计算机工程与应用
Computer Engineering and Applications
ISSN 1002-8331,CN 11-2127/TP

《计算机工程与应用》网络首发论文

题目： 改进 RT-DETR 的航拍小目标检测算法
作者： 刘思元，高凯，雍龙泉
网络首发日期： 2024-10-30
引用格式： 刘思元，高凯，雍龙泉．改进 RT-DETR 的航拍小目标检测算法[J/OL]．计算机工程与应用. https://link.cnki.net/urlid/11.2127.tp.20241029.1605.002

网络首发时间：2024-10-30 09:06:36
网络首发地址：https://link.cnki.net/urlid/11.2127.tp.20241029.1605.002

改进 RT-DETR 的航拍小目标检测算法

刘思元， 高凯， 雍龙泉

陕西理工大学 数学与计算机科学学院，陕西 汉中 723001

摘 要：针对现有的目标检测算法在航拍图像中小目标中易出现的漏检和误检问题，提出了基于改进 RT-DETR 的算法。首先在主干网络中引入了部分卷积 PConv，设计了 PConvBlock 结构，并通过由 PConvBlock 组成的 BasicBlock-PConvBlock 模块替代原有 BasicBlock，有效减少了模型参数。其次，采用双向特征金字塔网络 BiFPN 结构优化特征融合模块，并引入 S2 特征进一步提升小目标的检测能力。最后，引入 CARAFE 上采样算子，增强了多尺度特征的快速融合。实验表明，在 VisDrone 测试集上，改进后的模型在参数量上比 RT-DETR 模型降低了 13.9%，同时在 mAP0.5 和 mAP0.5:0.9 指标上分别提升了 2.4%和 1.9%。在 TT100K 和 DOTA 数据集上均优于 RT-DETR 算法。改进模型在保持较小参数量和计算量的同时，提高了检测精度，满足了无人机航拍图像实时检测的应用需求。

关键词：小目标检测；轻量化；RT-DETR；部分卷积

文献标志码：A **中图分类号：**TP391.41 doi：10.3778/j.issn.1002-8331.2407-0399

Improved RT-DETR Algorithm for Aerial Small Object Detection

LIU Siyuan, GAO Kai, YONG Longquan

School of Mathematics and Computer Science, Shaanxi University of Technology, Hanzhong, Shaanxi 723001, China

Abstract: Aiming to address the issue of missed and false detection of small objects in aerial photography images by existing object detection algorithms, an improved algorithm based on RT-DETR was proposed. Firstly, Partial Convolution (PConv) was introduced into the backbone network, and a PConvBlock structure was designed. Then, a BasicBlock-PConvBlock module composed of PConvBlocks replaced the original BasicBlock, effectively reducing the number of model parameters. Secondly, the Bidirectional Feature Pyramid Network (BiFPN) structure was adopted to optimize the feature fusion module. Furthermore, the S2 feature was introduced to enhance the detection ability of small objects. Finally, the CARAFE upsampling operator was introduced to strengthen the fast fusion of multi-scale features. Experimental results show that the improved model has a 13.9% reduction in parameter number compared to the RT-DETR model, and the mAP0.5 and mAP0.5:0.9 indicators are improved by 2.4% and 1.9%, respectively, on the VisDrone test set. On the TT100K and DOTA datasets, the improved model outperforms the RT-DETR algorithm. The improved model significantly enhances detection accuracy while maintaining a smaller parameter number and computational cost, meeting the real-time detection application requirements for drone aerial photography images.

Key words: Small Object Detection; Lightweight; RT-DETR; Partial Convolution

随着无人机技术的快速发展，其在航拍、监控、 和应急救援等多个领域的应用正变得越来越广泛。在

基金项目：陕西省自然科学基础研究计划项目(2024JC-YBMS-014);陕西省教育厅青年创新团队项目(23JP024);陕西理工大学科研项目(SLGNL202409)。

作者简介：刘思元(2001—)，男，硕士研究生，CCF 学生会员，研究方向为图像处理、目标检测；高凯(1981—)，通信作者，男，副教授，硕士生导师，主要研究方向为计算机应用技术、智能优化算法，E-mail: gaokai@snut.edu.cn; 雍龙泉(1980—)，男，教授，博士，CCF 普通会员，主要研究方向为最优化理论与算法，智能优化算法。

这些应用中，目标检测是一项至关重要的任务，它可以帮助无人机快速准确地识别图像中的目标物体，从而实现各种功能，如监视区域安全、追踪移动目标、执行搜索与救援等任务。然而，在实际应用中，无人机航拍的图像常常具有一定的挑战性，例如拍摄高度、光照条件、目标尺寸变化等因素会影响目标检测的性能，容易受环境干扰，导致难以被常规目标检测算法检测出来[1]。尤其是针对小目标的检测任务，由于小目标通常具有低分辨率、低对比度和复杂背景等特点，更具有挑战性。

随着深度学习在目标检测领域的广泛应用，目标检测技术的准确性和速度迅速提高。最初的 R-CNN 系列算法，如 R-CNN[2]、Fast R-CNN[3]、Faster R-CNN[4]、Mask R-CNN[5]等算法，这些算法都是先生成候选框，然后识别框内物体的两阶段算法，优先考虑检测精度，但是由于计算量过大导致运行速度慢。随后出现了代表单阶段算法的 SSD[6]算法和 YOLO 系列算法[7-12]，这些算法在精度和速度之间取得了良好的平衡，获得了广泛的应用。然而，无论是两阶段还是单阶段算法，大多数都依赖非极大值抑制 NMS 技术来处理冗余的边界框，这种技术存在效率低、无法并行处理，降低了推理速度，在不同的场景下要选择合适的 NMS 阈值避免丢失目标的问题。

Transformer[13]模型最初在自然语言处理领域表现优异，近年来被逐渐应用于计算机视觉领域。DETR[14]首次将 Transformer 应用于计算机视觉领域，直接对图像进行集合预测，取消了 anchor 机制和 NMS 处理，简化了目标检测流程，在目标检测领域表现出了卓越的学习能力。然而，DETR 在训练时需要大量时间，限制了其在实时应用中的使用。许多研究人员对 DETR 进行了改进，但注意力机制本身引入了更多的参数和计算复杂性，阻碍了 DETR 模型的实时应用。LV[15]等人提出了实时检测模型 RT-DETR，这是一款高实用性的端到端实时目标检测模型，保持高精度的同时实现了实时性能，在速度和精度方面优于同等规模的 YOLO 系列检测模型。

为了提高小目标检测的精度，SUN[16]等人将经典路径聚合网络 PAN[17]与聚合和分发 GD(Gather-and-Distribute)[18]机制集成在一起，组成新的特征融合架构应用在 RT-DETR 中，利用两种融合方法的优势，提高了对小目标的检测。MUZAMMUL[19]等人将 RT-DETR 模型与切片辅助超推理 SAHI(Slicing Aided Hyper Inference)[20]方法结合，有效提升了无人机航拍图像中的目标检测和识别能力。Li[21]等人在 RT-DETR 中引入自注意力上采样 SAU[22]模块，并且提出了基于 EIoU(Efficient-IoU)改进的 IDIoU 损失函数，以减少小缺陷匹配过程中的不稳定性，提升小目标的检测能力。张储[23]等人使用 Inner-IoU 损失函数和 OrthoBasicBlock 模块优化 RT-DETR 模型，OrthoBasicBlock 模块使用通过 Gram-Schmidt 过程正交化的滤波器集合，提供更好的注意力模块初始化状态，有利于收敛到更优解。李亦涵[24]等人将 RT-DETR 模型引入可变核卷积 AKConv[25]来适应遥感图像中小目标的大小和形状变化，并且引入 Mosaic 数据增强技术对数据集进行预处理，增强了模型对复杂背景和光照变化的适应性。庞玉东[26]等人在 RT-DETR 模型使用部分卷积 PConv(Partial Convolution)[27]设计 FastNet-Block，替换原始 BackBone 中的 BasicBlock，以提升模型性能。胡佳乐[28]等人在 RT-DETR 中设计引入多尺度注意力模块 EMA(Efficient Multi-scale Attention)[29]对 BasicBlock 模块进行改进，并且结合动态上采样模块 Dysample[30]和 SSFF[31]模块提出 DySSFF 模块,替换原有的特征融合模块，避免小目标特征信息丢失。

综上所述，本文在 RT-DETR-R18 模型上进行改进，实验表明，改进算法在 VisDrone2019、TT100K 和 DOTA 数据集上比 RT-DETR-R18 模型均表现出色。具体的改进工作体现在以下三个方面：

(1)在主干网络，使用部分卷积 PConv 构建出 PConvBlock 结构，将原始模型中的 BasicBlock 模块替换为由 PConvBlock 组成的 BasicBlock- PConvBlock，保持了高效的特征提取能力，有效地减少了模型的参数量。

(2)在特征融合模块，将传统的 FPN 替换为双向特征金字塔网络 Bi-FPN[32]，并且引入额外的高分辨率特征图 S2，与 S3 特征进行融合。改进后的融合结构在小目标检测任务中表现出了更高的精度，为小目标检测提供了更丰富的特征支持，从而提升了检测性能。

(3)使用 CARAFE[33]轻量级上采样替换最近邻插值方法，解决语义信息丢失问题，提升特征金字塔网

络性能。

1 RT-DETR 概述

RT-DETR 是基于 Transformer 的实时端到端目标检测器，主要包含骨干网络(Backbone)、高效混合编码器(Efficient Hybrid Encoder)和带有检测头的解码器(Transformer Decoder)三部分。

其中高效混合编码器由两个主要模块组成：基于注意力的同尺度特征交互(Attention-based Intra-scale Feature Interaction, AIFI)和基于 CNN 的跨尺度特征融合(CNN-based Cross-scale Feature Fusion, CCFF)。这两个模块协同工作，从骨干网络的最后三个阶段(S3、S4、S5)中提取并处理多尺度特征。AIFI 通过使用单尺度 Transformer 编码器和骨干网络的最高特征 S5 上执行同尺度特征交互，降低了计算成本，同时保持对高层语义信息的敏感性。CCFF 由多个融合块(Fusion Block)组成，这些融合块通过卷积层将相邻尺度的特征融合到一起形成新的特征。其中每个融合块包含两个 1×1 的卷积层和三个 RepConv，1×1 的卷积层用于调整通道数，RepConv 用于特征融合。最后通过逐元素相加的方式融合两个路径的输出，生成融合后的新特征。

解码器将编码器的输出转换为最终的目标检测结果，首先通过一个 IOU 感知查询模块来实现，该模块负责从编码器输出的特征序列中选择一组最具代表性的图像特征。这些特征随后被用作初始对象查询，是解码器后续生成预测结果的基础。为了提高查询的初始质量和预测的准确性，RT-DETR 提出了不确定性最小查询选择机制，该机制通过量化预测的类别和定位之间的不确定性，选择那些具有最低不确定性的查询作为初始对象查询。一旦选择了初始对象查询，解码器将通过一系列的迭代优化步骤来精细预测结果，经过若干次迭代优化后，解码器最终生成一组预测框和相应的置信度分数。

本文选择 RT-DETR-R18 作为基线模型，模型结构如图 1 所示。

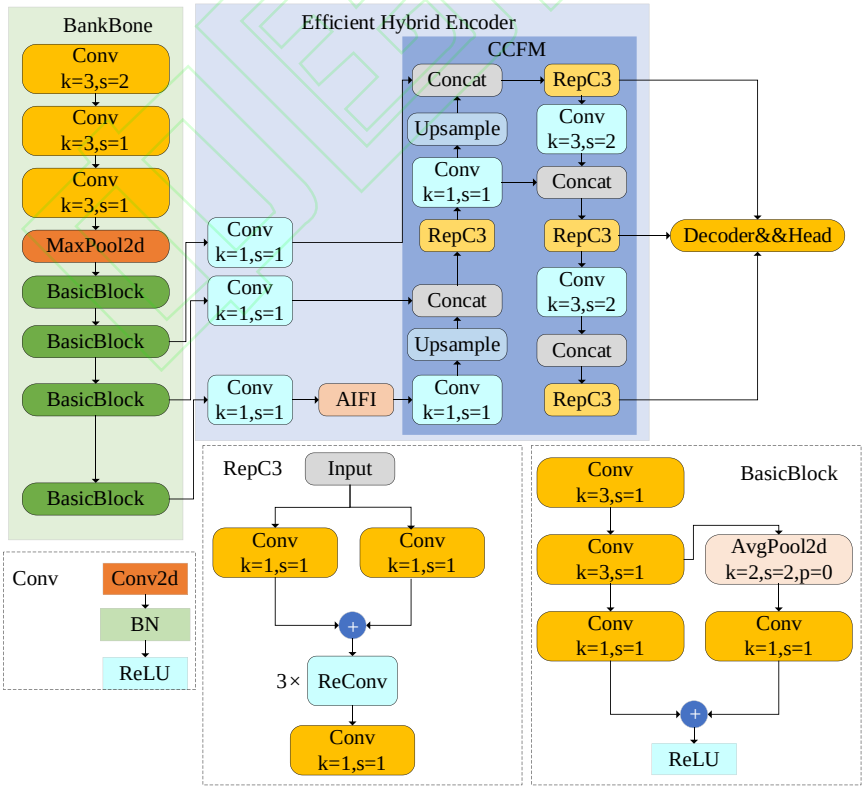


图 1 RT-DETR-R18 模型结构

Fig.1 RT-DETR-R18 model structure

2 RT-DETR 改进

本文对 RT-DETR-R18 网络进行了一系列的改进，旨在提升小目标检测算法的精度，同时降低模型的参数量。首先在主干网络中引用了 PConv 构建了 PConvBlock 结构，随后采用由 PConvBlock 构成的 BasicBlock-PConvBlock 替代了原有的 BasicBlock 模块，这一改进减少了模型的参数量，而不影响其高效的特征提取能力。进一步地，在特征融合模块采用双向特征金字塔网络 BiFPN 结构，并通过引入 S2 特征进一步对其进行了优化，这种改进增强了网络对小目标的检测精度，使改进后的模型能够捕捉到图像中的细微特征。最后，引入 CAREFE 上采样算子，能够快速且有效地进行多尺度特征融合，进一步提升了模型的性能。改进后的 RT-DETR 结构如图 2 所示。

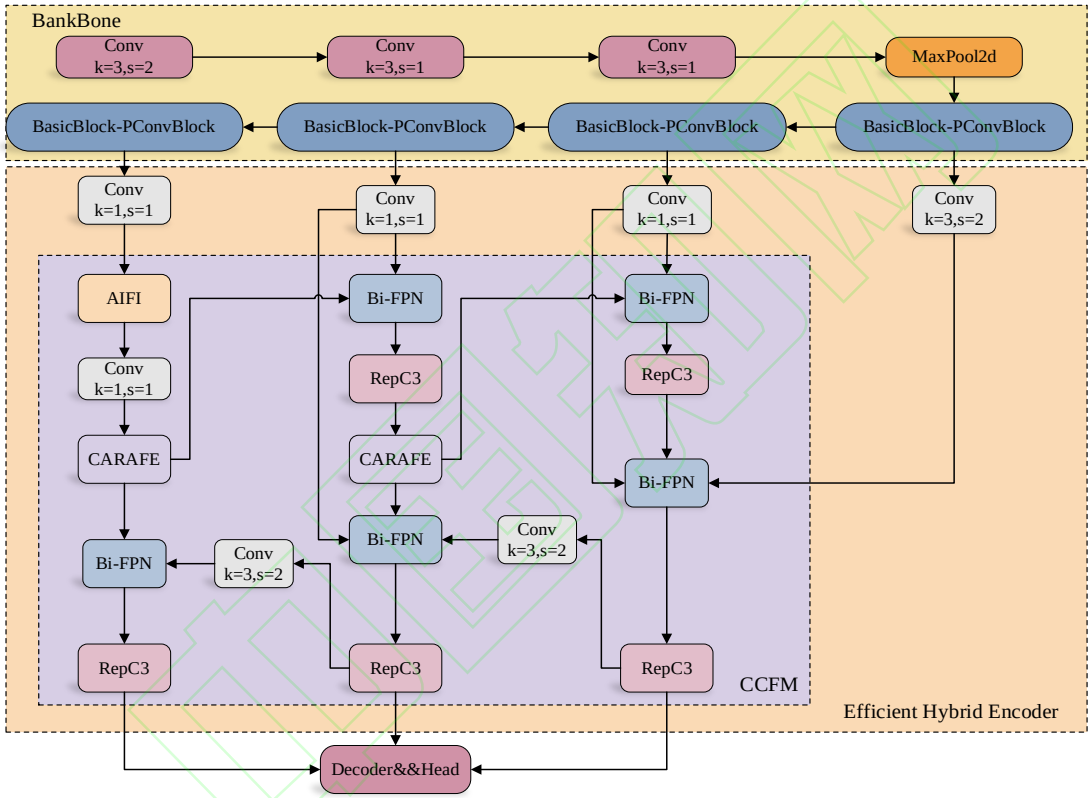


图 2　改进的 RT-DETR-R18 模型结构

Fig.2　Structure of improved RT-DETR-R18 network

2.1　结合 PConv 的 BasicBlock-PConvBlock

为解决主干网络中普通卷积带来的参数膨胀问题，同时确保网络能高效捕捉空间特征，本文引入了 FasterNet 提出的 PConv 部分卷积技术。PConv 的核心优势在于其轻量化设计：仅对输入特征图的一部分通道应用卷积运算，而不是全部通道，从而减少了整体的计算量。对于连续或规则排列的数据，可以通过选取特征图中的第一个或最后一个连续通道，将其作为整个特征图的代表性样本进行计算。这种方法减少了不必要的计算和内存访问，优化了资源的使用效率。PConv 的 FLOPs 浮点运算量仅为传统卷积的 1/16，内存访问量也降低到传统卷积的 1/4。这种计算上的优化，不仅减轻了模型的负担，还提高了运算效率。

为了增强网络捕捉目标关键特征的能力，并有效控制过拟合，对 BasicBlock 进行改进，如图 3 所示。此设计中，在 PConv 之前融入了 1×1 卷积层以及跳跃连接，其中 1×1 卷积的作用在于调整通道数，确保输入输出的维度匹配，保持网络的一致性。通过引入残差连接，网络能够在训练过程中整合来自多分支的丰富特征信息，提升了网络对特征的提取和表达能力。为了进一步优化训练过程，在残差结构前增设了 DropPath 机制，这一策略缓解了过拟合的现象，确保网络在面对复杂数据时的泛化能力。

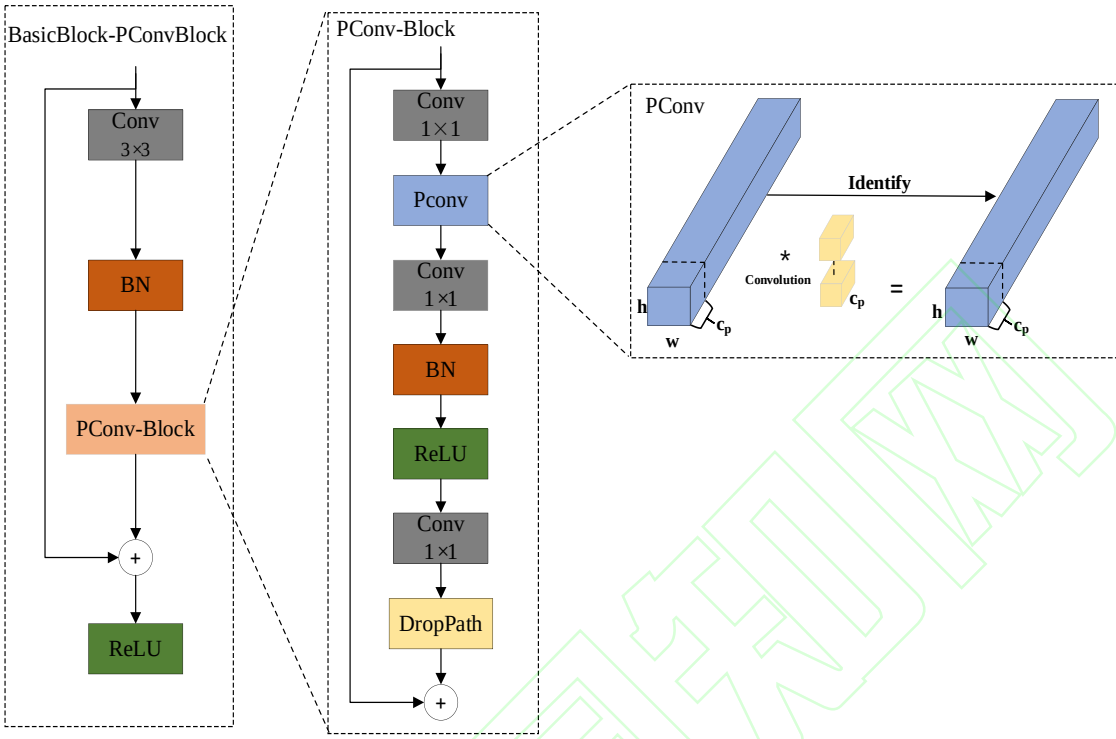


图 3 改进的 BasicBlock 网络结构

Fig.3 Improved BasicBlock network structure

2.2 改进 BiFPN 的多尺度特征融合

RT-DETR 模型通过融合骨干网络最后三个阶段的特征图，提供了一种适用于目标检测任务的通用解决方案。然而，在处理航拍图像这类特殊场景时，由于小目标在图像中占据的像素较少，它们的特征信息在早期的特征提取阶段容易被忽略，导致关键目标信息的丢失。

为了克服这一局限并提升模型在航拍图像中小目标检测的性能，改进后的结构如图 4 所示。将传统的 FPN 结构替换为 BiFPN，BiFPN 通过其双向连接机制，实现了自顶向下的高层语义信息传递和自底向上的低层精细空间信息的上采样，从而丰富了特征的上下文感知。此外，BiFPN 的交叉连接层加强了同层和跨层特征的融合，提升了模型对多尺度目标的检测精度。这一改进不仅增强了特征表达，也优化了小目标的识别效果。

在 BiFPN 的基础上，首先引入了特征图 S2，由于特征图 S2 采样尺度产生的 160 × 160 像素的特征图，其精细的分辨率捕捉到了更为丰富的细节信息，对于小目标的特征提取至关重要。其次通过将 S2 通道提取的高分辨率特征与 S3 通道的特征进行融合，这种跨尺度的特征整合不仅增强了模型对小目标的感知，还提高了特征的表达能力。实验结果如表 1 所示，这种改进提升了网络对小目标的捕捉能力，改进后的 BiFPN 在精度上比未改进的 BiFPN 有所提高。

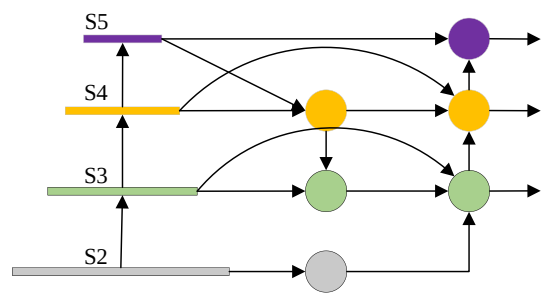


图 4 改进的 BiFPN 结构

Fig.4 Improved BiFPN structure

表 1 在 VisDrone 测试集上 S2-BiFPN 和原始 BiFPN 的对比

Table 1 Comparison between S2-BiFPN and the original BiFPN on the VisDrone2019-DET-test-dev

RT-DETR-R18	BiFPN	S2-BiFPN	Precision/%	Recall/%	mAP0.5/%	mAP0.5:0.9/%	Params/MB
√			56.4	39.9	38.8	22.4	20.1
	√		56.2	39.2	39.1	22.5	19.6
		√	**56.7**	**41.4**	**40.1**	**23.5**	20.6

2.3 上采样算子 CARAFE

在 RT-DETR 框架中,上采样过程默认采用的最近邻插值方法，虽然计算简洁高效，但它主要关注局部亚像素区域的直接邻近像素，没有充分考虑特征图的全局内容信息。这种方法无法充分捕捉密集检测任务，特别是在处理航拍图像时，由于图像中的目标分布密集并且特征细节微小，仅依赖最近邻插值导致在上采样过程中损失关键的空间信息，增加了漏检的风险。

为了解决这一挑战,引入 CARAFE 作为上采样策略。CARAFE 通过内容感知的方式进行特征重组，不仅考虑了局部区域的像素，还能够整合更广泛的上下文信息，从而生成更加精确和语义丰富的特征表示。这种上采样机制能够减少航拍图像中由于特征微小导致的空间信息损失，有效降低漏检情况，提高检测的准确性和完整性。CARAFE 整体框架如图 5 所示。

在 CARAFE 中,核预测模块负责生成内容感知的重组内核，这一模块由三个子模块构成，通道压缩器接收一个尺寸为 H × W × C 的特征图，通过 1 × 1 卷积层将输入特征图的通道数从 C 压缩到 C_m，从而降低后续步骤的参数量和计算成本。其次内容编码器使用卷积核大小为$k_{up} \times k_{up}$进行内容编码，生成大小为 H × W × σ^2 × k_{up}^2重组上采样核，其中 σ 是上采样的比例因子，随后这些核在空间维度上进行展开，形成 σH × σW × k_{up}^2大小的上采样核。最终核归一化器通过 softmax 函数对每个重组内核进行空间归一化，确保所有上采样核中的权重值之和为 1。

内容感知重组模块利用核预测模块生成的上采样核进行特征重组。首先对于输出特征图的每一个目标位置使用一个$k_{up} \times k_{up}$大小的局部窗口，该窗口的源像素来计算目标像素的值。对局部窗口内的每个像素，根据其对应的上采样中的权重进行加权，然后将这些加权的像素值求和，得到目标位置的最终值。计算公式如下：

$$\boldsymbol{X}_{l'}^{'} = \sum_{n=-\frac{k}{2}}^{\frac{k}{2}} \sum_{m=-\frac{k}{2}}^{\frac{k}{2}} \boldsymbol{W}_{l'}(n,m) \cdot \boldsymbol{X}(i+n, j+m) \quad (1)$$

在上式中，$\boldsymbol{X}_{l'}^{'}$ 是上采样后的特征图在位置l' 的值，$\boldsymbol{W}_{l'}(n,m)$是上采样核在位置 n,m 的权重，$\boldsymbol{X}(i+n, j+m)$是输入特征图在局部区域中的原始像素值。

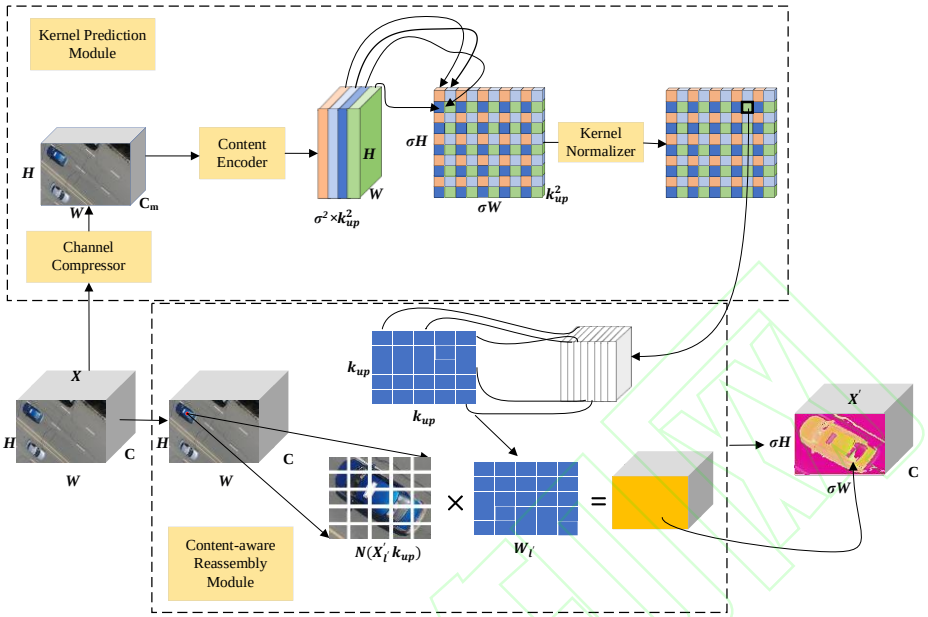


图 5 CARAFE 整体框架

Fig.5 CARAFE overall framework

3 实验与结果分析

3.1 实验环境

本文所有的实验环境均在 NVIDIA A800 80GB GPU 上进行，使用 Python 3.9 编译器和 Pytorch 1.12.1 深度学习框架。为了公平比较，实验中的模型均未使用预训练权重，批量大小 batch size 设置为 16，训练轮次 epoch 设置为 300，早停轮数设置为 50，初始学习率为 0.0001，其他训练超参数均采用默认值。

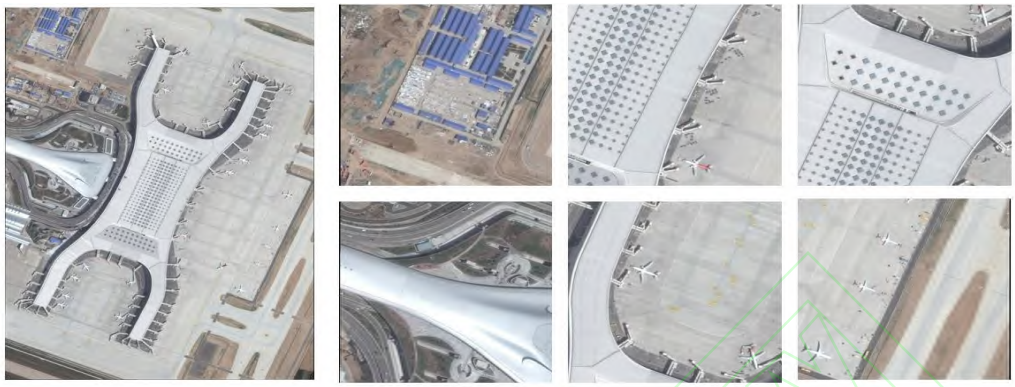
3.2 数据集介绍

本文在 VisDrone2019[34]、TT100K[35]和 DOTA[36]数据集上进行小目标检测实验。VisDrone2019 是一个专门为无人机航拍设计的小目标数据集，航拍图像包含复杂多变的背景和大量的小目标。TT100K 是一个用于交通标志检测和识别的大规模数据集，图像包含多类别、高分辨率图像和多样化场景而具有很高的挑战性。DOTA 数据集是一个专门用于遥感图像中目标检测的大规模数据集，其图像目标密集，目标实例的尺寸变化较大。VisDrone2019 作为本文实验的主要实验数据集，进行消融实验和对比实验。在 TT100K 和 DOTA 数据集上进行泛化实验，验证模型的普适性。

VisDrone2019 数据集由天津大学机器学习和数据挖掘实验室 AISKYEYE 团队收集，包含了 8629 张无人机航拍图像。其中 6471 张图片为训练集，548 张图片为验证集，1610 张图片为测试集。

TT100K 数据集由清华大学发布的公开数据集，对原数据集使用 Python 语言处理，只保留包含图片数超过 100 的类别，并按照 7:2:1 的比例重新划分数据集，其中训练集有 6598 张，验证集有 1889 张，测试集有 970 张。

DOTA 数据集是遥感图像目标检测的大规模数据集，由于数据集中的图片尺寸大小不一致，直接送入网络中难以训练，将 DOTA 数据集通过切割原始图像的方法，以相邻图像间隔 200 像素将其分割成 1024×1024 像素的子图，使用 Python 语言处理后共得到 21046 张图片，处理后的效果如图 6 所示。其中训练集包含 15749 张图片，测试集 5297 张图片。



(a)原始图片 (b)处理后部分效果

图 6 DOTA 数据集处理效果对比

Fig.6 Comparison of processing effects on DOTA dataset

3.3 评估指标

本文使用精确率 (Precision)、召回率 (Recall)、平均精确度(mAP)和参数量(Parameters)作为评估指标，其中 Precision ,Recall，mAP 计算公式如下：

$$\text{Precision} = \frac{\text{TP}}{\text{FP} + TP} \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{FN} + TP} \tag{3}$$

$$AP = \int_0^1 P(R)dR \tag{4}$$

$$mAP = \frac{1}{n}\sum_{i=0}^{n} AP_i \tag{5}$$

在上式中，TP 为真正例，即正确检测到的目标数；FP 为假正例，即错误检测为目标的数目；FN 为假负例，即未被模型检测到的目标数。AP 用于衡量单个类别的检测精度，是精确率-召回率(P-R)曲线下的面积。mAP 用于衡量模型在所有类别上的综合检测性能，是所有类别的平均精确度的平均值,其中 n 为类别数。

3.4 消融实验

本文基于 RT-DETR 模型进行了一系列的改进，通过消融实验验证所提算法模块的有效性。实验在 VisDrone 测试集上进行，实验结果如表 2 所示。

实验结果表明，使用结合 PConv 的 BasicBlock-PConvBlock 替换了原有的 BasicBlock 进行轻量化处理，减少了模型的参数量，相较于原始的 RT-DETR-R18 的参数量降低了 16.4%，在精度上影响甚微。进一步地，引入结合了 S2 特征图的 BiFPN 结构，模型在 mAP0.5 和 mAP0.5:0.9 指标上分别提升了 1.1%和 0.6%。最后，引入了 CARAFE 上采样算子，进一步提升了模型的性能,相较于原始的 RT-DETR-R18 模型，精度提高了 1.8%，召回率提升了 2.5%，mAP0.5 提升了 2.4%，mAP0.5:0.9 提升了 1.9%，同时模型的参数量也降低了 13.9%，证明了改进措施的有效性。综上所述，这一系列的改进不仅优化了模型的结构，提高了检测精度，同时也降低了模型的参数量。

表 2 在 VisDrone 测试集上的消融实验结果

Table 2 Results of ablation experiments on the VisDrone2019-DET-test-dev

BasicBlock-PConvBlock	S2-BiFPN	CARAFE	Precision/%	Recall/%	mAP0.5/%	mAP0.5:0.9/%	Params/MB
×	×	×	56.4	39.9	38.8	22.4	20.1
√	×	×	56.1	39.9	38.7	22.3	16.8
×	√	×	56.7	41.4	40.1	23.5	20.6
×	×	√	57.1	40.5	39.6	23.1	20.0
√	√	×	56.8	41.2	39.9	23.0	17.5
√	×	√	57.7	41.1	39.8	22.7	16.9
×	√	√	57.1	40.8	40.2	23.6	20.4

√	√	√	**58.2**	**42.4**	**41.2**	**24.3**	17.3

3.5 对比实验

为了验证本文提出的算法在小目标检测上的优越性，在 VisDrone 测试集上将本文算法与当前主流算法进行对比，包括 YOLOv5m、YOLOv5l、YOLO8m、YOLO8l、RT-DETR-R18、RT-DETR-R34 以及 YOLOv9-c，训练集实验结果的对比如图 7 所示。在 VisDrone 验证集上，本文提出的算法与文献[16]、文献[28]、文献[37]和文献[38]的结果进行对比分析。文献[37]通过采用部分卷积 PConv 和高效多尺度注意力 EMA 机制，构建了创新的 F_C2f_EMA 模块。与之相比，本文设计的基于部分卷积 PConv 的 BasicBlock- PConvBlock 模块展现出更优的性能，在特征提取和融合方面的增强，提升了检测精度。文献[38]引入了上采样算子 CARAFE，以减少通道压缩过程中的信息丢失。与该文献中的方法相比，本文将 CARAFE 上采样算子与改进的 S2-BiFPN 相结合，通过双向特征融合增强了模型对小目标的检测能力，对比结果如表 3 所示。综上所述，本文提出的改进算法不仅在 Precision、Recall 和 mAP 这三个指标上高于其他主流算法和其他改进算法，而且在训练过程中的损失值始终维持在较低水平。

测试集实验结果如表 4 所示。改进后 RT-DETR 模型在 VisDrone 测试集上 mAP0.5 能够达到 41.2%，均优于其他模型，其中 mAP0.5 比 YOLOv5m 高出 6.3%、比 YOLOv5l 高出 3.7%、比 YOLO8m 高出 5.9%、比 YOLOv8l 高出 3.4%，在参数量上比 RT-DETR-R34 减少了 44%，同时在 mAP0.5 上还提升了 1.0%。改进后的 RT-DETR 模型相较于其他模型参数量更小，精度更高。

为了评估本文提出的改进算法在不同小目标数据集上的检测效果和泛化能力。在 TT100K 和 DOTA 数据集上进行进一步的对比实验。对比结果如表 5 所示。从表中可以看出，在 TT100K 数据集上，改进后的模型相较于 RT-DETR-R18 在 Precision、Recall、mAP0.5 以及 mAP0.5:0.9 上提高了 2.8%、0.8%、1.7%、2%。在 DOTA 数据集上，改进后的模型比 RT-DETR-R18 在 Precision、Recall、mAP0.5 和 mAP0.5:0.9 上提高了 0.9%、0.5%、0.9%、1.2%。综合两个数据集的测试结果，改进后的算法在小目标检测方面，具有高精度、低漏检率以及更小的模型尺寸。

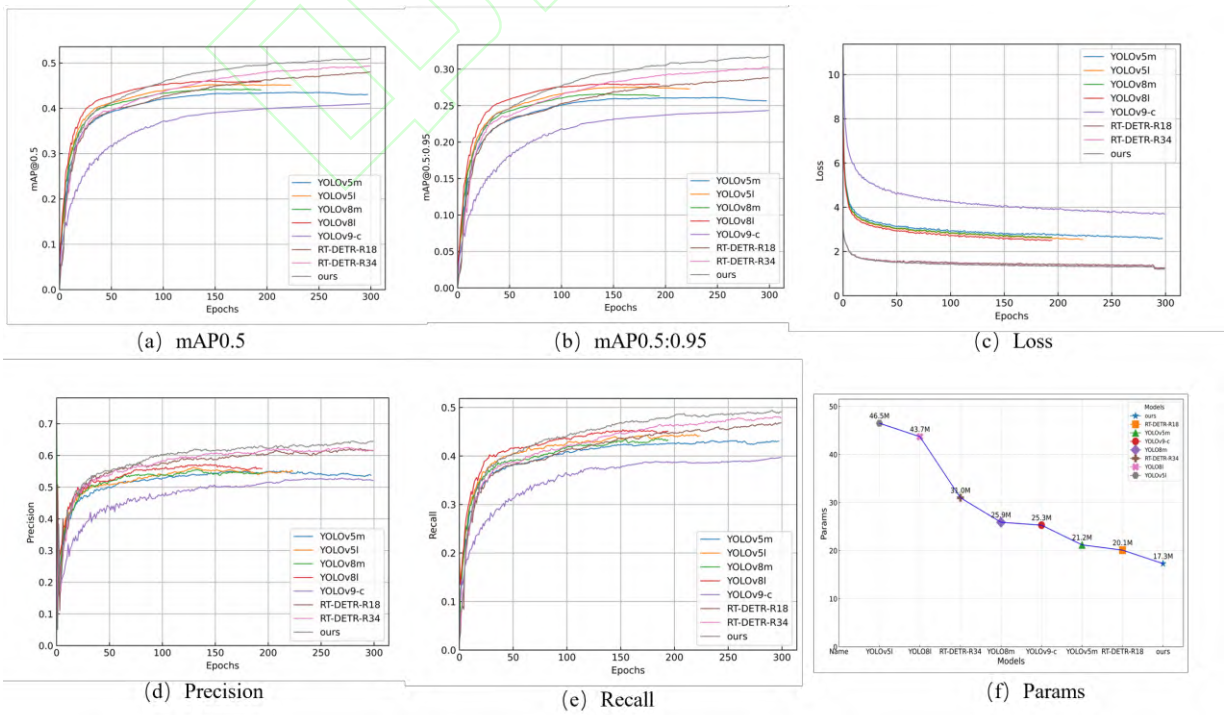


图 7 各主流算法对比

Fig.7 Comparison of various algorithms

表 3 其他改进算法在 VisDrone 验证集上的对比实验

Table 3 Comparative experiments of other improved algorithms on the VisDrone2019-DET-val

Model	Precision/%	Recall/%	mAP0.5/%	mAP0.5:0.9/%	Params/MB
文献[16]	-	-	47.1	28.8	28.1
文献[28]	63.2	47.3	49.4	30.2	33.8
文献[37]	58.5	46.0	48.7	29.7	5.94
文献[38]	-	-	43.1	25.7	12.1
ours	**64.4**	**49.0**	**51.0**	**31.7**	17.3

表 4 其他主流算法在 VisDrone 测试集上的对比实验

Table 4 Comparative experiments of other mainstream algorithms on the VisDrone2019-DET-test-dev

Model	Precision/%	Recall/%	mAP0.5/%	mAP0.5:0.9/%	Params/MB
YOLOv5m	47.8	36.8	34.9	20.1	21.2
YOLO5l	49.8	39.0	37.5	21.9	46.5
YOLO8m	47.9	37.6	35.3	20.3	25.9
YOLO8l	51.8	38.5	37.8	22.1	43.7
RT-DETR-R18	56.4	39.9	38.8	22.4	20.1
RT-DETR-R34	56.1	41.3	40.2	23.5	31.0
YOLOv9-c	47.5	35.4	34.2	19.9	25.3
ours	**58.2**	**42.4**	**41.2**	**24.3**	**17.3**

表 5 在 TT100K 和 DOTA 数据集上的对比实验

Table 5 Comparative experiments on TT100K and DOTA datasets

数据集	Model	Precision/%	Recall/%	mAP0.5/%	mAP0.5:0.9/%	Params/MB
TT100K	RT-DETR-R18	86.3	82.7	85.9	65.3	20.1
	ours	**89.1**	**83.5**	**87.6**	**67.3**	**17.3**
DOTA	RT-DETR-R18	76.7	69.1	71.4	47.3	20.1
	ours	**77.6**	**69.6**	**72.3**	**48.5**	**17.3**

3.6 可视化分析

为了验证本文改进算法在真实复杂场景中的有效性，在 VisDrone2019 测试集上挑选了包含密集、昏暗以及高空视角的代表性图像进行检测，这些图像中的检测目标均为小尺寸目标，检测效果如图 8 所示。通过图 8 的对比检测图可知，在第一幅图像中，由于受到栏杆和植被的遮挡，RT-DETR 算法错误的将路灯识别为汽车，而本文改进算法则避开了这一常见误检。第二幅图像的暗光复杂背景下，RT-DETR 未能检测到行人目标，而本文改进算法却成功地识别并定位了行人。在第三幅图像中，目标间的密集和遮挡问题导致 RT-DETR 漏检了多辆汽车，而本文改进算法依然能够准确地检测出这些小目标。与 RT-DETR 相比，本文的改进算法减少了漏检和误检的发生，提升了模型对小目标的检测精度。

(a)RT-DETR (b)本文改进模型

图 8 改进前后的检测效果对比

Fig.8 Comparison of detection effects before and after improvement

4 结束语

在本文研究中，针对航拍图像中普遍存在的小目标众多、相互遮挡导致漏检和误检的问题，提出了一种改进 RT-DETR 的航拍图像小目标检测算法。本文研究从模型轻量化的角度出发，采取了以下解决方案：首先，在主干网络中引入了 PConv，构建了高效的 PConvBlock 结构。随后，将由 PConvBlock 构成的 BasicBlock-PConvBlock 模块替代了原有的 BasicBlock，有效减少了模型的参数量。其次，在特征融合模块，采用了双向特征金字塔网络 BiFPN 结构，并通过引入 S2 特征进一步优化了这一结构，提升了模型对小目标的检测精度。最后，引入了 CARAFE 上采样算子，以快速且有效地进行多尺度特征融合，增强了模型的特征表达能力。

实验结果相比于 RT-DETR 模型在 VisDrone 测试集上，本文改进模型在参数量上降低了 13.9%，而在 mAP0.5 和 mAP0.5:0.9 指标上分别提升了 2.4%和 1.9%。与其他目标检测算法相比，本文改进模型在检测精度上取得了优势，同时保持了较小的参数量和计算量，满足了无人机航拍图像实时检测的需求。此外，在 TT100K 数据集上，本文改进模型在 mAP0.5 和 mAP0.5:0.9 指标上相较于 RT-DETR-R18 分别提高了 1.7%和 2%。在 DOTA 数据集上，改进后的模型在相同指标上比 RT-DETR-R18 提高了 0.9%和 1.2%。

综上所述，本研究提出的改进算法不仅在小目标检测精度上取得了突破，同时在参数量和计算效率上也实现了优化，充分证明了其在航拍图像小目标检测领域的有效性和实用性。

参考文献：

[1] 谢椿辉,吴金明,徐怀宇.改进 YOLOv5 的无人机影像小目标检测算法[J].计算机工程与应用,2023,59(9):198-206. XIE C H, WU J M, XU H Y.Small object detection algorithm based on improved YOLOv5 in UAV image[J].Computer Engineering and Applications,2023,59(9): 198-206.

[2] GIRSHICK R, DONAHUE J, DARRELL T, et al. Ri-ch feature hierarchies for accurate object detection andsemantic segmentation[C]//Proceedings of the IEEE C-onference on Computer Vision and Pattern Recognitio-n,2014: 580-587.

[3] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IE-EE international conference on computer vision. 2015:1440-1448.

[4] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region propos-al networks[J].IEEE Transactions on Pattern Analysis &Machine Intelligence, 2017, 39(6): 1137-1149.

[5] HE K, GKIOXARI G, DOLLÁR P, et al. Mask r-cnn[C]//Proceedings of the IEEE international conference on computer vision, 2017: 2961-2969.

[6] LIU W, ANGUELOV D, ERHAN D, et al. SSD: sin-gle shot multibox detector[C]//European Conference onComputer Vision, 2016: 21-37.

[7] REDMON J, DIVVALA S, GIRSHICK R, et al. You only once: unified, real-time object detection[C]//Proc-eedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.

[8] LI C, LI L, JIANG H, et al. YOLOv6:A single-stage object detection framework for industrial applications[J].arxiv preprint arxiv:2209.02976, 2022.

[9] GE Z, LIU S, WANG F, et al. Yolox:Exceeding yolo series in 2021[J].arxiv preprint arxiv:2107.08430, 2021.

[10] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. Y-OLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 7464-7475.

[11] WANG C Y, YEH I H, LIAO H Y M. YOLOv9:Le-arning What You Want to Learn Using Programmable Gradient Information[J]. arxiv preprint arxiv:2402.13616, 2024.

[12] WANG A, CHEN H, LIU L, et al. YOLOv10: Real-Time End-to-End Object Detection[J]. arxiv preprint a-rxiv:2405.14458, 2024.

[13] VASWANI A, SHAZEER N, PARMAR N, et al. Att-ention is all you need[C]//Proceedings of the 31st Int-ernational

Conference on Neural Information Processi-ng Systems. Red Hook,NY： Curran Associates Inc.2017: 6000-6010.
[14] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]// Europeanconference on computer vision. Cham: Springer Intern-ational Publishing, 2020: 213-229.
[15] LV W, XU S, ZHAO Y, et al. Detrs beat yolos on r-eal-time object detection[J].arXiv preprint arXiv:2304.08069,2023.
[16] SUN F, HE N, LI R, et al. GD-PAN: a multiscale f-usion architecture applied to object detection in UAV aerial images[J]. Multimedia Systems, 2024, 30(3): 143-155.
[17] LIU S, QI L, QIN H F, et al. Path aggregation netw-orkfor instance segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recogn-ition. 2018: 8759-8768.
[18] WANG C, HE W, NIE Y,et al. Gold-YOLO: Efficientobject detector via gather-and-distribute mechanism[J]. Advances in Neural Information Processing Systems, 2024, 36: 51094-51112.
[19] MUZAMMUL M, ALGARNI A M, GHADI Y Y, et al. Enhancing UAV Aerial Image Analysis: IntegratingAdvanced SAHI Techniques with Real-Time DetectionModels on the VisDrone Dataset[J]. IEEE Access, 2024, 12:21621-21633.
[20] AKYON F C, ALTINUC S O, TEMIZEL A. Slicing aided hyper inference and fine-tuning for small object detection[C]//2022 IEEE International Conference on Image Processing (ICIP). IEEE, 2022: 966-970.
[21] LI D, YANG P, ZOU Y. Optimizing Insulator Defect Detection with Improved DETR Models[J].Mathematic-s, 2024, 12(10): 1507-1524.
[22] XIE Y, ZHENG S, LI W. Feature-guided spatial atten-tion upsampling for real-time stereo matching network[J]. IEEE MultiMedia, 2020, 28(1): 38-47.
[23] 张储,徐伟悦,杨如雪,等.一种基于优化后的 RT-DETR 模型的红花目标检测方法和装置:202410039910[P].2024-04-09.
ZHANG C, XU W Y, YANG R X, et al. A method and device for detecting red flower targets based on an optimized RT-DETR model:202410039910[P].2024-04-09.
[24] 李亦涵,张秀再,沈涛.一种改进 RT-DETR 算法的遥感图像目标检测方法及系统:202410609716[P].2024-06-14.
LI Y H, ZHANG X Z, SHEN T. An improved RT-DETR algorithm for remote sensing image object det-ection method and system:202410609716[P].2024-06-14.
[25] ZHANG X, SONG Y, SONG T, et al. AKConv: Con-volutional kernel with arbitrary sampled shapes and a-rbitrary number of parameters[J].arXiv preprint arXiv:2311.11587, 2023.
[26] 庞玉东,李志星,刘伟杰,等. 基于改进实时检测 Transfor-mer 的塔机上俯视场景小目标检测模型 [J/OL]. 计算机应用 (2024-04-27)[2024-08-04]. http://kns.cnki.net/kcms/detail/51.1307.TP.20240402.2133.013.html.
PANG Y D, LI Z X, LIU W J, et al. Small target detection model in overlookingscenes on tower cranes based on improved real-time detection Transformer[J/OL].Computer Engineering and Applications(2024-04-27)[2024-08-04]. http://kns.cnki.net/kcms/detail/51.1307.TP.20240402.2133.013.html.
[27] CHEN J, KAO S-H, HE H, et al. Run, Don't walk: Chasing higher FLOPS for faster neural networks[C]//Proceedings of the IEEE/CVF Conference on Comput-er Vision and Pattern Recognition, 2023: 12021-12031.
[28] 胡佳乐,周敏,申飞. 面向无人机小目标的 RTDETR 改进检测算法[J/OL].计算机工程与应用(2024-06-25)[2024-08-03]. http://kns.cnki.net/kcms/detail/11.2127.TP.20240624.1520.010.html.
HU J L, ZHOU M, SHEN F. Improved detection alg-orithm of RTDETR for UAV small target[J/OL].Comp-uter Engineering and Applications(2024-06-25)[2024-08-03].http://kns.cnki.net/kcms/detail/11.2127.TP.20240624.1520.010.html.
[29] OUYANG D, HE S, ZHANG G, et al. Efficient multi-scale attention module with cross-spatial learning[C]//ICASSP 2023-2023 IEEE International Conference on A-coustics, Speech and Signal Processing (ICASSP). IEEE, 2023: 1-5.
[30] LIU W, LU H, FU H, et al. Learning to Upsample by Learning to Sample[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 6027-6037.
[31] KANG M, TING C M, TING F, et al. ASF-YOLO: A Novel YOLO Model with Attentional Scale Seque-nce Fusion for Cell Instance Segmentation[J]. arXiv:2312.06458, 2023.
[32] TAN M, PANG R, LE Q V. Efficientdet:Scalable and efficient object detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern reco-gnition, 2020: 10781-10790.
[33] WANG J, CHEN K, XU R, et al. Carafe: Content-a-ware reassembly of features[C]//Proceedings of the IEEE/CVF international conference on computer vision, 2019: 3007-3016.
[34] DU D W, ZHU P F, WEN L Y, et al. VisDrone-DET2019: the vision meets drone object detection in ima-ge challenge results[C]//2019 IEEE/CVF lnternational Conference on Computer Vision Workshop. Seoul: lEEE, 2019: 213-22.
[35] ZHU Z, LIANG D, ZHANG S, et al. Traffic-sign de-tection and classification in the wild[C]//Proceedings of the IEEE conference on computer vision and patte-rn recognition. 2016: 2110-2118.
[36] XIA G S, BAI X, DING J, et al. DOTA: A large-sc-ale dataset for object detection in aerial images[C]//Pr-oceedings of the IEEE conference on computer vision and pattern recognition. 2018: 3974-3983.
[37] 雷帮军,余翱,余快. 基于YOLOv8s改进的小目标检测算法[J].无线电工程, 2024, 54(04): 857-870.

LEI B J, YU A, YU K.Small object detection algorit-hm based on improved YOLOv8s[J].2024, 54(04): 857-870.

[38] 李岩超,史卫亚,冯灿. 面向无人机航拍小目标检测的轻量级 YOLOv8 检测算法[J].计算机工程与应用, 204, 60(17): 167-178.

LI Y C, SHI W Y, FENG C. Lightweight YOLOv8 detection algorithm for small object detection in UAV aerial photography[J].Computer Engineering and Appli-cations, 204, 60(17): 167-178.