•计算机科学与技术•

DOI:10.15961/j.jsuese.202300593

本刊网刊

基于改进YOLOv7–tiny的无人机航拍图像小目标检测算法

张光华[1],李聪发[1*],李钢硬[1],卢为党[2]

(1.东北石油大学 电气信息工程学院,黑龙江 大庆 163318;2.浙江工业大学 信息工程学院,浙江 杭州 310014)

摘　要:无人机航拍图像目标检测是无人机应用的一项重要技术,针对无人机航拍图像中目标尺度变化大、小尺寸目标分布密集、背景复杂而导致的漏检和误检问题,本文提出一种基于YOLOv7−tiny带ConvMixer检测头的无人机航拍图像小目标检测算法。首先,将激活函数LeakyReLU替换为SiLU,弥补LeakyReLU缺少的非线性表达,提升模型训练时的收敛速度与模型泛化能力;其次,为了增强对多尺度目标的特征提取能力,额外设计了小目标检测层,并衍生出一个微小目标检测头,增大了模型感受野,更好地解决目标尺度剧烈变化带来的大尺度方差问题,提升了小目标的检测能力;此外,在预测头部分集成ConvMixer层,ConvMixer中的深度卷积和逐点卷积有助于找到传递给预测头的特征信息中的空间和通道关系,提升对微小目标的处理能力;最后,将YOLOv7−tiny的耦合检测头替换为更高效的解耦头,对定位与分类任务解耦出单独的特征通道,增强对目标的分类和定位能力。为了全面验证每个改进点的有效性,本文从两个方向设计了消融实验,并对比分析了改进算法与其他算法的检测性能。实验结果表明,本文算法在Visdrone2021数据集上平均精度均值(mAP)达到40.9%,较基线算法提升了3.7%,模型内存为28.2 MB,检测速度达到35.8 帧/s,改进算法综合性能与对比的主流先进算法相比更优。通过检测效果分析可知,本文算法在无人机航拍图像检测上的误检和漏检问题得到较大改善。综上,本文算法的准确性和实时性能胜任航拍图像小目标检测任务。

关键词:无人机航拍图像;小目标检测;SiLU;ConvMixer;更高效的解耦头

中图分类号:TP391.4　　**文献标志码:**A　　**文章编号:**2096-3246(2025)03-0235-12

无人机具备简便操控、灵活机动、性能出色等优势[1],其数据采集能力强大,运营成本低廉,且便于运输,适用范围广泛,可在多样化场地和复杂环境中高效执行任务。随着无人机技术与计算机视觉技术的迅速发展,配备深度学习技术的无人机被广泛应用于城市交通、军事侦察、智慧农业[2–4]等实际应用领域,然而由于无人机航拍角度多变、成像环境复杂,导致无人机航拍图像检测存在如下难题:1)航拍视野大,背景复杂,目标尺度差异大,导致小目标不易定位,难以辨别检测;2)小目标像素信息少,缺乏区分自身与背景的外观信息,导致漏检和误检的情况;3)在精度和实时性平衡中遇到难题。

目前主流的目标检测算法主要基于深度学习模型,可分为双阶段算法和单阶段算法两类。双阶段算法通常需要生成目标区域候选框,然后通过卷积神经网络对候选框进行分类和定位,例如Fast−RCNN[5]、Faster R−CNN[6]、Cascade R−CNN[7]。单阶段算法结构更简单,以YOLO[8]和SSD[9]算法为例,仅使用卷积神经网络提取特征,直接预测不同目标的类别和位置。因此,单阶段算法相较于双阶段算法在实时性方面更优。

Liu等[10]提出一种基于空间坐标自注意力机制的无人机航拍检测器,对空间注意模块和坐标注意模块进行修改和组合,形成一个新颖的空间坐标自注意模块,并提出Slim−BiFPN网络使算法参数量较原来减少了25.8%,且有效避免了精度损失,但检测速度下降了12.1%。Luo等[11]针对无人机航拍图像检测存在的因目标小、排列密集、分布稀疏、背景复杂等挑战而表

收稿日期:2023 – 08 – 03　　**修回日期:**2023 – 11 – 29　　**网络出版日期:**2024 – 04 – 19

基金项目:国家自然科学基金项目(62271447)

作者简介:张光华(1979—),男,副教授,博士. 研究方向:信号处理. E-mail:dqzgh@nepu.edu.cn

* **通信作者:**李聪发,E-mail:18296611707@163.com

现不佳的问题，提出了使用非对称卷积的特征提取模块对 YOLOv5 主干中不同位置的残差块进行了相应的替换，有效提升了检测精度，模型参数量和计算复杂度有一定增加但在合理范围。Jiang 等[12]提出一种轻量级的实时目标检测网络（VC-YOLO），在主干网络引入有限数量卷积层，在特征金字塔网络增加更多的横向连接，在检测部分引入通道和空间注意模块，与基线模型 YOLOv3 相比，模型以较少的计算开销实现了检测精度的提升，但检测性能的提升幅度不大。Huang 等[13]基于 YOLOv4 提出 BLUR-YOLO 算法，使用基于模糊滤波器的池化结构（BlurPool）代替原有下采样方法，并提出特征金字塔网络（Blur-PANet）来有效融合多层特征，检测精度提升了 1.2%，但没有平衡模型复杂度和检测精度。冒国韬等[14]基于 YOLOv5 引入多尺度分割注意力单元与自适应加权特征融合模块，平均精度均值达到 34.7%，但模型内存和复杂度巨大，没考虑精度和实时性的平衡。徐坚等[15]基于 YOLOv5 网络结构在骨干网络中加入可变性卷积，并设计特征平衡金字塔结构和交叉自注意力模块，在保证实时性的同时改善了小目标的检测效果。徐光达等[16]采取添加小目标检测层、多层级特征融合层和解耦检测头的措施提升了检测精度，缺点是改进模型单方面追求性能而导致参数量和复杂度计算量太大，几乎没有考虑模型的轻量化部署。陈卫彪等[17]采用加深特征融合网络深度、增加额外的小目标检测头和将普通卷积模块替换为深度可分离卷积模块 3 种方式改进 YOLOv5s 模型，达到提升小目标检测能力和减小模型参数量的目的，缺点是网络层数较多，检测精度有待提升。在天津大学机器学习与数据挖掘实验室发布的 VisDrone-DET2020 数据集目标检测挑战赛[18]中，该实验室团队收集了 29 种检测算法，其中性能较优的算法有 DroneEye2020、DMNet 等。DroneEye2020 在 Cascade R-CNN 模型的基础上进行改进，在颈部网络（Neck）使用递归特征金字塔（RFP），并使用可切换多孔卷积（SAC）以获得更好的性能；DMNet 则使用密度裁剪加均匀裁剪的策略来训练基线模型，并采用融合检测以获得更好的检测效果。Baidya 等[19]基于 YOLOv5 网络结构在预测头中加入 ConvMixer 模块，并额外增加一个检测头来处理无人机航拍图像中的微小目标，其检测效果与在 VisDrone 2021 无人机视觉挑战赛中获得第 5 名的 TPH-YOLOv5[20]效果相当。

但上述研究存在以下问题：首先，上述方法大都通过引入注意力机制获取目标上下文信息，以此增加有效信息的权重，达到增强小目标的检测能力的目的，但忽视了网络的参数量增加和算力需求变大问题；其次，为解决深层网络在提取特征后会削弱对小目标位置和分类等细节信息的感知问题，上述绝大多数算法模型都提出了特征融合模块。此外，现阶段方法没有很好地平衡检测性能与实时性，一方面，追求最优检测性能往往导致模型复杂度极高；另一方面，轻量化后的模型检测性能提升不大。

针对参数量和算力需求增加问题，本文弃用注意力机制而采用 ConvMixer 模块，通过分离空间和通道维度的混合，捕获全局信息和上下文信息，提升对小目标的处理能力，同时在整个处理过程中保持输入特征图大小和分辨率不变。针对深层网络削弱细节感知问题，在网络检测头部分采用更高效的解耦头，为检测小目标提供更精确的边界框坐标和目标类别信息；此外，为进一步加强对无人机航拍图像小目标的特征提取能力，将激活函数 LeakyReLU 换为 SiLU，加快收敛速度，增强特征提取网络性能；为处理无人机航拍图像目标尺度剧烈变化而难以辨别问题，额外设计了小目标检测层并增加一个检测头。针对性能和实时性平衡问题，本文在保持模型参数量和计算复杂度较小的前提下，有效地提升了模型检测精度。

1 YOLOv7-tiny 网络结构

YOLOv7-tiny 是 YOLOv7 网络简化而来，是为了在嵌入式设备、移动设备中部署而设计的网络模型，由输入端、骨干网络（Backbone）、颈部网络（Neck）及预测头（Prediction head）4 部分组成，其结构如图 1 所示。

输入端采用 Mosaic 和 Mixup 等数据增强技术丰富数据集，加快训练速度，以及对图像进行归一化、通道顺序调整等预处理，再输入至骨干网络进行特征提取。骨干网络部分主要是特征提取网络。

用 CBL 模块提取原始特征，CBL 模块由卷积层（Conv）、批量归一化层（BN）和 LeakyReLU 激活函数组成，使用高效层聚合网络（ELAN）代替扩展高效层聚合网络（E-ELAN），通过控制最短和最长的梯度路径，使网络能够学习到更多的特征，通过最大池化层（MP）对特征图进行下采样，以减小尺寸并提取主要特征信息，同时提高模型的计算效率。Neck 部分主要是特征融合网络，其中，SPPCSP 模块是结合空间金字塔池化（SPP）和跨级局部网络（CSP）结构的改进模块，用于增强网络的多尺度特征提取能力；通过上采样（UpSample）和张量拼接操作（Concat），将来自不同层级的特征图进行融合，以提升目标检测性能；利用特征金字塔网络进行多尺度特征融合，提升目标检测的准确性并增大多尺度目标的感受野。与 YOLOv7 不同，YOLOv7-tiny 预测头部分用标准卷积（Conv）替代重参数化卷积（RepConv）进行通道数调整，并采用隐式检测头（IDetect head），Prediction1~3 为预测模块。

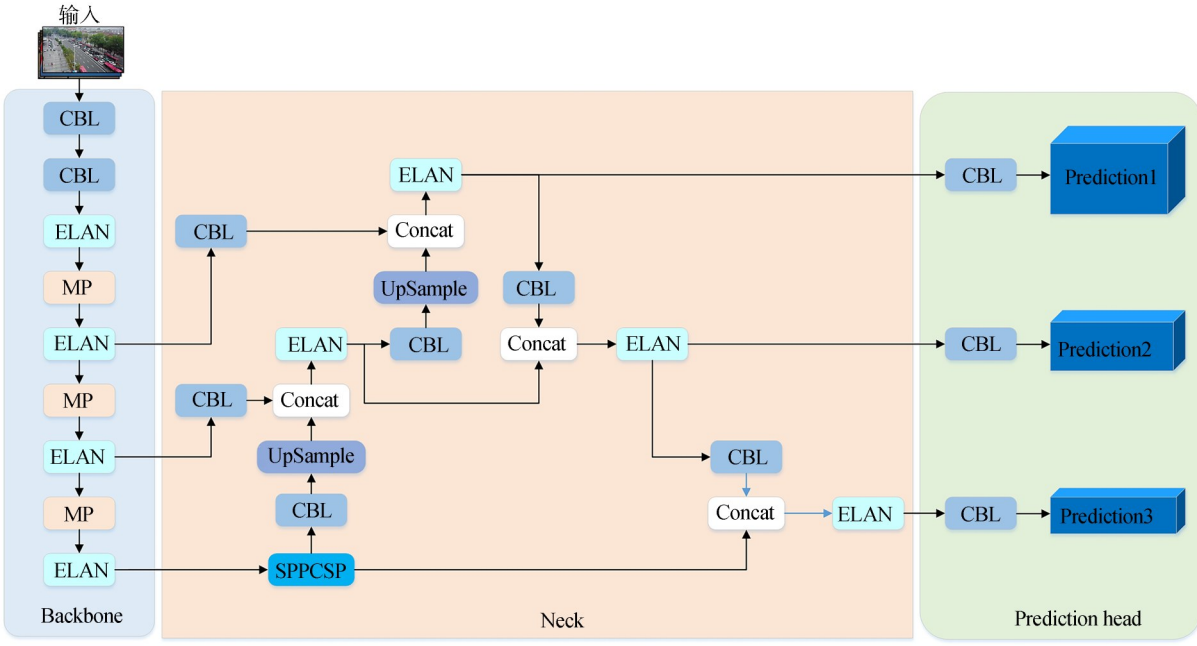


图 1　YOLOv7–tiny 网络结构

Fig. 1　Network structure of Yolov7–tiny algorithm

YOLOv7–tiny 通过简化结构牺牲了一定的精度，优点是提升了检测速度，在轻量化部署方面具有优势。但也存在不足：1）使用的 LeakyReLU 激活函数缺少非线性表达，特征越向下传输时，梯度更新曲线不够光滑，影响检测精确度；2）主干网络部分与 Neck 部分大量使用 ELAN 网络，使模型架构和参数量复杂化，在特征融合时也会导致特征信息的冗余；3）预测头部分缺乏处理小目标的能力。为此，本文主要围绕改善特征融合网络、预测头部分来提升对航拍图像小目标的检测性能。

2　改进 YOLOv7–tiny 算法网络结构

本文算法详细架构如图 2 所示，主要对特征融合网络部分和预测头部分进行了改进。

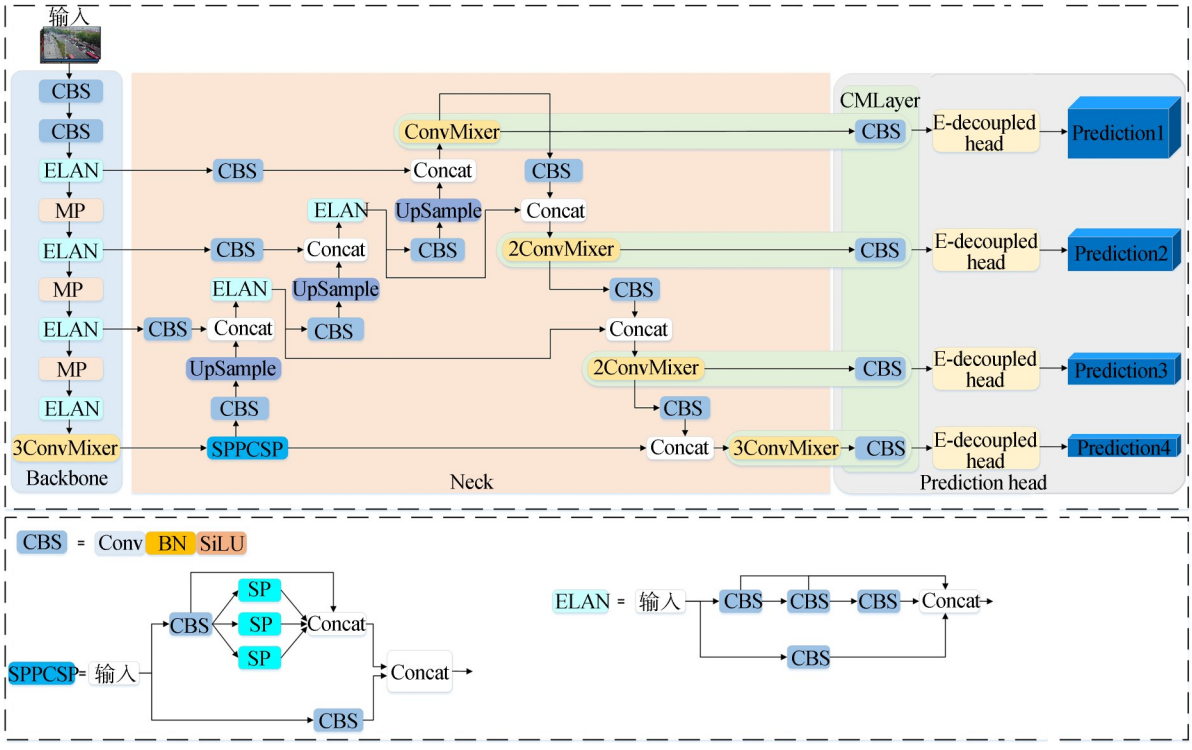


图 2　改进后的 YOLOv7–tiny 网络结构

Fig. 2　Improved YOLOv7–tiny network structure

图 2 中，将 CBL 集成模块中的激活函数替换为 SiLU 激活函数，构成 CBS 模块，其中，SPPCSP 及 ELAN 结构中的卷积层也改为 CBS 模块，SPPCSP 模块中的空间金字塔层（SP）增强了对多尺度目标的检测能力。在模型 Neck 端引入了一层小目标检测层，并增加了一个预测头，提升极小目标的检测能力。另外，在算法模型的主干网络末端及预测头中集成相应数量的 ConvMixer 层，与 CBS 模块构成图 2 中的 CMLayer，ConvMixer 层可以帮助模型捕获全局信息以及丰富上下文小目标特征信息，增强对小目标的处理能力。为了使模型更好地处理航拍图像中小目标的定位与分类任务，提升两类任务的精确度，本文采用更高效结构的解耦头（E-decoupled head）预测目标，解耦出两个单独的特征通道用于边界框坐标回归和对象分类任务。图 2 中，输入 Prediction1~4 的特征图尺寸依次减小，对应的感受野依次增大，Prediction1 用于检测极小尺寸的目标。

2.1 SiLU 激活函数

为了确保特征提取网络的准确性，本文针对 LeakyReLU 激活函数存在的问题进行了改进。算法选择 SiLU 激活函数作为替代，SiLU 激活函数在零点处具有最小值，能够有效地缓冲权重并保持网络的稳定性。相比 LeakyReLU，SiLU 在 0 点附近表现出更加平滑和较强的可导性，这对于梯度计算和更新非常有帮助，加快了模型的收敛速度；此外，SiLU 增加了更多的非线性表达，从而增强了模型泛化性。总体上，SiLU 激活函数的使用提升了模型的准确度。两者的曲线平滑度对比如图 3，LeakyReLU 和 SiLU 激活函数的表达式如下：

$$\mathrm{LeakyReLU}(x)=\begin{cases}x, x>0;\\ \alpha x, x\le 0\end{cases} \tag{1}$$

$$\mathrm{SiLU}(x)=x\frac{1}{(1+\mathrm{e}^{-x})} \tag{2}$$

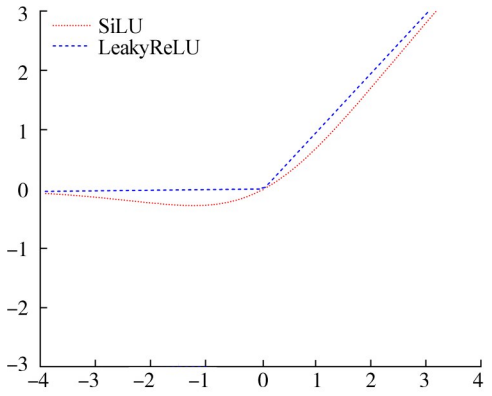


图 3 LeakyReLU 和 SiLU 激活函数对比

Fig. 3 Comparison of LeakyReLU and SiLU activation functions

式（1）~（2）中：x 为经过卷积和批归一化处理后的多维特征张量的元素；α 为 LeakyReLU 函数的斜率系数，$0<\alpha<1$，用于控制负值区域的输出行为。

2.2 ConvMixer 架构

ConvMixer[21] 通过构建纯卷积的图像块（Patch）处理框架，验证了 Vision Transformer（ViT）[22] 性能提升的关键因素是图像块划分策略，而非自注意力机制。ConvMixer 是最近计算机视觉应用研究中出现的一种新型架构，旨在对 ViT 架构性能进行挑战，其主要思想是通过在空间和通道上混淆数据来提升性能。在足够的实验数据支持下，ConvMixer 架构在性能方面证明与传统的 ViT 架构相当，同时在计算时间和资源占用方面更加高效。

在本文的改进 YOLOv7-tiny 模型的检测头部分采用了 ConvMixer 架构增强小目标检测性能。预测头中的 ConvMixer 有助于捕捉传递给预测头特征中的空间和通道关系，这种关系是通过 ConvMixer 中的深度卷积和逐点卷积来发现的。ConvMixer 中的逐点卷积还能增强预测头对微小物体的检测能力，这是因为它处理的是单个数据点级别的信息。此外，与传统的卷积神经网络不同，ConvMixer 在整个混合器层中保持输入结构的完整性，使其非常适用于目标检测架构中的预测头部分。ConvMixer 层的架构如图 4 所示。

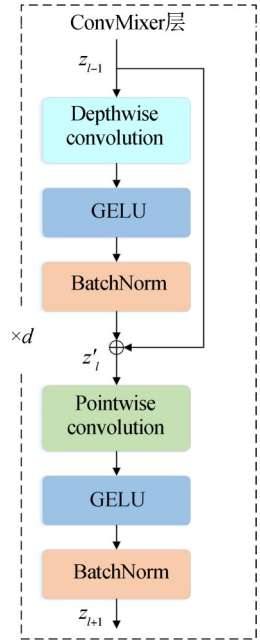


图 4 ConvMixer 架构

Fig. 4 Architecture of the ConvMixer

图 4 中，ConvMixer 层本身由深度卷积（Depthwise convolution）和逐点卷积（Pointwise convolution）组成，深度卷积为分组卷积，使用大卷积核对特征进行空间混合，其组数等于通道数，逐点卷积是核大小为 1×1 的

卷积。每个卷积之后都用激活函数GELU和激活后的批量归一化函数（BatchNorm）处理。ConvMixer层可复用，d为重复次数。过程表达式如下：

$$z_l'=\mathrm{BN}(\sigma(\mathrm{ConvDepthwise}(z_{l-1})))+z_{l-1} \quad (3)$$

$$z_{l+1}=\mathrm{BN}(\sigma(\mathrm{ConvPointwise}(z_l'))) \quad (4)$$

式（3）~（4）中，ConvDepthwise 和 ConvPointwise 分别为深度卷积和逐点卷积操作，z_{l-1}为前一层（即l–1层）输入的特征图，z_l'为经过深度卷积空间混合和残差连接后的中间特征，z_{l+1}为经过逐点卷积通道混合后的特征，σ为GELU激活函数，BN为对特征图进行归一化处理。

2.3　Efficient解耦头

YOLOv7–tiny的检测头为耦合检测头，其分类和定位任务共享相同参数。耦合头与Efficient解耦头对比如图5所示。图5中，n_a为预定义锚框数量，耦合头将特征金塔网络（FPN）的特征信息通过卷积改变通道维数，将锚框（anchor）、边界框坐标、目标类别等信息进行耦合处理，完成对图像中目标的定位与分类任务。文献[23]通过对比经过全连接头（Fully connected head）和卷积头（Convolution head）的输出特征图得出：卷积头更适合定位任务，但由于缺乏空间敏感性，对于分类任务并不具有鲁棒性；但全连接头比卷积头更具空间敏感性及更强的能力来区分完整对象和部分对象，所以更适合分类任务。因此，作者提出了集成全连接头和卷积头的双检测头机制来更好地完成目标检测任务。相较于耦合检测头，FCOS[24]和YOLOX[25]模型将检测头的分类和定位任务解耦为两个分支，并且在每个分支中引入了额外的两个卷积核大小为3×3的卷积层以提高性能，取得了较好的分类与定位性能。上述文献论证说明分类任务与定位任务专注的信息不同，需要各自适合的通道才能得到最佳结果。

无人机航拍图像具有目标尺度小、类别多、尺度剧烈变化、分布密集、背景复杂的特点，所以无人机航拍图像目标检测任务需要更精确的定位信息和丰富的分类信息，所以在改进的YOLOv7–tiny网络架构中采用更高效的解耦合头（Efficient decoupled head）预测目标，其优点是采用混合通道（Hybrid channels）策略设计，并对定位与分类任务解耦出单独的特征通道。相比普通解耦头，本文解耦头进行了精简设计：首先，将中间卷积核大小为3×3的卷积层的数量减少到只有1个；同时，相比普通解耦头扩张通道数，本文解耦头维持通道数不变以降低计算量，并综合考虑了相关算子表征能力和硬件上的计算开销这两者的平衡，在维持精度的同时降低了延时。图5中，以高H、宽W、通道数为256的特征图作为输入。首先，经过卷积核大小为1×1的卷积降低特征通道数或通道数不变（本文保持不变，仅特征变换）；然后，输出分为两个分支，上分支经过一个核大小为3×3的卷积完成特征信息的提取后，再调整特征通道数降至为数据集目标类别数C后完成分类（Cls）任务。下分支则是经过3×3卷积特征信息提取后，再分为两条支路，回归（Reg）支路负责获取中心2维坐标、边界框的高度及宽度4个参数，置信度（Obj）支路负责获取1个置信度参数。

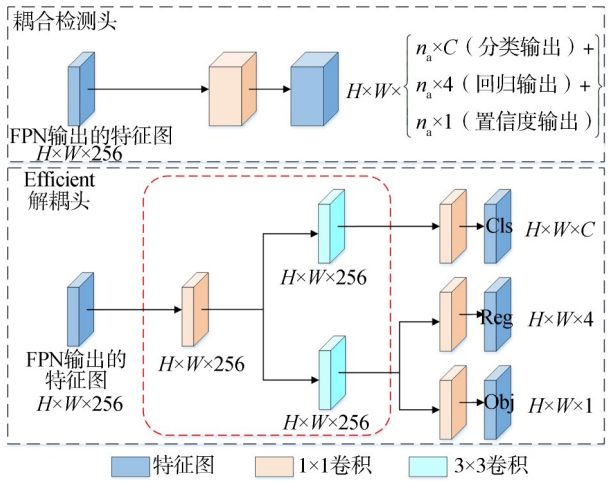


图5　耦合头与Efficient解耦头对比

Fig. 5　Comparison of coupling head and Efficient decoupled head

3　实验结果与分析

3.1　实验环境与配置

本文模型训练的实验环境搭载的操作系统为Ubuntu18.04.6，处理器型号为Intel Xeon（R）Bronze 3204 CPU@1.9 GHZ×12，显卡为显存大小24 G的NVIDIA Quadro P6000，深度学习框架为Pytorch，并采用CUDA 11.3作为加速训练架构。模型训练之前需要优化相关参数，主要有批量大小、初始学习率、权重衰减值、训练轮数及图像尺寸。具体参数配置见表1。

表1　实验参数配置

Tab. 1　Experimental parameters configuration

批量大小	初始学习率	权重衰减值	训练轮数	图像尺寸/像素×像素
8	0.01	0.000 5	300	640×640

3.2　数据集

本文算法在天津大学机器学习与数据挖掘实验室发布的Visdrone2021[26]目标检测数据集上进行实验。该数据集图像由无人机装载摄像头在中国14个不同城市采集，采集环境复杂多变：包括不同场景（城市和乡村、目标稀疏和密集场景等）、天气情况、光照

条件。Visdrone2021 数据集有 10 209 幅图像，其中，训练集 6 471 幅，验证集 548 幅，测试集 1 610 幅及不附带标签的测试挑战集 1 580 幅，共有 10 类目标：行人（pedestrian）、人（people）、自行车（bicycle）、汽车（car）、面包车（van）、卡车（truck）、三轮车（tricycle）、遮阳三轮车（awning-tricycle）、公共汽车（bus）、摩托车（motor）。图 6 展现了 Visdrone2021 数据集中所有目标的宽高分布。图 6 中，目标宽和高的值已归一化，区域颜色越深表示数量越多越密集。从图 6 中可以看出，目标主要集中分布在宽高极小的区域，大尺度目标分散分布，说明该数据集主要由小目标组成。

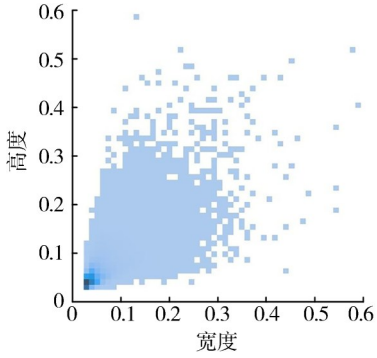


图 6 目标宽高分布

Fig. 6 Width and height distribution of all targets

3.3 评价指标

评估模型性能的主要指标有精确率（记为 P）、召回率（记为 R）、平均精度（AP，记为 P_A）、平均精度均值（mAP，记为 P_{mA}）、每秒传输帧数（FPS）、网络参数量、模型复杂度计算量（FLOPs），主要指标计算公式如下：

$$P=\frac{n_{TP}}{n_{TP}+n_{FP}} \tag{5}$$

$$R=\frac{n_{TP}}{n_{TP}+n_{FN}} \tag{6}$$

$$P_A=\int_0^1 PR\mathrm{d}R \tag{7}$$

$$P_{mA}=\frac{1}{N}\sum_{i=1}^{N}(P_A)_i \tag{8}$$

式（5）~（8）中，n_{TP} 为预测正确的正样本数量，n_{FP} 为预测错误的正样本数量，n_{FN} 为预测错误的负样本数量，N 为 P_A 的总类别数，$(P_A)_i$ 为第 i 个类别的 P_A 值。

3.4 消融实验

为探讨各改进点对网络模型的有效性，本文将 YOLOv7-tiny 激活函数替换为 SiLU，命名为 YOLOv7-tiny-S；同理，将增加第 4 个检测头的 YOLOv7-tiny 模型命名为 YOLOv7-tiny-F，将引入 ConvMixer 架构的模型命名为 YOLOv7-tiny-C，将引入更高效解耦头的模型命名为 YOLOv7-tiny-E。

本文从两个方向设计了消融实验：一方面，在原有 YOLOv7-tiny 算法基础上，仅增加一种改进模块验证该模块的对基准模型的影响；另一方面，在最终改进模型 YOLOv7-tiny-SFCE 上，仅消除一种改进方法并验证其对最终改进模型的影响。

在同等实验条件下进行了 10 组消融实验，结果如表 2 所示。对表 2 分析如下：

表 2 改进点消融实验

Tab. 2 Ablation experiments of improved points

方法	P/%	R/%	网络参数量/10^6	FLOPs/10^9	P_{mA}/%	FPS/(帧·s^{-1})
YOLOv7-tiny	47.5	39.7	6.03	13.3	37.2	82.6
YOLOv7-tiny-S	49.2	40.2	6.03	13.3	37.8	86.9
YOLOv7-tiny-F	49.8	40.7	6.10	15.5	37.9	62.1
YOLOv7-tiny-C	49.1	39.3	6.39	14.1	37.8	64.1
YOLOv7-tiny-E	49.2	39.8	12.50	25.5	38.3	69.4
YOLOv7-tiny-FC	48.7	40.2	7.96	19.4	38.8	40.6
YOLOv7-tiny-SFC	50.2	39.7	7.96	19.4	39.0	45.2
YOLOv7-tiny-SFE	50.6	41.1	13.90	34.3	40.2	53.1
YOLOv7-tiny-SCE	50.6	40.0	14.30	26.1	38.5	48.3
YOLOv7-tiny-FCE	49.2	41.9	14.50	35.5	39.7	36.6
YOLOv7-tiny-SFCE	51.2	41.9	14.50	35.5	40.9	35.8

1）第 1 组为基准模型的实验，即 YOLOv7-tiny 算法模型，其结果作为改进模型的性能参考，其在 Visdrone2021 数据集的 mAP 值为 37.2%。

2）第 2 组实验把基准模型的激活函数替换为 SiLU（YOLOv7-tiny-S），参数量不变，与基准模型相比，mAP 提升了 0.6%，准确率提升了 1.7%，召回率提升了 0.5%，检测速度 FPS 提升 4 帧/s。模型精度的提升原因主要有：首先，SiLU 激活函数具有非线性特性和更优的梯度传播，可以有助于模型更有效的学习特征和表达模型学习到的特征；其次，SiLU 激活函数信息传递效果更强，有助于网络的收敛过程，更快的收敛伴随着更好的泛化性能。因此，模型在学习复杂特

征时能避免信息损失，提高检测性能。检测速度的提升原因主要为计算效率改善，虽然模型网络参数量和复杂度计算量没有改变，但SiLU激活函数具有更高的计算效率，使得模型在推理时减少一些冗余计算，提高整体检测速度。

3）第3组为增加第4个检测头的实验（YOLOv7−tiny−F），与基准模型相比，mAP提升了0.7%，精确率提升了2.3%，召回率提升了1.0%。参数量以及模型复杂度的增加导致检测速度的降低。指标mAP、精确率和召回率的提升的可能原因如下：首先，新增的小目标检测头直连至浅层网络，而浅层网络的底层特征图具有更高的分辨率和更多的细节信息，增强了对小目标的表征能力，有助于有效检测小目标；其次，第4个检测头能够使模型更大范围地捕捉上下文特征信息，从而更准确地预测小目标的位置和类别；此外，衍生出第4个检测头的小目标检测层为模型带来了更好的特征融合机制，在网络中引入多尺度信息，模型可以更好地适应目标的尺度变化，提高了检测的鲁棒性，从而提升了检测精度。

4）第4组实验为在预测头集成ConvMixer层（YOLOv7−tiny−C），其mAP提升了0.6%，精确率提升了1.6%。mAP和精确率的提升是因为ConvMixer是多层架构，每个层次包含全连接的局部注意力结构和一个多层感知器，有助于模型更好地捕获图像中小目标的局部特征，完成图像目标检测任务。模型FPS的降低是因为ConvMixer中的全连接的局部注意力架构，其中的全连接结构会导致模型的计算复杂度相较于普通卷积有少量增加，在处理复杂场景的大量目标时，会降低模型的推理速度。

为了进一步探索ConvMixer层对微小目标的处理能力，在YOLOv7−tiny−F基础上增加该模块进行实验（即第6组实验YOLOv7−tiny−FC），第6组实验表明ConvMixer层与小目标检测层结构的有机融合对小目标的检测性能提升较明显，更能发挥该模块在检测头部分处理微小目标的能力。

5）第5组实验增加更高效的解耦头（YOLOv7−tiny−C），与基准模型相比，其mAP提升了1.1%，精确率提升了1.7%。更高效的解耦头提供了更精确的边界框坐标回归和分类信息；此外，其混合通道策略设计能够使网络更适应不同尺度、形状和密度分布的目标，从而提升了mAP等指标。参数量增加的原因可能是：相比于YOLOv7−tiny的耦合检测头，高效解耦头将分类与回归分支进行解耦，将分类与检测任务单独进行，导致参数量增加；此外，模型复杂度计算量的增加也导致推理速度变慢。但相较第4组实验可知，模型参数量和FLOPs与检测速度不一定负相关。

6）从第2组至第10组消融实验发现，更高效的解耦头的引入对基准模型检测精度的提升最为明显，mAP提升了1.1%。与最终改进模型YOLOv7−tiny−SFCE相比，第4个小目标检测头的消除对检测性能的影响最为明显，检测精度降低了2.4%。

3.5　对比试验

为了验证本文算法的有效性，选取了MSA−YOLO[14]、VC−YOLO[12]、DSM−YOLOv5[17]等各种无人机航拍图像目标检测先进算法与本文改进算法在Visdrone2021测试集上进行检测性能指标的对比分析，结果如表3所示。表3中，awn-tri表示awning-tricycle。

表3　本文算法与先进算法检测性能指标对比

Tab. 3　Comparison of the detection performance indicators between the proposed algorithm and advanced algorithms

模型	P_A/%										P_{mA}/%
	pedestrian	people	bicycle	car	van	truck	tricycle	awn-tri	bus	motor	
VC−YOLO[12]	31.5	26.2	5.44	72.0	31.3	19.2	15.7	7.60	41.4	33.9	28.4
BLUR−YOLO[13]	25.2	3.4	6.29	65.4	31.9	40.9	11.7	8.13	47.1	18.5	25.9
MSA−YOLO[14]	33.4	17.3	11.20	76.8	41.5	41.4	14.8	18.40	60.9	31.0	34.7
DMNet[18]	28.5	20.4	15.90	56.8	37.9	30.1	22.6	14.00	47.1	29.2	30.3
DroneEye2020[18]	35.7	18.3	14.00	56.5	42.9	37.6	35.4	25.90	50.4	28.9	34.6
Cascade R−CNN[27]	22.2	14.8	7.60	54.6	31.5	21.6	14.8	8.60	34.9	21.4	23.2
YOLOv5s	35.8	30.5	10.10	65.0	31.5	29.5	20.6	11.10	41.0	35.4	31.1
YOLOv3−LITE[28]	34.5	23.4	7.90	70.8	31.3	21.9	15.2	6.20	40.9	32.7	28.5
SDS−YOLO[29]	46.7	35.4	14.40	82.0	45.1	35.8	26.5	12.70	54.3	47.4	40.0
CenterNet[30]	22.6	20.6	14.60	59.7	24.0	21.3	20.1	17.40	37.9	23.7	26.2
DSM−YOLOv5[17]	42.6	32.7	11.90	79.3	41.3	35.6	21.4	11.00	49.7	42.0	36.8
本文模型	49.0	41.4	13.90	82.4	42.4	35.6	26.1	13.90	52.9	51.0	40.9

由表 3 可知：本文算法在 pedestrian、people、car、motor 4 个目标类别的检测性能均取得了最优 AP 值，分别为 49.0%、41.4%、82.4%、51.0%，在 pedestrian、people、motor类别的AP值均取得了较大领先，bus类别AP值仅次于MSA-YOLO，其他目标类别如van、bicycle的AP值也领先大多数对比算法，其中本文算法对van目标类别的检测性能仅次于 DroneEye2020 和 SDS-YOLO，对bicycle类别的检测性能仅次于DMNet、SDS-YOLO和CenterNet；综合对比所有算法的mAP值，本文模型比 VC-YOLO 提升 12.5%，比次优的 SDS-YOLO 算法高 0.9%。综上，本文算法 YOLOv7-tiny-SFCE 与其他先进算法相比，综合性能最优。以上性能参数提升主要归因于不同改进点的有机融合增强了对小尺度目标处理能力和对复杂场景下背景噪声信息的抗干扰能力。由以上分析可知，本文算法对无人机航拍图像中的小目标有较强的检测能力，适合无人机航拍目标检测任务。

3.6 检测效果分析

为了验证本文改进算法在实际场景中的检测效果，选取 Visdrone2021 测试集和测试挑战集中不同复杂场景的无人机航拍图像进行检测，检测效果如图 7 所示。由图 7 可以看出，本文算法在目标稀疏与密集、白天与黑夜和复杂背景场景下，对汽车、行人等小目标，能够有效抑制图像复杂背景中的建筑物、树木等噪声信息的干扰，正确分类和定位每个目标。这说明本文算法在不同背景、目标分布、光照条件等实际场景下都具有较好的检测性能，能胜任无人机航拍图像目标检测任务。

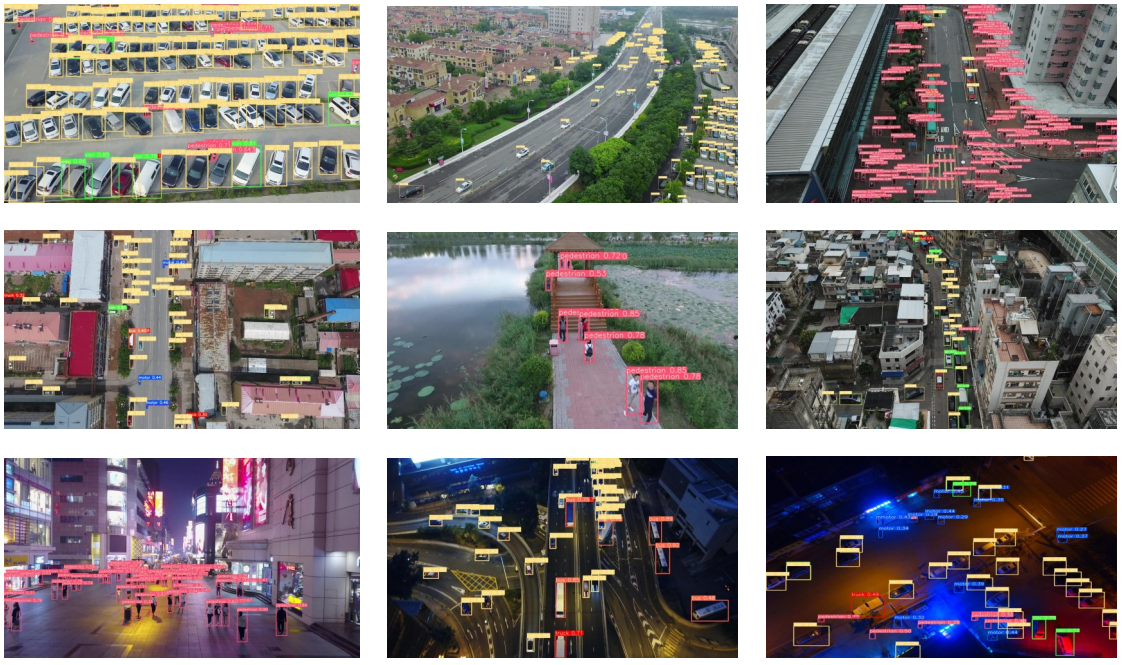


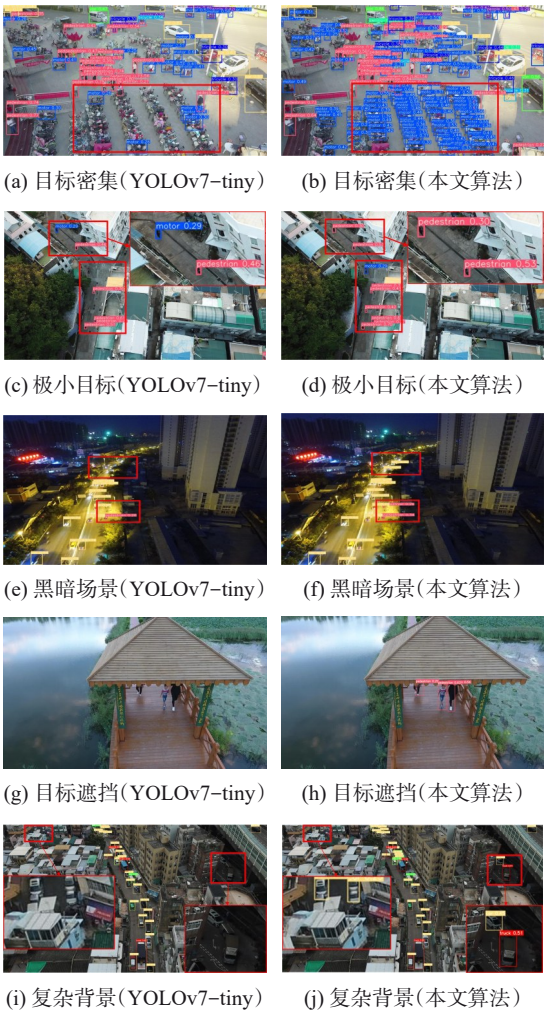
图 7 改进算法在不同场景检测效果

Fig. 7 Detection effect of improved algorithm in different scenarios

为了进一步评价本文算法对无人机航拍图像的检测性能，从 Visdrone2021 测试挑战集随机选取了目标密集、极小目标、黑暗场景、目标遮挡和复杂背景的 5 幅不同场景下的图像，将基线算法 YOLOv7-tiny 和本文改进算法的可视化检测结果进行对比，结果如图 8 所示。

从图 8(a)和(b)可以看出：在目标分布密集场景下，YOLOv7-tiny 对摩托车目标类别有大量漏检情况，对面包车目标类别存在误检为汽车类别的情况；而本文算法能正确定位该类目标并正确识别，这得益于改进后算法网络新增的检测头和解耦头，新增小目标检测头直连至浅层网络，能有效提取密集场景下的目标特征信息，避免了多次下采样后导致的密集目标特征和背景噪声信息混杂在一起的问题，解耦头则提供了更精确的定位和分类信息。对比图 8(c)和(d)可知：对于像素点少的极小目标的检测，YOLOv7-tiny 将行人误检为摩托车，并存在行人和摩托车目标类别的漏检情况；本文算法整体改进后加强了对底层细节信息的提取，对于图像中的所有微小目标都能正确识别。对比图 8(e)和(f)可知：对于黑夜低光照场景下，本文算法能准确识别并检测出来远处微小目标，说明其对弱光照场景具有适应性；而 YOLOv7-tiny 则对汽车和行人目标存在漏检情况。对比图 8(g)和(h)可知：在目标特征被建筑物遮挡的情况下，YOLOv7-tiny

完全漏检3个行人目标，而本文算法则能精确定位并检测出所有行人目标。对比图8(i)和(j)可知：在复杂建筑物背景下，本文算法依旧展现出强大的目标定位与辨识能力，能精确识别出图像边缘和弱光照条件下的车类目标，YOLOv7–tiny则受复杂背景噪声信息的干扰出现漏检情况。



(a) 目标密集(YOLOv7–tiny)　(b) 目标密集(本文算法)

(c) 极小目标(YOLOv7–tiny)　(d) 极小目标(本文算法)

(e) 黑暗场景(YOLOv7–tiny)　(f) 黑暗场景(本文算法)

(g) 目标遮挡(YOLOv7–tiny)　(h) 目标遮挡(本文算法)

(i) 复杂背景(YOLOv7–tiny)　(j) 复杂背景(本文算法)

图8　检测效果对比

Fig.8　Comparison of detection effects

由上述检测效果对比结果可知，本文改进算法在不同场景下对于无人机航拍图像多尺度目标的特征信息的提取能力有较大提升。算法模型特征融合网络部分增加额外的小目标检测层，丰富了特征信息的融合，使模型提取更有利于检测小尺度目标的特征信息。通过替换更高效的解耦头，在目标分布密集、目标尺度小、目标遮挡、不同光照条件、背景复杂等情况下，能够更精确地提取定位信息和丰富的分类信息。网络在增加第4个检测头和集成ConvMixer层的加持下，增强了对于极小目标的检测能力。整体来看，本文算法相比YOLOv7–tiny对无人机图像中多尺度小目标辨识能力更强，有效抑制了背景噪声干扰，目标漏检和误检情况有较大改善。进一步说明本文改进算法对于无人机航拍图像目标检测的有效性。

4　结　论

为了解决无人机航拍图像目标检测难题，本文提出一种YOLOv7–tiny改进模型，对其网络结构进行了关键改进，具体改进如下：替换激活函数为SiLU，提升模型的泛化能力；为了增强微小目标处理能力，增加了额外的检测头以及在预测头上集成使用ConvMixer模块；为了减少误检率与漏检率，使用更高效的解耦头增强在复杂背景下对多尺度目标的分类与定位能力。以上改进措施能够相互促进、有机融合。经过在Visdrone2021数据集上的大量实验证明，本文改进算法在不同航拍场景具有较好的检测效果，与先进算法相比在4个类别(pedestrian、people、car、motor)取得了最优的检测效果。但算法还存在一些问题，如存在对自行车和摩托车相似目标类别错误识别问题和极小目标的少量漏检情况。进一步提升极小目标的检测精度和对模型轻量化将是未来的研究方向。

参考文献：

[1] Luo Xudong,Wu Yiquan,Chen Jinlin.Research progress on deep learning methods for object detection and semantic segmentation in UAV aerial images[J].Acta Aeronautica et Astronautica Sinica,2024,45(6):241–270.[罗旭东,吴一全,陈金林.无人机航拍影像目标检测与语义分割的深度学习方法研究进展[J].航空学报,2024,45(6):241–270.]

[2] Tan Li,Huang Xiaokai,Lv Xinyue,et al.Strong interference UAV motion target tracking based on target consistency algorithm[J].Electronics,2023,12(8):1773.

[3] Wu Lizhen,Li Hongnan,Niu Yifeng.Occlusion and confusion targets recognition method for UAV under small sample conditions[J].Journal of National University of Defense Technology,2022,44(4):13–21.[吴立珍,李宏男,牛轶峰.无人机小样本条件下遮挡和混淆目标识别方法[J].国防科技大学学报,2022,44(4):13–21.]

[4] Bao Wenxia,Xie Wenjie,Hu Gensheng,et al.Wheat ear counting method in UAV images based on TPH–YOLO[J].Transactions of the Chinese Society of Agricultural Engineering,2023,39(1):155–161.[鲍文霞,谢文杰,胡根生,等.基于TPH–YOLO的无人机图像麦穗计数方法[J].农业工程学报,2023,39(1):155–161.]

[5] Girshick R.Fast R–CNN[C]//Proceedings of the IEEE International Conference on Computer Vision.New York:IEEE,2015:1440–1448.

[6] Ren Shaoqing,He Kaiming,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.

[7] Cai Zhaowei, Vasconcelos N. Cascade R-CNN: Delving into high quality object detection[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.Salt Lake City:IEEE,2018:6154–6162.

[8] Redmon J,Divvala S,Girshick R,et al. You only look once: Unified, real-time object detection[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR).Las Vegas:IEEE,2016:779–788.

[9] Liu Wei,Anguelov D,Erhan D,et al.SSD:Single shot multibox detector[C]//European Conference on Computer Vision.Cham:Springer,2016:21–37.

[10] Liu Chen, Yang Degang, Tang Liu, et al. A lightweight object detector based on spatial-coordinate self-attention for UAV aerial images[J].Remote Sensing,2023,15(1):83.

[11] Luo Xudong,Wu Yiquan,Wang Feiyue. Target detection method of UAV aerial imagery based on improved YOLOv5 [J].Remote Sensing,2022,14(19):5063.

[12] Jiang Bo,Qu Ruokun,Li Yandong,et al. VC−YOLO:Towards real-time object detection in aerial images[J]. Journal of Circuits,Systems and Computers,2022,31(8):2250147.

[13] Huang Tongyuan,Zhu Jinjiang,Liu Yao,et al.UAV aerial image target detection based on BLUR−YOLO[J].Remote Sensing Letters,2023,14(2):186–196.

[14] Mao Guotao,Deng Tianmin,Yu Nanjing.Object detection in UAV images based on multi-scale split attention[J]. Acta Aeronautica et Astronautica Sinica, 2023, 44(5): 268 – 278. [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[J].航空学报,2023,44(5):268–278.]

[15] Xu Jian, Xie Zhengguang, Li Hongjun. Feature-balanced UAV aerial image target detection algorithm[J]. Computer Engineering and Applications,2023,59(6):196–203.[徐坚,谢正光,李洪均.特征平衡的无人机航拍图像目标检测算法[J].计算机工程与应用,2023,59(6):196–203.]

[16] Xu Guangda, Mao Guojun. Aerial image object detection of UAV based on multi-level feature fusion[J]. Journal of Frontiers of Computer Science & Technology, 2023, 17(3): 635–645.[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[J].计算机科学与探索,2023,17(3):635–645.]

[17] Chen Weibiao, Jia Xiaojun, Zhu Xiangbin, et al. Target detection for UAV image based on DSM−YOLOv5[J].Computer Engineering and Applications,2023,59(18):226–233.[陈卫彪,贾小军,朱响斌,等.基于DSM−YOLOv5的无人机航拍图像目标检测[J].计算机工程与应用,2023,59(18):226–233.]

[18] Du Dawei,Wen Longyin,Zhu Pengfei,et al. VisDrone-DET 2020: The vision meets drone object detection in image challenge results[C]//Proceedings of the 2020 16th European Conference on Computer Vision.Glasgow:Springer,2020: 692–712.

[19] Baidya R, Jeong H. YOLOv5 with ConvMixer prediction heads for precise object detection in drone imagery[J].Sensors,2022,22(21):8424.

[20] Zhu Xingkui,Lyu Shuchang,Wang Xu,et al.TPH−YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW).Montreal:IEEE, 2021:2778–2788.

[21] Trockman A,Kolter J Z.Patches are all you need?Leakey−ReLU[EB/OL].(2022−01−24)[2023−08−01].https://arxiv.org/abs/2201.09792.

[22] Alexey D, Lucas B, Alexander K, et al. An image is worth 16×16 words: Transformers for image recognition at scale [EB/OL].(2021–06–03)[2023–08–01].https://arxiv.org/abs/2010.11929.

[23] Wu Yue, Chen Yinpeng, Yuan Lu, et al. Rethinking classification and localization for object detection[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020: 10186–10195.

[24] Tian Zhi,Shen Chunhua,Chen Hao,et al.FCOS:Fully convolutional one-stage object detection[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision(ICCV).Seoul:IEEE,2019:9627–9636.

[25] Ge Zheng,Liu Songtao,Wang Feng,et al.YOLOX:Exceeding YOLO series in 2021[EB/OL].(2021–08–06)[2023–08–01].https://arxiv.org/abs/2107.08430.

[26] Cao Yaru,He Zhijian,Wang Lujia,et al.VisDrone−DET2021: The vision meets drone object detection challenge results [C]//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Montreal:IEEE,2021:2847–2854.

[27] Yu Weiping,Yang T,Chen Chen.Towards resolving the challenge of long-tail distribution in UAV images for object detection[C]//Proceedings of the 2021 IEEE Winter Confe-

rence on Applications of Computer Vision(WACV). Waikoloa:IEEE,2021:3257–3266.

[28] Zhao Haipeng, Zhou Yang, Zhang Long, et al. Mixed YOLOv3−LITE: A lightweight real-time object detection method[J]. Sensors,2020,20(7):1861.

[29] Wang Hengtao, Zhang Shang, Chen Xiang, et al. Target detection algorithm of lightweight UAV aerial photography [J]. Electronic Measurement Technology,2022,45(19):167–174.[王恒涛,张上,陈想,等.轻量化无人机航拍目标检测算法[J].电子测量技术,2022,45(19):167–174.]

[30] Albaba B M, Ozer S. SyNet: An ensemble network for object detection in UAV images[C]//Proceedings of the 2020 25th International Conference on Pattern Recognition(ICPR). Milan:IEEE,2021:10227–10234.

Small Target Detection Algorithm for UAV Aerial Images Based on Improved YOLOv7−tiny

ZHANG Guanghua[1], *LI Congfa*[1*], *LI Gangying*[1], *LU Weidang*[2]

(1.School of Electrical and Information Engineering, Northeast Petroleum University, Daqing 163318, China;
2.College of Information Engineering, Zhejiang University of Technology, Hangzhou 310014, China)

Abstract:

Objective UAVs provide advantages such as easy control, low cost, and good performance, and efficiently perform tasks in diverse sites and complex environments. UAV aerial image target detection is widely applied in practical scenarios, including urban transportation, military reconnaissance, and smart agriculture. This study proposes a small target detection algorithm for UAV aerial images using a ConvMixer detection head based on the improved YOLOv7−tiny to address the problems of missed detection and false detection caused by significant variations in target scale, densely distributed small-sized targets and complex backgrounds in UAV aerial images.

Methods First, the activation function LeakyReLU is replaced with SiLU to compensate for the limited nonlinear expression of LeakyReLU and to enhance convergence speed and model generalization during training. Second, to strengthen the feature extraction capability for multi-scale targets and improve the detection of small targets, a small-target detection layer is designed, leading to a tiny-target detection head that increases the model receptive field and better addresses the scale variance problem caused by drastic target size changes. In addition, the ConvMixer layer is integrated into the prediction head; the depthwise and pointwise convolutions in ConvMixer capture the spatial and channel relationships in the feature information, improving the processing capability for small targets. Finally, the coupled detection head of YOLOv7−tiny is replaced with a more efficient decoupled head, which separates feature channels for localization and classification tasks and enhances both classification and localization accuracy. Regarding experiments, ablation experiments are designed from two directions to comprehensively verify the effectiveness of each improvement. Comparative experiments are also conducted to assess and analyze the detection performance of the improved algorithm against other algorithms.

Results and Discussions This study mainly addresses the following aspects: 1) The network structure of the improved algorithm is proposed, and the principles and components of each improvement are introduced. Based on the YOLOv7−tiny network, the LeakyReLU activation function in the convolution block CBL is replaced by the SiLU activation function. A small target detection layer is introduced at the neck of the network, and a prediction head is incorporated. Several ConvMixer layers are also integrated into the end of the backbone network and the detection head. Finally, the efficient decoupled head structure is adopted for target prediction. All these enhancements to the baseline form the improved YOLOv7−tiny algorithm network structure. 2) Ablation experiments are designed to verify the effectiveness of each modification. This includes, firstly, adding a single improved module to the original YOLOv7−tiny algorithm to observe its impact and, secondly, removing individual modules from the final improved YOLOv7−tiny−SFCE model to evaluate their effect. Ten sets of ablation experiments are conducted under identical conditions. Results indicate that introducing the efficient decoupled head leads to the most significant accuracy improvement, increasing mAP by 1.1%. Removing the fourth small target detection head results in the most obvious performance degradation, reducing detection accuracy by 2.4%. 3) Comparative experiments are conducted to verify the comprehensive performance of the improved algorithm. More than ten recently proposed advanced algorithms are selected for comparison in terms of AP and mAP values across ten target categories. Results show that the proposed algorithm achieves the highest mAP value of 40.9% and performs best in detecting the categories pedestrian, people, car, and motor. Among these, the pedestrian, people, and motor categories show especially strong detection performance. 4) The detection performance of the improved algorithm is verified in real-world scenarios through comparative analysis. Detection results are demonstrated in various conditions, including sparse and

dense distributions and day and night scenarios. Five images featuring dense targets, minimal targets, dark scenes, occluded targets, and complex backgrounds are randomly selected from the Visdrone2021 test challenge set to evaluate detection performance in UAV aerial images. Comparative visual detection results with the baseline YOLOv7−tiny show that the proposed algorithm significantly improves the identification of multi-scale small targets and reduces both missed and false detections.

Conclusions This study mainly addresses and improves the issues of missed and false detections caused by large-scale variations, dense small target distributions, and complex backgrounds in UAV aerial images. Key contributions include enhancing the model's feature extraction capabilities, providing more accurate localization and classification, and improving small target detection. However, limitations remain: 1) Some missed detections still occur for small targets with minimal pixel information and insufficient features to distinguish them from the background. 2) A balance between detection accuracy and real-time performance has not yet been achieved. The model's parameter count and computational complexity require reduction. Future research will focus on further improving the detection of very small targets and optimizing the model for lightweight applications.

Key words: UAV aerial images; small target detection; SiLU; ConvMixer; efficient decoupled head

（编辑　吴芝明）

引用格式:Zhang Guanghua,Li Congfa,Li Gangying,et al.Small target detection algorithm for UAV aerial images based on improved YOLOv7−tiny[J].Advanced Engineering Sciences,2025,57(3):235–246.[张光华,李聪发,李钢硬,等.基于改进YOLOv7–tiny的无人机航拍图像小目标检测算法[J].工程科学与技术,2025,57(3):235–246.]