle	76.1		75.3	
	Helicopter	56.0		58.7	
	Roundabout	62.2		69.8	
	Soccer ball field	46..2		47.3	
	Swimming pool	52.5		53.5	
HRSID	Ship	90.9	90.9	93.7	93.7

3 结论与展望

由于无人机图像背景复杂、小目标较多, 由于无人机图像小目标检测易出现漏检误检现象, 本文提出了一种高阶深度可分离无人机图像小目标检测算法以增强模型检测效果. 该算法在主干网络部分, 融合 ConvMixer 与 CSPNet 结构, 增大模型感受野, 大大提升了模型检测性能; 结合 C3 模块与递归门控卷积构建 C3HB 模块, 使模型的高阶空间交互能力提升, 增强了模型的特征信息提取和表达能力, 提高算法对无人机图像小目标的高阶建模能力; 此外, 将 YOLOv5 算法原本的预测部分替换为精简的解耦头部, 加快网络收敛速度, 提高网络的收敛效果; 最后, 边界框损失函数采用 EIoU, 分别对预测和真实框之间宽和高的预测结果进行惩罚, 提高收敛速度和边界框定位准确度, 精确定位检测框. 在 VisDrone2019 数据集上的实验结果有效表明, 本文所提出的模型有效提升了检测精度, 大大降低了无人机图像小目标的错检率和漏检率, 模型性能提升明显, 但模型的参数量有所增加, 在轻量化方面仍有进步空间, 在以后的工作中, 需继续研究探索如何实现一个实时的高性能检测模型. 本文算法可充分与实际结合, 在工农业、城市规划等实际应用中有较高实用价值.

参考文献

1 朱华勇, 牛轶峰, 沈林成, 等. 无人机系统自主控制技术研究现状与发展趋势. 国防科技大学学报, 2010, 32(3): 115–120. [doi: 10.3969/j.issn.1001-2486.2010.03.022]

2 徐光达, 毛国君. 多层级特征融合的无人机航拍图像目标检测. 计算机科学与探索, 2023, 17(3): 635–645. [doi: 10.3778/j.issn.1673-9418.2205114]

3 向昌成, 黄成兵, 罗平, 等. 基于 YOLO 算法的无人机航拍图像车辆目标检测系统研究. 计算机与数字工程, 2021, 49(8): 1566–1570. [doi: 10.3969/j.issn.1672-9722.2021.08.012]

4 冒国韬, 邓天民, 于楠晶. 基于多尺度分割注意力的无人机航拍图像目标检测算法. 航空学报, 2023, 44(5): 326738.

5 Girshick R, Donahue J, Darrell T, *et al*. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014. 580–587.

6 Girshick R. Fast R-CNN. Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015. 1440–1448.

7 Ren SQ, He KM, Girshick R, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks. Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal: MIT Press, 2015. 91–99.

8 Wang JL, Luo JX, Liu B, *et al*. Automated diabetic retinopathy grading and lesion detection based on the modified R-FCN object-detection algorithm. IET Computer Vision, 2020, 14(1): 1–8. [doi: 10.1049/iet-cvi.2018.5508]

9 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.

10 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.

11 Redmon J, Farhadi A. YOLO9000: Better, faster, stronger. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 6517–6525.

12 Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv:1804.02767, 2018.

13 Bochkovskiy A, Wang CY, Lia OHM. YOLOv4: Optimal speed and accuracy of object detection. arXiv:2004.10934, 2020.

14 Lin TY, Goyal P, Girshick R, *et al*. Focal loss for dense

object detection. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 2999–3007.

15 程江川, 王伟, 康林, 等. 基于改进 YOLOv5 与嵌入式平台的多旋翼无人机检测算法. 兵工自动化, 2023, 42(4): 74–78.

16 李利霞, 王鑫, 王军, 等. 基于特征融合与注意力机制的无人机图像小目标检测算法. 图学学报, 2023, 44(4): 658–666.

17 韩俊, 袁小平, 王准, 等. 基于 YOLOv5s 的无人机密集小目标检测算法. 浙江大学学报 (工学版), 2023, 57(6): 1224–1233

18 李杨, 武连全, 杨海涛, 等. 一种无人机视角下的小目标检测算法. 红外技术, 2023, 45(9): 925–931.

19 奉志强, 谢志军, 包正伟, 等. 基于改进 YOLOv5 的无人机实时密集小目标检测算法. 航空学报, 2023, 44(7): 327106.

20 Ng D, Chen YQ, Tian B, *et al*. Convmixer: Feature interactive convolution with curriculum learning for small footprint and noisy far-field keyword spotting. Proceedings of the 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Singapore: IEEE, 2022. 3603–3607.

21 Rao YM, Zhao WL, Tang YS, *et al*. Hornet: Efficient high-order spatial interactions with recursive gated convolutions. Proceedings of the 36th Conference on Neural Information Processing Systems. New Orleans: OpenReview.net, 2022. 10353–10366.

22 Zhang YF, Ren WQ, Zhang Z, *et al*. Focal and efficient IoU loss for accurate bounding box regression. Neurocomputing, 2022, 506: 146–157. [doi: 10.1016/j.neucom.2022.07.042]

23 Wang CY, Liao HYM, Wu YH, *et al*. CSPNet: A new backbone that can enhance learning capability of CNN. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Seattle: IEEE, 2020. 1571–1580.

24 Ge Z, Liu ST, Wang F, *et al*. YOLOX: Exceeding YOLO Series in 2021. arXiv:2107.08430, 2021.

(校对责编: 牛欣悦)