

◎YOLOv8改进及应用专题◎

优化改进YOLOv8实现实时无人机车辆检测的算法

史涛¹, 崔杰¹, 李松²

1. 天津理工大学 电气工程与自动化学院, 天津 300384

2. 华北理工大学 电气工程学院, 河北 唐山 063210

摘要:针对现有无人机车辆检测算法精度低、易受背景环境干扰、难以检测微小目标车辆问题,提出了一种改进YOLOv8的无人机车辆检测算法YOLOv8-CX。结合Deformable Convolutional Networks v1-3的优点,提出一种能够灵活采样特征的C2f-DCN模块,以更好地提取不同尺寸大小车辆之间的特征。利用Large Separable Kernel Attention的思想,提出了具有长程依赖性和自适应能力的SPPF-LSKA模块,可以有效减少背景对于车辆检测的干扰。在颈部网络,采用CF-FPN(feature network for tiny object detection)特征融合结构,通过结合上下文信息和抑制不同尺度特征之间的冲突信息,提升了对小目标的检测精度。最后,将原始YOLOv8的头部替换为Dynamic Head检测头。通过将尺度、空间和任务三种注意力机制结合统一,进一步提升了模型的检测性能。实验结果表明,在Maphai数据集上,改进算法与原算法相比准确率(P)、召回率(R)、平均精度(mAP)分别提升了8.5、11.2和6.2个百分点,且算法检测速度达到72.6 FPS,满足无人机车辆检测实时性的要求。通过与其他主流目标检测算法比较,验证了该方法的有效性和卓越性。

关键词:无人机车辆检测;YOLOv8;可变形卷积;注意力机制;特征融合

文献标志码:A **中图分类号:**TP391.4 **doi:**10.3778/j.issn.1002-8331.2312-0291

Algorithm for Real-Time Vehicle Detection from UAVs Based on Optimizing and Improving YOLOv8

SHI Tao¹, CUI Jie¹, LI Song²

1. School of Electrical Engineering and Automation, Tianjin University of Technology, Tianjin 300384, China

2. College of Electrical Engineering, North China University of Science and Technology, Tangshan, Hebei 063210, China

Abstract: To address the problems of low accuracy, easy interference from background environment and difficulty in detecting small target vehicles of existing UAV vehicle detection algorithms, an improved UAV vehicle detection algorithm YOLOv8-CX is proposed based on YOLOv8. By integrating the advantages of Deformable Convolutional Networks v1-3, a C2f-DCN module is proposed to flexibly sample features and better extract features between vehicles of different sizes. Utilizing the idea of large separable kernel attention, a SPPF-LSKA module is proposed with long-range dependency and self-adaptability, which can effectively reduce background interference on vehicle detection. In the neck network, a CF-FPN (feature network for tiny object detection) feature fusion structure is adopted to enhance the detection accuracy of small targets by combining contextual information and suppressing conflicts between features at different scales. Finally, the original YOLOv8 head is replaced with a Dynamic Head detection head. By unifying scale, space and task, the three types of attention mechanisms, the model detection performance is further improved. Experimental results show that on the Maphai dataset, compared with the original algorithm, the improved algorithm increases the accuracy (P), recall (R) and mean average precision (mAP) by 8.5, 11.2 and 6.2 percentage points respectively, and the algorithm detection speed reaches 72.6 FPS, meeting the real-time requirements of UAV vehicle detection. By comparing with other mainstream target detection algorithms, the effectiveness and superiority of this method are validated.

Key words: unmanned vehicle detection; YOLOv8; deformable convolution; attention mechanism; feature fusion

基金项目:国家自然科学基金(61203343)。

作者简介:史涛(1980—),男,博士,讲师,研究方向为类脑智能机器人、机器人视觉、生物启发的智能计算;崔杰(2000—),通信作者,男,硕士研究生,研究方向为深度学习、目标检测,E-mail:2864717531@qq.com;李松(2000—),男,硕士研究生,CCF学生会员,研究方向为深度学习、目标检测。

收稿日期:2023-12-19 **修回日期:**2024-02-23 **文章编号:**1002-8331(2024)09-0079-11

无人机车辆检测任务是利用无人机技术监测和识别地面车辆。它可以为交通监管、安全防护、突发事件处理和城市规划等方面提高管理效率,为监测周边车辆情况、提供实时交通情况和为城市规划等方面提供数据支持。无人机车辆检测任务对于提高社会管理效率和城市运行效率具有重要的应用价值。

无人机车辆检测算法是利用无人机拍摄的图像或视频进行车辆识别和定位的技术。目前主流的无人机车辆检测算法包括传统的基于图像处理和特征提取的方法,以及基于深度学习的方法。传统的基于图像处理和特征提取的方法包括 Haar 特征、HOG 特征等,这些方法通过提取图像中的局部特征并利用分类器进行车辆检测。例如,关晓斌等^[1]提出结合使用 HOG 特征与 SIFT 特征作为车辆检测的特征提取算法。张凯等^[2]基于梯度分布直方图(HOG)和支持向量机(SVM),开发了一种视频车辆识别算法。然而,这些方法需要大量手工设计特征,难以提取图像中的高级抽象特征。仅仅依靠低级特征识别效果有限。最近几年,随着深度学习在计算机视觉领域的广泛应用,目标检测技术取得长足进步。目标检测算法分为多阶段和单阶段两大类。常用的多阶段算法包括 R-CNN、Fast R-CNN、Faster R-CNN、U-net 等。魏相站等^[3]在特征提取网络中嵌入卷积模块的注意力机制模块结构,提出了一种基于改进 Faster RCNN 模型的检测方法,曹磊等^[4]基于 VGG-16 网络提取出图片的多层特征,提出一种基于改进 RPN 的 Faster-RCNN 网络 SAR 图像车辆目标检测方法。另一类是单阶段检测方法,采用回归和分类分支预测边界框和类别,实现快速端到端的检测。代表性的单阶段目标检测算法包括 YOLO(you only look once)和 SSD(single shot multibox detector)等。对于无人机车辆检测任务实时性的要求,单阶段算法相比较于多阶段算法因其快速的检测速度而更受关注。赵宇航等^[5]提出了一种基于 YOLO V3 深度卷积神经网络的车辆检测方法,通过构建一个基于残差网络的卷积神经网络提取每帧视频的深度特征,提高了车辆检测过程中的目标置信度,降低了漏检率。宋世奇等^[6]提出的改进 SSD 车辆检测算法利用 K-means++ 聚类算法获取适应航拍车辆目标的默认候选框参数信息,并在特征提取网络骨架中加入了具有抗混叠效应的可学习低通滤波层提高了小尺度目标车辆特征的有效传递提取。范江霞等^[7]提出的改进 YOLOv5s 算法对锚框尺寸进行修正,并通过增加注意力机制网络,提升了模型识别车辆的速度和准确性。

近段时间,针对无人机车辆检测算法精度低、易受背景环境干扰、难以检测微小目标车辆等问题,赵倩等^[8]提出一种基于多重金字塔和多尺度注意力的轻量级 YOLOv5 算法。在主干网络中减少下采样次数,提高了小目标检测能力。张利丰等^[9]提出了一种改进 YOLOv8

的车辆目标检测算法 RBT-YOLO,使用 SoftNMS(soft non maximum suppression)替换原有 NMS,使模型对候选框的处理方式更为温和,增强了模型对目标的检测能力。孙庆^[10]提出了一种基于深度学习的车辆检测算法,在特征融合部分使用深度可分离卷积(depthwise separable convolution, DSConv)替代 BiFPN 网络中的普通卷积,深度可分离卷积能够提高特征融合效率,进而减少背景环境对于车辆特征的干扰。张河山等^[11]提出一种改进的 YOLOX 网络用于无人机航拍图像的检测。在特征融合部分引入自适应空间特征融合(adaptively spatial feature fusion, ASFF)模块增强网络的特征学习能力,将二元交叉熵损失(BCE loss)函数替换为变焦距损失(varifocal loss)函数来加强网络对正样本的学习。但上述作者的研究只局限于增强小目标车辆的提取效果或者减少环境背景对于车辆特征提取的干扰等方面的一种,且提升改进效果有限。

自从单阶段目标检测算法 YOLO 面世以来,研究人员就在不断对其进行改进。Ultralytics 团队作为该算法的主要开发者,他们在 2023 年提出了 YOLO 的最新的版本 YOLOv8,算法在满足实时性的同时,还具备了较高的检测精度与鲁棒性,适合运用于对无人机目标车辆的检测进行针对性改进。因此,本文基于 YOLOv8n 对模型进行优化,提出了一种基于 YOLOv8n 的无人机实时车辆检测算法 YOLOv8-CX,本文的主要工作如下:

(1)引入 Deformable ConvNets v1-3^[12-14]改进 C2f 模块,针对不同尺寸大小的车辆,更加灵活地提取其特征。

(2)将 large separable kernel attention^[15]与 YOLOv8 中 SPPF 模块结合,通过捕获长程依赖性和适应性,来减少复杂背景对车辆检测的干扰。

(3)在 YOLOv8 颈部网络采用(ment network for tiny object deteciton, CF-FPN)^[16]特征融合结构以增强对小目标车辆的检测效果,CF-FPN 结构能够有效地结合上下文信息,并抑制不同尺度特征之间的冲突信息。

(4)将 YOLOv8 头部替换为 Dynamic Head^[17]检测头,通过结合尺度、空间和任务三种注意力函数,使得模型在上述三种改进下的性能效果进一步提升。

1 相关原理

1.1 YOLOv8 算法介绍

YOLO 是一种实时目标检测算法,它将图像划分为网格单元,并在每个单元内同时预测目标的边界框和类别,以实现快速准确的目标检测。YOLOv1^[18]是首个版本,采用单一尺度的全卷积神经网络进行目标检测,具有实时性和高效性的特点。YOLOv2^[19]在 YOLOv1 的基础上引入了多尺度特征图和 Anchor Boxes,提高了检测精度和鲁棒性。YOLOv3^[20]进一步改进了网络结构,采用了更深的 Darknet-53 作为特征提取网络,并引入了多

尺度预测和跨尺度连接,提高了检测速度和准确性。YOLOv4^[21]在YOLOv3的基础上引入了CSPDarknet53和SPP结构,进一步提高了检测性能和速度。YOLOv5^[22]通过引入更轻量级的网络结构和多尺度训练,提高了检测速度和鲁棒性。而YOLOv7^[23]则引入了可扩展的高效层聚合网络E-ELAN、创新的过渡模块和重参数化结构等策略,以增强特征提取和语义信息表达的能力,进一步优化了目标检测效果。

YOLO算法^[24-26]历经迭代,目前已经升级到YOLOv8,且基于缩放系数提供了N/S/M/L/X尺度的不同大小模型。YOLOv8算法结构参考了YOLOv7 ELAN的设计思想,在骨干和颈部网络中采用了YOLOv5的C2f结构,以获得更丰富的梯度流。在头部网络中,YOLOv8采用了解耦头结构,将分类和检测头分离。此外,YOLOv8采用了TaskAlignedAssigner正样本分配策略,并引入了Distribution Focal Loss作为回归损失。图1展示了YOLOv8算法结构。通过这些改进,YOLOv8在损失计算和网络结构方面取得了显著的进展。

1.2 算法存在的问题

考虑到无人机设备有限的硬件性能,本文采用权值与参数量最小的YOLOv8n作为算法模型。但在实际的检测过程中,YOLOv8n算法存在一些问题。首先,YOLOv8n算法采用的标准卷积在提取不同尺度大小车辆的特征时,效果不是很好。其次,在检测任务中,检测背景环境的复杂性对车辆检测造成了很大的干扰,特别是对于小目标车辆的检测有很大的挑战。但YOLOv8n

算法在这种情况下,对车辆和小目标车辆的检测结果并不理想,容易出现误检以及漏检的情况。

2 算法改进

为了解决YOLOv8n算法存在的问题,提出了一种改进YOLOv8n的检测网络模型YOLOv8-CX。图2为改进后的YOLOv8n结构示意图,首先在YOLOv8n的骨干网络,采用C2f-DCN模块替换第八层的C2f模块,SPPF-LSKA模块替换SPPF模块。其次在YOLOv8n的颈部网络使用CF-FPN结构进行特征融合。最后将YOLOv8n的头部网络更换为DyHead模块进行预测输出。C2f-DCN模块增强了对不同尺度车辆特征的提取能力。SPPF-LSKA模块目的是减少无关背景特征对车辆检测的干扰,提高模型在复杂环境的准确率。而CF-FPN特征融合结构通过结合上下文信息,提高了对小目标车辆的检测能力。DyHead模块则通过结合尺度、空间和任务三种注意力机制,使模型在识别不同尺度、复杂背景和微小目标车辆情况下的性能进一步提升。

2.1 C2f-DCN模块

2.1.1 可变形卷积DCN

传统卷积的优点是在处理一些静态或较为规则的低复杂度目标时具有良好的性能,但无人机车辆检测往往是动态任务,且不同类别的车辆尺寸差异较大,传统卷积可能无法很好地适应这些尺寸变化。而可变形卷积(DCN)可以根据目标的空间和形状位置进行动态调整,更好地捕捉目标特征来适应不同车辆动态尺寸的变

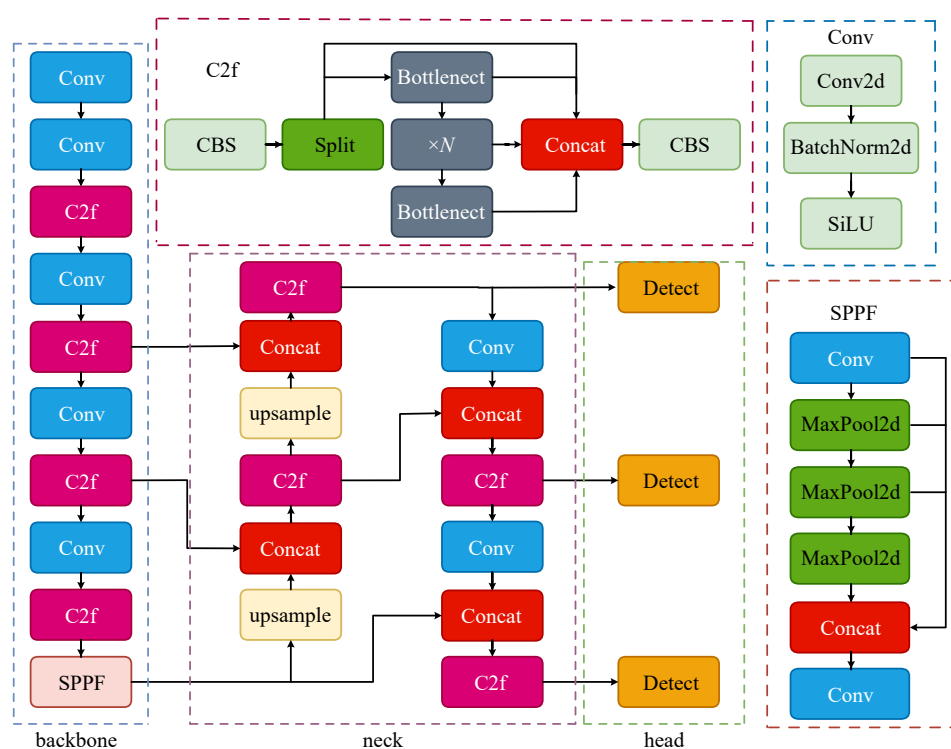


图1 YOLOv8网络结构

Fig.1 YOLOv8 network structure

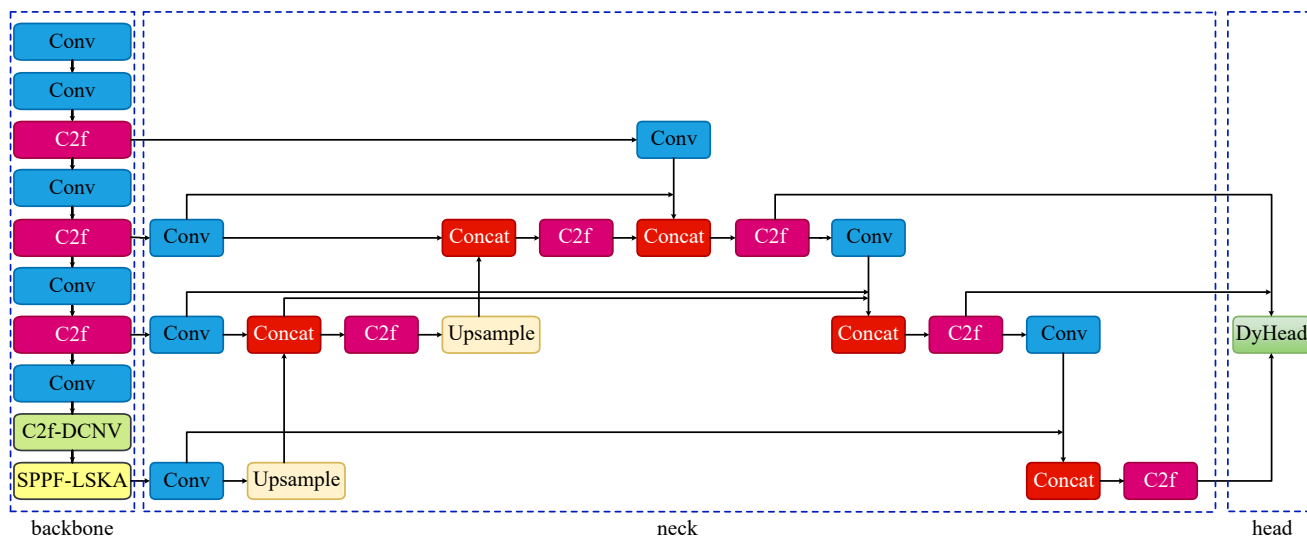


图2 YOLOv8-CX网络结构

Fig.2 YOLOv8-CX network structure

化。在可变形卷积(DCN)中,采样位置不再局限于规则的网格位置,而是通过引入一个偏移量(offset)来自由采样。过程如图3所示。图片左侧代表输入特征,右侧代表输出特征。首先卷积(Conv)对输入特征进行预处理,生成偏移量和调制量,然后利用这些偏移量将规则分布的 3×3 像素点(黑色)转换为不规则任意分布的9个像素点(虚线连接的像素点)。接着,利用生成的新像素点对输入图像进行采样,得到采样后的特征图。最后,将采样后的特征图与卷积核进行逐元素相乘,并进行求和操作,得到最终的卷积结果。

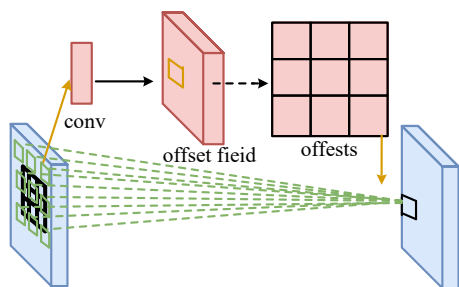


图3 可变形卷积过程

Fig.3 Deformable convolution process

2.1.2 可变形RoI池化

池化层是卷积神经网络中常见的层级结构,用作降低输入数据的维度,减少模型参数和计算量,同时提取数据中的关键特征。传统的池化层是在固定大小的窗口内进行最大值或平均值的操作,以降低特征图的尺寸。相比之下,可变形RoI(region of interest)池化层是一种针对目标检测任务设计的池化层,它可以根据目标的形状和大小来动态调整池化窗口的形状和大小,从而更好地适应不同尺寸变化目标的特征。可变形RoI池化过程如图4所示。首先,针对输入特征图,可变形RoI池化会生成对应的感兴趣区域,并将其划分为 $K \times K$ 个子区域。其次,针对每一个子区域的位置,计算其各自

对应的偏移量,对于每个子区域,根据偏移量调整采样位置,并得到调整采样位置后新子区域的特征值。接着,对每个新子区域的采样特征值进行自适应最大池化操作,以获得固定大小的输出。最后将所有新子区域的池化结果按照一定的顺序拼接在一起,形成最终的可变形RoI池化结果。与可变形卷积DCN不同的是,可变形RoI池化的偏移量是通过全连接(fc)层得到的。

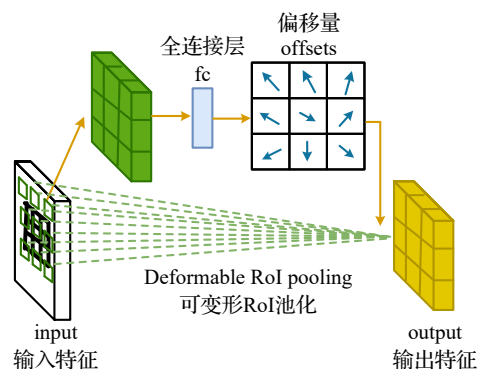


图4 可变形RoI池化过程

Fig.4 Deformable RoI pooling process

2.1.3 C2f-DCN 结构

提升多尺度特征提取能力C2f模块需要整合不同尺度的特征图。Deformable ConvNets可以学习非线性的变形field,更好捕捉特征点在不同尺度间的对应关系。这有利于提取多尺度特征的上下文信息。强化特征点对应关系学习Deformable ConvNets通过变形卷积核可以建立特征点在不同尺度间更复杂的对应关系。这比普通卷积更适用于C2f模块学习跨尺度特征点的变换。

在深层网络中,目标的形变特征往往更加复杂,因此,本文仅对YOLOv8骨干网络中第八层的C2f结构进行改进,将C2f结构中Bottleneck结构的常规卷积改为可变形卷积并融入可变形RoI池化层。改进后的C2f-DCN结构如图5所示。

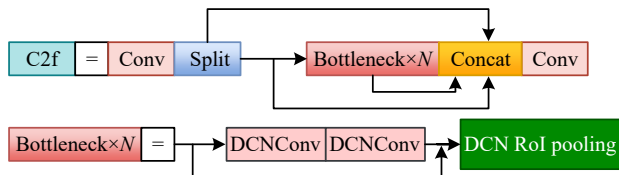


图5 C2f-DCN 结构

Fig.5 Structure of C2f-DCN

在改进的 Bottleneck 结构中,引入两层可变形卷积,使模型能够自由地对输入特征图进行采样,从而更好地学习到目标物体的尺度、背景、和形变等信息。经过卷积层后,利用残差连接将输出特征图与输入特征图进行融合。最后,采用可变形 RoI 池化,通过学习来自前面特征映射和 RoI 池化的偏移量,扩大模型的感受野,提取更复杂的特征。这样一来,改进后的模块能够自适应学习不同尺寸大小的物体,更好地理解学习不同尺寸下的车辆特征,具有更强的鲁棒性和泛化能力。

2.2 SPPF-LSKA 模块

由于无人机检测的车辆所处环境复杂多变,为了提高模型对车辆关键特征的提取能力,在 YOLOv8 骨干网络中的 SPPF 模块引入 LSKA (large separable kernel attention) 注意力机制,网络忽略无关背景信息干扰,注意到更多有效车辆特征信息。以往的注意力机制 self-attention 与 large kernel attention (LKA) 等存在许多不足, self-attention 注意力具有较强的长程依赖性和适应性,但忽略了图像的二维结构。LSA 注意力解决了 self-attention 注意力忽略图像二维结构的问题,但其在处理

大尺寸卷积核时计算量过大。本文引入的 LSKA 注意力机制很好地解决了 LSA 注意力在处理大尺寸卷积核时计算量过大的问题,以较小的计算成本实现了较高的性能。

LSKA 是一种创新的大型可分离核注意力模块,它通过对大核卷积操作进行分解来捕获长程依赖性和适应性,同时又将分解得到的 2D 卷积核再次分解为串联的 1D 核,降低了计算复杂度和内存需求。其过程结构如图 6 所示。首先将一个 $K \times K$ 的卷积分解为 $(2d-1) \times (2d-1)$ 深度卷积、 $K/d \times K/d$ 深度扩张卷积和 1×1 卷积,其次将 2D 的深度卷积核与深度扩张卷积核继续分解为 1D 的水平(横向)卷积核与垂直(纵向)卷积核,最后将分解后的卷积核依次串联。

改进后的 SPPF-LSKA 结构如图 7 所示。

2.3 CF-FPN 特征融合结构

在无人机车辆检测任务中,由于小目标车辆分辨率低、体积小,所以小目标车辆通常比大目标车辆更加难以检测。YOLOv8 中的颈部网络通过采用 PANet 结构融合不同尺度特征之间的信息,提高了对小目标车辆的检测能力。但这种特征融合的结构忽略了不同尺度特征之间冲突信息的存在,同时缺乏上下文信息,尤其是对于微小物体,它们很容易被冲突信息淹没。为了解决 YOLOv8 颈部网络中存在的问题,本文采用了将上下文增强和特征细化相结合的 CF-FPN 特征融合结构来替代 PANet 结构。CF-FPN 整体结构如图 8 所示。

CEM 模块采用不同膨胀率的空洞卷积来获取不同

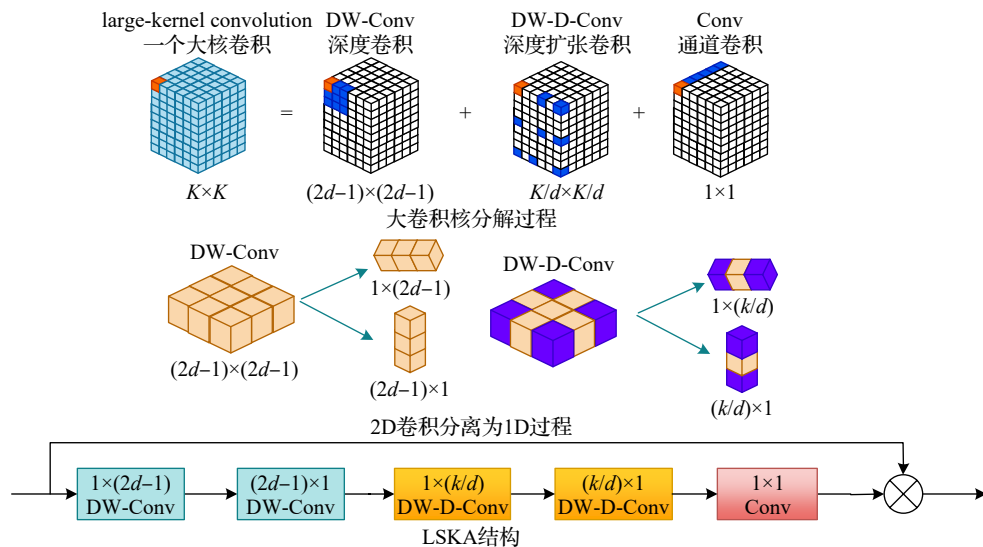


图6 LSKA 注意力结构

Fig.6 LSKA attention structure



图7 SPPF-LSKA 模块结构

Fig.7 SPPF-LSKA modular structure

下文信息,每个预测位置的预测都独立进行,缺乏全局视野。(3)预测头部参数量少,表达能力有限。难以深度挖掘特征中的空间结构信息。所以总体来说,YOLOv8原有头部的单尺度预测结构和缺乏动态学习能力,给目标检测尤其是多尺度目标检测带来一定限制。而Dynamic Head采用动态路由机制,可以更好地融合上下文信息,识别多尺度目标。它可以动态调整不同特征层的权重,有利于提取多尺度特征。

为进一步提升模型在识别不同尺度、复杂背景和小目标车辆时的性能,本文在YOLOv8头部网络引入DyHead(Dynamic Head)模块。DyHead模块通过利用自我注意力机制在尺度感知注意力、空间感知注意力与任务感知注意力三者上实现了统一。其中,在检测层上给定三维特征张量 $F \in \mathbb{R}^{L \times S \times C}$,该注意力计算公式(1)如下所示:

$$W(F) = \pi_C(\pi_S(\pi_L(F) \cdot F) \cdot F) \cdot F \quad (1)$$

其中, F 代表一个输入的三维张量 $L \times S \times C$, $\pi_L(\cdot)$ 、 $\pi_S(\cdot)$ 和 $\pi_C(\cdot)$ 分别为尺度感知注意力模块、空间感知注意力模块与任务感知注意力模块。并分别只作用到三维张量 F 的 L 、 S 和 C 维度上。 $\pi_L(\cdot)$ 、 $\pi_S(\cdot)$ 和 $\pi_C(\cdot)$ 注意力模块以及DyHead结构如图11所示。

图11中,图(a)表示了 $\pi_L(\cdot)$ 、 $\pi_S(\cdot)$ 与 $\pi_C(\cdot)$ 三种注意力模块的内部结构及三者串联构成的单个DyHead结构,这三种注意力模块应用与检测头部,可以多次堆叠使用。图(b)则表示多个DyHead模块叠加使用的过程。考虑到计算开销和模型性能等多方面因素,本文只使用一个DyHead模块。

3 实验与结果

3.1 数据集

Mapsai数据集是由Spoon Kutlery 2023年7月在Roboflow平台上发布的公开汽车数据集。该数据集将被检测车辆分为小型货车、野营车、汽车、摩托车、小型卡车、大型卡车、拖拉机、大型货车与其他车等九个类别,共有9357张图像。其中8085张图像作为训练集,774张图像作为验证集,398张图像作为测试集。

3.2 实验设备及评价指标

本模型所有实验环境都是在基于ubuntu20.04系统下进行的,硬件配置为CPU,14vCPU, Intel® Xeon® Gold 6330 CPU; GPU, RTX 3090, 24 GB GPU, 编译环境为 python 3.8.10+pytorch 1.13+cuda 11.3。

在实验时,将初始学习率设置为0.01,批量大小设置为32,权重的衰减率设置为0.0005,热身轮数设置为3,色相的增强系数设置为0.015,饱和度的增强系数为0.7,亮度的增强系数为0.4,训练周期设置为150。

为了方便评价模型改进后的性能,本文将准确率(P)、召回率(R)、平均精度(mAP)、每秒帧数(FPS)、Giga浮点运算每秒(GFLOPs)、权重文件(MB)大小作为本文的评价指标。

准确率(P)为所有预测为正样本的结果中,预测正确的比率。定义如下:

$$P = \frac{TP}{TP + FP} \quad (2)$$

召回率(R)为是指在所有实际正例中,模型正确识别为正例的比例。定义如下:

$$R = \frac{TP}{TP + FN} \quad (3)$$

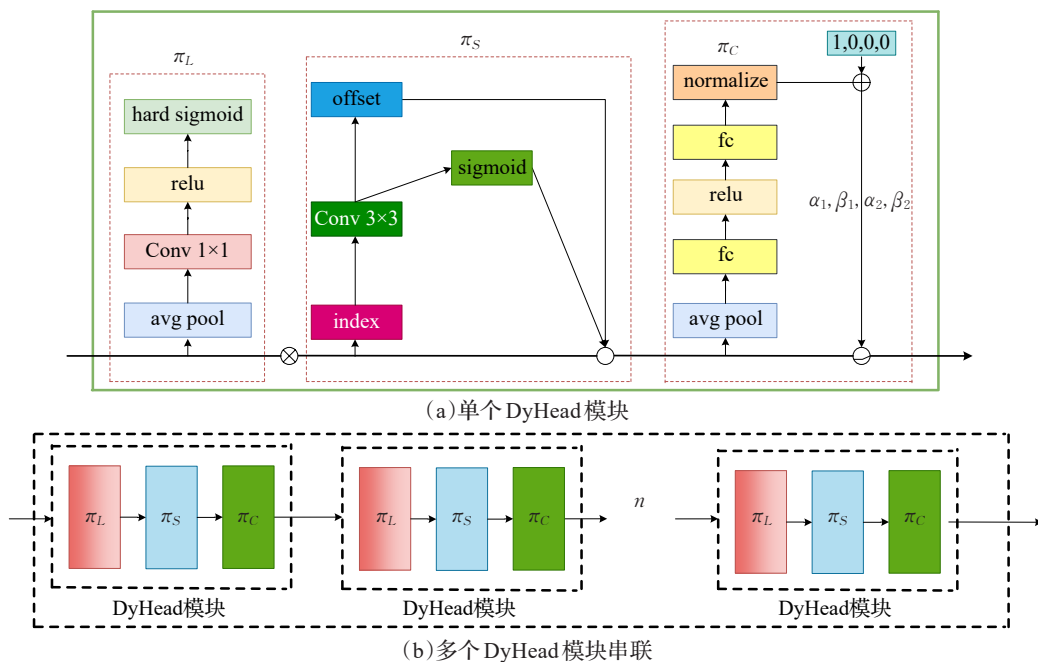


图11 DyHead模块结构图

Fig.11 DyHead module structure diagram

其中, TP 表示检测结果中预测正确的样本数量、FP 表示检测结果中预测错误的样本数量、FN 表示样本所有正确目标中未被检测到的数量。

平均精度(mAP)对 m 个类别的准确率(P)与召回率(R)综合考虑,较全面地反应网络的性能。定义如下:

$$mAP = \frac{1}{m} \sum_{i=1}^m \int_0^1 P(R) dR \quad (4)$$

每秒帧数(FPS)反映了模型的推理速度,当FPS大于30时,模型就满足了实时性检测的要求。Giga浮点运算每秒(GFLOPs)与权重文件(MB)是度量模型复杂度的指标。GFLOPs与权重文件(MB)越小,表示模型所需要的计算能力就越小。

3.3 CF-FPN 结构中不同特征融合方式性能对比

在CF-FPN特征融合结构CEM模块中支持加权融合(weighted fusion)、自适应融合(adaptive fusion)和串联融合(concatenation fusion)三种特征融合方式。为了验证本文使用串联融合方式的性能优越性,与另外两种融合方式进行了实验对比,实验结果如表1所示。

表1 不同特征融合方式性能对比

Table 1 Performance comparison of different feature fusion methods

Fusion	mAP50/%	FPS	GFLOPs	权重文件/MB
Weighted Fusion	56.3	105.0	7.5	4.4
Adaptive Fusion	55.9	90.2	7.5	4.4
Concatenation Fusion	56.4	126.7	7.3	4.4

从表1中可以看出,串联融合的特征融合方式相比于其他两种方式,串联融合有着更高的检测精度与检测速度,且模型所需的计算能力最小。

3.4 串联不同数量DyHead模块的性能对比

DyHead模块串联的数量决定着模型检测的性能。本文对比了串联1~6块DyHead模块对于模型性能的影响。由于串联操作增加了模型计算的复杂度,所以每添加一块DyHead模块就要多付出0.5 GFLOPs左右的计算力。对比实验结果如表2所示。

表2 不同数量DyHead模块性能对比

Table 2 Performance comparison of different numbers of DyHead modules

模块数量	mAP50/%	FPS	GFLOPs	权重文件/MB
1	59.2	142.2	9.6	7.2
2	58.9	89.2	10.2	8.2
3	56.4	50.1	10.7	9.2
4	57.6	44.2	11.2	10.2
5	57.9	37.1	11.8	11.2
6	57.8	35.3	12.3	12.2

从表2可以看出,并不是串联的Dyhead模块越多性能就会越好,反而随着串联模块的增加,模型的性能却逐渐降低。因此,本文只串联1块DyHead模块进行预测输出。

3.5 部分改进模块横向对比实验

为了验证模块的性能和选择最佳模块,本文设计了对SPPF-LSKA和CF-FPN模块的横向对比实验。通过横向对比实验,将SPPF与不同的注意力机制如DAattention、self-attention、LSA和LSKA进行融合。实验结果如表3所示。将YOLOv8的颈部网络采用EfficientRepBiPAN、AFPN、RepGFPN和CF-FPN四种不同的特征融合结构。实验结果如表4所示。

表3 SPPF与不同注意力机制融合横向对比

Table 3 Horizontal comparison of SPPF fused with different attention mechanisms

SPPF融合的注意力机制	mAP50/%	FPS	GFLOPs	权重文件/MB
SPPF-DAattention	53.1	122.1	8.3	6.8
SPPF-self-attention	54.1	116.5	8.3	6.8
SPPF-LSA	55.6	111.3	10.8	7.6
SPPF-LSKA	56.7	122.2	8.3	6.8

表4 颈部网络不同特征融合结构横向对比

Table 4 Horizontal comparison of different feature fusion structures in neck networks

融合结构	mAP50/%	FPS	GFLOPs	权重文件/MB
EfficientRepBiPAN	49.1	94.1	7.2	6.0
AFPN	49.3	72.9	8.4	5.6
RepGFPN	53.8	110.5	8.3	6.9
CF-FPN	56.4	126.7	7.3	4.4

从表3可以看出,在mAP50指标上,将LSKA注意力机制与SPPF融合后取得了最好的性能为56.7%,表明LSKA是与SPPF融合时最有效的注意力机制。在每秒帧数(FPS)指标上,SPPF-DAattention的速度最快为122.1 FPS,而SPPF-LSA的速度最慢为111.3 FPS,说明LSA机制与SPPF融合时计算成本最大。在模型大小方面,LSA注意力机制最大,其他三者保持相同。综合考虑mAP50、FPS和模型大小,SPPF-LSKA表现最佳,同时取得最高准确率和较高速度,证明LSKA是与SPPF融合时的最佳选择。

从表4可以得到,在准确率mAP50指标上,CF-FPN达到56.4%,高于其他三种结构,表明CF-FPN在提取和融合特征的能力上更强。在运行效率FPS指标上,CF-FPN的126.7 FPS也最高,说明CF-FPN在实时性能上优于其他结构。在计算复杂度GFLOPs指标上,CF-FPN虽然不如RepGFPN低,但比其他两种结构低,计算效率较好。在模型大小指标权重文件大小上,CF-FPN仅为4.4 MB,比其他三种结构更小,部署和应用更方便。总体来说,CF-FPN特征融合结构在检测准确率,运行效率,计算效率和模型大小四个方面都表现出色,证明其特征提取和融合能力最强,是一种优秀的实时目标检测特征融合结构。

3.6 消融实验

为了更好地验证改进的各模块以及改进模块之间

表5 消融实验结果
Table 5 Fusion experiment results

C2f-DCN	SPPF-LSKA	DyHead	CF-FPN	P/%	R/%	mAP50/%	FPS	GFLOPs	权重文件/MB
×	×	×	×	57.2	52.9	54.7	115.3	8.1	6.3
√	×	×	×	57.3	56.2	55.2	111.8	8.0	6.1
×	√	×	×	59.4	54.7	56.7	122.2	8.3	6.8
×	×	√	×	61.2	57.4	59.2	142.2	9.6	7.2
×	×	×	√	54.1	63.1	56.4	126.7	7.3	4.4
√	√	×	×	57.4	63.2	57.6	118.4	8.2	6.6
√	√	√	×	60.4	63.1	60.2	80.3	9.8	7.6
√	√	×	√	58.7	57.5	57.7	115.8	7.5	4.8
√	×	√	√	64.0	59.3	60.4	78.7	9.1	5.7
√	√	√	√	65.7	64.1	60.9	72.6	9.4	6.3

注:其中√表示有加入该模块,×则表示没有加入该模块。

的组合对于原模型性能提升的影响。本文设计了消融实验。实验结果如表5所示。

通过表5实验结果证明了随着各个改进模块的加入,模型在准确率(P)、召回率(R)和平均精度(mAP)等方面都有不同程度的提升。C2f-DCN模块和SPPF-LSKA模块分别对模型准确率和召回率有所提升。加入C2f-DCN模块可以提升准确率,但会降低FPS。SPPF-LSKA模块可以提高准确率,但会增加计算量和模型大小。CF-FPN模块再将GFLOPS降低0.8的同时且对于模型召回率提升了10个百分点左右,提升幅度较大,证明了CF-FPN特征融合结构不仅不会增加运算成本,而且会降低运算成本。且对于小目标车辆的检测有很好的效果。而DyHead模块由于三种注意力的统一,在准确率、召回率和平均精度上都有较为明显的提升。组合使用多个模块(如C2f-DCN、SPPF-LSKA、DyHead)可以进一步提升准确率,但会影响FPS和模型大小。最后将各个模块结合起来,将它们的优势取长补短。结合后的模型在准确率、召回率和平均精度上都达到了最高,且FPS为72.6仍然满足实时性检测的需求。

3.7 实际检测效果验证

原YOLOv8n算法在面对不同尺寸、复杂背景和小目标的车辆时效果不是很好。为了验证改进后YOLOv8-CX算法在面对这些条件下的检测效果,本文挑选了一些具有针对性的图片来进行测试。YOLOv8-CX算法测试效果如图12所示。

从图12中可以看出,改进后YOLOv8-CX算法在面对不同背景环境与不同尺寸的车辆时都有很好的检测效果,且能够有效地检测到小目标车辆。证明改进后YOLOv8-CX算法具有较强的鲁棒性和准确性。

3.8 不同算法之间的对比实验

为了进一步验证算法的性能,且考虑到无人机车辆检测任务实时性的要求对于多阶段目标检测算法不适合应用,所以本次实验将YOLOv8-CX算法与当前一些主流单阶段目标检测算法做了对比,结果如表6所示。



图12 YOLOv8-CX检测效果图

Fig.12 YOLOv8-CX detection rendering

表6 不同算法对比结果

Table 6 Comparison of results from different algorithms

Algorithm	P/%	R/%	mAP50/%	FPS	GFLOPs	权重文件/MB
Efficientnet-b0	52.7	51.2	52.3	81.5	13.2	9.7
RT-DETR-L	58.7	49.9	48.7	58.4	57.0	40.5
YOLOv3n	49.7	51.2	48.7	133.7	12.0	8.3
YOLOv5n	53.2	49.1	50.1	129.4	7.1	5.4
YOLOv6n	39.0	53.4	44.4	133.8	11.8	8.7
YOLOv7n	56.7	57.3	49.7	117.3	13.1	7.5
YOLOv8n	57.2	53.9	54.7	115.3	8.1	6.3
GOLD-YOLO	57.7	54.1	55.2	119.3	10.5	7.6
RT-DETR-r18	59.3	50.1	51.2	55.2	58.3	47.3
YOLOv8-CX	65.7	64.1	60.9	72.6	9.4	6.3

与Efficientnet和RT-DETR算法相比较,YOLOv8-CX在更低的GFLOPs与权重文件下,准确率(P)分别提升了13和7个百分点,召回率(R)提升了12.9和14.2个百分点,mAP50提升了8.6和12.2个百分点。与最新的实时目标检测算法RT-DTER-18和华为推出的GOLD-YOLO相比,YOLOv8-CX仍然具有最高的检测性能,相较于RT-DTER-18算法,YOLOv8-CX在准确率(P)、召回率(R)和mAP50分别提升了6.4、4.0和9.7个百分点。相较于GOLD-YOLO算法,YOLOv8-CX在准确率(P)、召回率(R)和mAP50分别提升了8.0、10.0和5.7个百分点。

而与YOLOv3、YOLOv5、YOLOv6、YOLOv7和YOLOv8等轻量级YOLO系列算法比较,YOLOv8-CX算法在仍然满足实时性检测的情况下,准确率(P)分别提升了16、12.5、26.7、9和8.5个百分点,召回率(R)提升了12.9、15、10.7、6.8和10.2个百分点,mAP50提升了12.2、10.8、16.5、11.2和6.2个百分点。涨幅明显。

实验结果表明,与其他单阶段目标检测算法比较,YOLOv8-CX算法以较低的模型体积实现了极高的检测精度。且仍然满足实时性的检测需求。具有高效性与优异性。

4 结论

本文提出了一种改进YOLOv8n的无人机车辆检测算法。解决了传统YOLOv8n算法在无人机车辆检测任务中存在的问题。在骨干网络,提出C2f-DCN和SPPF-LSKA模块,C2f-DCN模块增强了模型对于不同尺寸大小车辆特征的提取能力,SPPF-LSKA模块减少了背景环境对于车辆特征的干扰。在颈部网络,引入新的特征融合结构CF-FPN提高了模型对于小目标车辆的检测能力。在头部网络,使用集成了尺度、空间和任务三种注意力函数的DyHead检测头进一步提升了模型的性能。实验表明,与现有模型相比,改进后的YOLOv8-CX算法精度更高、对复杂环境和微小目标情况下汽车的检测效果更好,且满足实时性的检测需求,具有实际的应用价值。但考虑到成本及硬件资源,未来的研究重点是在保证模型性能不降低的情况下,对模型进行压缩,以更好地部署到无人机检测设备上。

参考文献:

- [1] 关晓斌, 李战明. 基于SIFT和HOG特征融合的视频车辆检测算法[J]. 计算机与数字工程, 2021, 49(6): 1113-1117.
GUAN X B, LI Z M. Vehicle detection algorithm based on video SIFT and HOG feature fusion[J]. Computer & Digital Engineering, 2021, 49(6): 1113-1117.
- [2] 张凯, 李华文. 一种基于SVM和HOG特征的视频车辆识别算法[J]. 电子世界, 2019(7): 74-75.
ZHANG K, LI H W. A video vehicle recognition algorithm based on SVM and HOG features[J]. Electronics World, 2019 (7): 74-75.
- [3] 魏相站, 邵丽萍, 周骅. 基于改进的Faster RCNN模型在车辆类型检测中的应用[J]. 智能计算机与应用, 2020, 10(7): 97-100.
WEI X Z, SHAO L P, ZHOU Y. Application of improved Faster RCNN model in vehicle type detection[J]. Intelligent Computer and Applications, 2020, 10(7): 97-100.
- [4] 曹磊, 王强, 史润佳, 等. 基于改进RPN的Faster-RCNN网络SAR图像车辆目标检测方法[J]. 东南大学学报(自然科学版), 2021, 51(1): 87-91.
CAO L, WANG Q, SHI R J, et al. Method for vehicle target detection on SAR image based on improved RPN in Faster-RCNN [J]. Journal of Southeast University(Natural Science Edition), 2021, 51(1): 87-91.
- [5] 赵宇航, 左辰煜, 朱俊杰, 等. 基于YOLO V3的无人机航拍车辆检测方法[J]. 电子世界, 2020(13): 110-111.
ZHAO Y H, ZUO C Y, ZHU J J, et al. Unmanned aerial vehicle detection method based on YOLO V3[J]. Electronic World, 2020 (13): 110-111.
- [6] 宋世奇, 李旭, 祝雪芬, 等. 基于改进SSD的航拍城市道路车辆检测方法[J]. 传感器与微系统, 2021, 40(1): 114-117.
SONG S Q, LI X, ZHU X F, et al. Urban road vehicle detection method by aerial photography based on improved SSD [J]. Transducer and Microsystem Technologies, 2021, 40(1): 114-117.
- [7] 范江霞, 张文豪, 张丽丽, 等. 改进YOLOv5的无人机影像车辆检测方法[J]. 遥感信息, 2023, 38(3): 114-121.
FAN J X, ZHANG W H, ZHANG L L, et al. Vehicle detection method of UAV imagery based on YOLOv5[J]. Remote Sensing Information, 2023, 38(3): 114-121.
- [8] 赵倩, 杨一聪. 多重金字塔的轻量化遥感车辆小目标检测算法[J]. 电子测量技术, 2023, 46(13): 88-94.
ZHAO Q, YANG Y C. Small object detection algorithm for lightweight remote sensing vehicles with multiple pyramids [J]. Electronic Measurement Technology, 2023, 46(13): 88-94.
- [9] 张利丰, 田莹. 改进YOLOv8的多尺度轻量化车辆目标检测算法[J]. 计算机工程与应用, 2024, 60(3): 129-137.
ZHANG L F, TIAN Y. Improved YOLOv8 multi-scale and lightweight vehicle object detection algorithm[J]. Computer Engineering and Applications, 2024, 60(3): 129-137.
- [10] 孙庆. 基于Transformer和BiFPN的轻量化车辆检测算法研究[D]. 西安: 长安大学, 2023.
SUN Q. Research on a lightweight vehicle detection algorithm based on Transformer and BiFPN[D]. Xi'an: Chang'an University, 2023.
- [11] 张河山, 范梦伟, 谭鑫, 等. 基于改进YOLOX的无人机航拍图像密集小目标车辆检测[J]. 吉林大学学报(工学版): 1-13[2023-12-28]. <https://doi.org/10.13229/j.cnki.jdxbgxb.20230779>.
ZHANG H S, FAN M W, TAN X, et al. Vehicle detection of dense small targets in UAV aerial images based on improved YOLOX[J]. Journal of Jilin University (Engineering Science Edition): 1-13[2023-12-28]. <https://doi.org/10.13229/j.cnki.jdxbgxb.20230779>.
- [12] DAI J, QI H, XIONG Y, et al. Deformable convolutional networks[C]//2017 IEEE International Conference on Computer Vision (ICCV), 2017: 764-773.
- [13] ZHU X Z, HU H, LIN S, et al. Deformable ConvNets V2: more deformable, better results[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),

- Long Beach, CA, USA, June 15-20, 2019. New York: IEEE Press, 2019: 9300-9308.
- [14] WANG W H, DAI J F, CHEN Z, et al. InternImage: exploring large-scale vision foundation models with deformable convolutions[J]. arXiv:2211.05778, 2022.
- [15] LAU K W, PO L M, UR REHMAN Y A. Large separable kernel attention: rethinking the large kernel attention design in CNN[J]. arXiv:2309.01439, 2023.
- [16] XIAO J S, ZHAO T, YAO Y T, et al. Context augmentation and feature refinement network for tiny object detection [C]//Under Review As a Conference Paper at ICLR 2022, 2022.
- [17] DAI X, CHEN Y, XIAO B, et al. Dynamic head: unifying object detection heads with attentions[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 7373-7382.
- [18] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [19] REDMON J, FARHADI A. YOL9000: better, faster, stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7272.
- [20] REDMON J, FARHADI A. YOLOv3: an incremental improvement[J]. arXiv:1804.02767, 2018.
- [21] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[J]. arXiv: 2004.10934, 2020.
- [22] 王鹏飞, 黄汉明, 王梦琪. 改进YOLOv5的复杂道路目标检测算法[J]. 计算机工程与应用, 2022, 58(17): 81-92.
- WANG P F, HUANG H M, WANG M Q. Complex road target detection algorithm based on improved YOLOv5[J]. Computer Engineering and Applications, 2022, 58(17): 81-92.
- [23] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: 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.
- [24] 刘卫光, 刘东, 王璐. 可变形卷积网络研究综述[J]. 计算机科学与探索, 2023, 17(7): 1549-1564.
- LIU W G, LIU D, WANG L. Survey of deformable convolutional networks[J]. Journal of Frontiers of Computer Science and Technology, 2023, 17(7): 1549-1564.
- [25] 赵珊, 郑爱玲, 刘子路, 等. 通道分离双注意力机制的目标检测算法[J]. 计算机科学与探索, 2023, 17(5): 1112-1125.
- ZHAO S, ZHENG A L, LIU Z L, et al. Object detection algorithm based on channel separation dual attention mechanism[J]. Journal of Frontiers of Computer Science and Technology, 2023, 17(5): 1112-1125.
- [26] 赵振兵, 王帆帆, 刘良帅, 等. 基于注意力特征融合YOLOv5模型的无人机输电线路航拍图像金具检测方法[J]. 电测与仪表, 2023, 60(3): 145-152.
- ZHAO Z B, WANG F F, LIU L S, et al. Transmission line image fitting detection method based on attention feature fusion YOLOv5 model[J]. Electrical Measurement & Instrumentation, 2023, 60(3): 145-152.