

基于高分辨率特征引导学习的小目标检测方法^{*}

涂晓光 李卓骏 刘建华 杨明 魏麟

(中国民用航空飞行学院航空电子电气学院 成都 641450)

摘要: 针对小目标检测中由于特征信息不足、特征图分辨率较低等问题导致检测精度和鲁棒性较差的技术瓶颈,提出一种基于高分辨率特征引导学习的小目标检测方法。该方法采用基于上下文聚合风车卷积的改进 YOLOv11 算法结构,构建了一个由高分辨率检测分支与低分辨率检测分支组成的双通道检测框架。训练过程中,通过高分辨率检测网络对低分辨率检测网络进行引导学习,缓解低分辨率图像中小目标语义信息不足的问题。在双通道网络的中间层引入多损失函数加权的多尺度特征对齐损失函数,增强了小目标特征的表达能力。实验结果表明,本文的方法在 PASCAL VOC 2012 小目标数据集上的 mAP50 相较于原始的 YOLOv11 提高了 4.11%,mAP50:95 提高了 4.07%;在 Visdrone2019 数据集上 mAP50 相较于原始的 YOLOv11 提高 2.24%,mAP50:95 提高了 1.50%。

关键词: 小目标检测;深度学习;双通道网络模型;多尺度特征对齐损失

中图分类号: TP391.41;TN911.73 **文献标识码:** A **国家标准学科分类代码:** 520.2050

Small object detection method based on high-resolution
feature-guided learning

Tu Xiaoguang Li Zhuojun Liu Jianhua Yang Ming Wei Lin

(College of Aviation Electronic and Electrical Engineering, Civil Aviation Flight University of China, Chengdu 641450, China)

Abstract: To address the technical bottlenecks of poor detection accuracy and robustness in small object detection caused by insufficient feature information and low-resolution feature maps, this paper proposes a small object detection method based on high-resolution feature-guided learning. The method adopts an improved YOLOv11 algorithm structure based on context aggregation pinwheel convolution and constructs a dual-channel detection framework consisting of a high-resolution detection branch and a low-resolution detection branch. During the training process, the high-resolution detection network guides the learning of the low-resolution detection network, alleviating the problem of insufficient semantic information of small objects in low-resolution images. A multi-scale feature alignment loss function with weighted multi-loss functions is introduced in the middle layer of the dual-channel network to enhance the expressive ability of small object features. Experimental results show that the proposed method achieves a 4.11% improvement in mAP50 and a 4.07% improvement in mAP50:95 compared to the original YOLOv11 on the PASCAL VOC 2012 small object dataset; on the Visdrone2019 dataset, the mAP50 increases by 2.24% and the mAP50:95 increases by 1.50% compared to the original YOLOv11.

Keywords: small object detection; deep learning; dual-channel network model; multi-scale feature alignment loss

0 引言

小目标检测作为计算机视觉领域的一个重要研究方向,广泛应用于多个行业和场景中,尤其在智能监控、无人驾驶、遥感影像分析等领域具有重要意义。小目标通常具有尺寸小、信息有限、特征信息不明显、模型易将小目标误

识别为背景、易受背景干扰等特点,这使得传统目标检测方法在处理小目标时面临较大的挑战^[1-3]。随着深度学习技术的发展,多种方法如通过多尺度特征融合、注意力机制、数据增强、Transformer^[4]架构等策略,被广泛应用于提升小目标检测的性能,同时社会对小目标检测的实时性和高效性的需求不断提升,因此,利用深度学习方法增强小目标

收稿日期:2025-09-02

^{*} 基金项目:国家自然科学基金(62406207)、四川省自然科学基金(2025ZNSFSC1502)、四川省通用航空器维修工程技术研究中心资助课题(GAMRC2023YB06)、中央高校基本科研业务费(25CAFUC03023)项目资助

检测的性能,在深度学习目标检测领域具有重要研究意义^[5]。

基于深度学习的目标检测算法主要分为 3 类:两阶段检测、单阶段检测和无锚点检测。两阶段方法(Fast R-CNN^[6]、Faster R-CNN^[7]等)首先生成候选区域再进行精细检测,精度较高但计算复杂;单阶段方法(SSD^[8]、YOLO^[9]系列等)通过端到端回归直接预测目标位置和类别,显著提升检测速度;无锚点方法(CornerNet^[10]、CenterNet^[11]等)摒弃传统锚框设计,通过关键点或中心点定位目标,在简化模型结构的同时更适应小目标检测需求,体现了目标检测技术从复杂先验约束向高效数据驱动的演进趋势。

安梦军等^[12]解决了水面波动、光照变化以及漂浮物部分遮挡等复杂场景中水面小目标漂流物检测问题,在 YOLOv11 模型的基础上,设计了 MSWTC 结构,提高了召回率,从而提升了小目标的捕捉能力。王浩等^[13]使用 FasterBlock 重构 YOLOv8 的 C2f 模块,既提升模型特征提取能力又更加轻量化,以解决在交通标志识别应用中,待检测目标多为小目标,易出现漏检、误检等问题。刘洋等^[14]修改了 YOLOv8 的头部功能并提出 LP-Detect 检测头,以提高网络对小目标的检测精度并降低漏检率。王迎龙等^[15]提出一种复杂大视场下低慢小无人机目标检测方法 BG-YOLO,该方法通过在高分辨率特征层添加检测头,有效减少了图像下采样过程中的信息丢失,提升了模型对小目标的检测能力,张莹等^[16]提出了一种多通道特征融合学习的印制电路板小目标缺陷检测网络 YOLOPCB,该方法通过改进 YOLOv7 网络,并设计了浅层特征融合模块与新的 anchors 匹配策略,大大提高模型性能。综上,尽管小目标检测算法通过不断提升和修改以适应更多的检测场景,但仍面临两个艰巨挑战:1)现有单通道网络结构难以有效捕获低分辨率图像中小目标的特征细节;2)缺乏有效机制对高分辨率特征信息的迁移和利用,大多数现有方法难以利用高分辨率图像中更丰富的细节信息去辅助低分辨率图像的小目标检测,缺少强有效的跨分辨率特征约束手段,导致低分辨率图像的检测效果受到限制。

在之前的研究中,Zhang 等^[17]提出了 SO-DETR 模型,结合双域特征和知识蒸馏策略,提升了小目标检测的效果;Ma 等^[18]提出了层次匹配的知识蒸馏网络,通过特征金字塔网络和注意力机制改善学生网络的小目标学习能力;Zhu 等^[19]提出了 ScaleKD 方法,通过尺度解耦特征蒸馏和跨尺度助手模块提高了小目标检测性能;Jobaer 等^[20]提出了一种新的知识蒸馏框架,专门针对模糊环境下的小目标检测,利用去模糊子网和注意力模块提高检测精度。

这些蒸馏学习方法均是通过大规模、复杂的教师网络和小规模、轻量化的学生网络进行目标检测,教师网络将小目标检测的知识传递给学生网络。这种方法在一定程度上提升了小目标检测的精度和效率。然而,基于蒸馏学习的

模型仍然存在一些局限性。首先,教师网络通常计算量大、训练成本高,教师网络和学生网络之间的知识传递效率较低,蒸馏过程需要额外的训练时间和计算资源。此外,教师网络需要特别设计关注小目标的机制,若教师网络本身无法有效识别小目标,蒸馏过程学生网络将无法超越教师网络获得更好效果。其次,由于教师网络输出的是高分辨率特征,而学生网络由于轻量化设计,对齐主要依赖蒸馏策略,可能会出现不完全对齐导致部分空间信息的丢失,这使得学生网络难以完全捕捉到小目标的细节特征。

通过改进双通道的蒸馏学习方法以更适应低分辨率下小目标的检测,本文首次提出了一种全新的方法——高分辨率特征引导学习。与传统的蒸馏学习方法不同,该方法并非单纯依赖教师网络的知识传递,而是通过高分辨率网络引导低分辨率网络。具体表现在:1)不需要额外大模型作为教师,高分辨率特征直接指导低分辨率网络,信息传递更直接更专注于小目标检测,网络也更加轻量化;2)由于两路网络使用结构相同,即可自然对齐特征空间与语义;3)该方法并不完全依赖教师网络对小目标的关注能力,通过专门设计低分辨率输入结合高分辨率引导,更贴合小目标特点。

针对双通道的蒸馏学习过程中信息丢失严重、中间传递过程计算量大、传递效果弱、高维教师网络缺乏对小目标的细致关注等问题,本文提出一种新颖的约束特征图方式——多尺度特征对齐损失(multi-scale feature alignment loss,MSFAL),以缓解小目标语义信息不足并解决蒸馏网络无法自然对齐的问题;此外,在数据预处理阶段,对整个数据集进行多倍降采样操作,从数据层面提升网络对多小目标场景的适应能力和学习效果;最后,本文采用了上下文聚合风车卷积(context aggregation pinwheel convolution, CAPC)来提高提升原本 YOLOv11 检测能力。

1 基于高分辨率特征引导学习的小目标检测方法

1.1 YOLOv11 算法结构

Ultralytics 公司在 2023 年推出 YOLOv8 模型,基于不同深度和宽度的缩放系数分成了 n/s/m/l/x 等版本,以用于不同场景的需求^[21]。而 YOLOv11 于 2024 年发布,它建立在以前 YOLOv8 成功的基础上,引入了 C3k2(cross-stage partial kernel 2)块和 C2PSA(cross-stage progressive spatial attention)块等新型组件。C3k2 取代了 backbone 和 neck 中的 C2f(cross-stage feature fusion)块,并采用两个较小的卷积而不是一个大型卷积,从而减少了处理时间。同时通过引入跨阶段局部空间注意力模块,以使模型能够更有效地关注图像中的重要区域,从而有可能提高检测准确性,实现了更快的特征聚合,巩固了 YOLOv11 作为迄今为止最有效和准确的 YOLO 模型的地位。YOLOv11 的整体网络架构如图 1 所示。

YOLOv11 的损失函数包括 2 个分支:分类和回归分

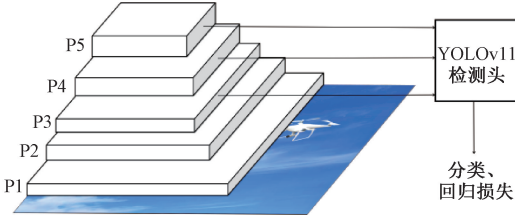


图1 YOLOv11 整体网络架构

Fig. 1 Overall network structure of YOLOv11

支,分类分支采用了二元交叉熵损失(binary cross-entropy loss, BCE Loss),回归分支使用了 DFL 损失(distribution focal loss)和 CIoU 损失(complete intersection over union loss)。

损失中, YOLOv11 检测结果其分类损失函数如式(1)所示,采用二元交叉熵损失,这里的 N 是样本数量, y_i 是第 i 个样本的真实标签, p_i 是模型预测的第 i 个样本为正类的概率。

$$BCE = -\frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \quad (1)$$

回归损失函数如式(2)所示,采用 DFL 损失。

$$DFL(S_i, S_{i+1}) = -((y_{i+1} - y) \log(S_i) + (y - y_i) \log(S_{i+1})) \quad (2)$$

其中, S_i 和 S_{i+1} 代表两个不同状态或时间点的模型输出或数据; y_i 和 y_{i+1} 表示在这两个状态或时间点的参数估计值。

CIoU 损失如式(3)所示。

$$Loss_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \quad (3)$$

其中, $\rho^2(b, b^{gt})$ 分别代表了预测框和真实框的中心点的欧式距离。 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。而 α 如式(4)所示, v 如式(5)所示, w, h 分别代表锚框的宽高。

$$\alpha = \frac{v}{1 - IoU + v} \quad (4)$$

$$v = \frac{4}{\pi^2} (\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h})^2 \quad (5)$$

最后将所有损失加权即可得到 YOLOv11 总损失。

1.2 高分辨率特征引导网络构建

本文提出了一种双通道网络结构,以解决低分辨率图像中小目标检测的难题。该网络结构包括两个通道:高分辨率目标检测网络和低分辨率目标检测网络。两者在网络架构上保持一致,区别在于高分辨率网络以高分辨率图像为输入,而低分辨率网络则使用由高分辨率图像多倍下采样生成的低分辨率图像。通过在相同中间特征层之间引入强约束(特征对齐)连接,实现了高低分辨率网络的联动,使低分辨率网络能够更有效地学习高分辨率网络的特征表达能力。受冉宁等^[22]改进 YOLOv8 时以小目标检测头替代大目标检测头的启发,本文为增大对小目标的感受野,强约束的中间层选择浅层特征进行优化。在训练过程中,这两个通道均基于改进的 YOLOv11 网络实现,并通过高分辨率网络对低分辨率网络的特征引导学习。网络的总损失函数由高分辨率网络和低分辨率网络的特征对齐损失以及目标检测损失按照一定权重比例加权求和得出。该设计显著提升了低分辨率图像中小目标的检测性能,整体的网络结构如图 2 所示。

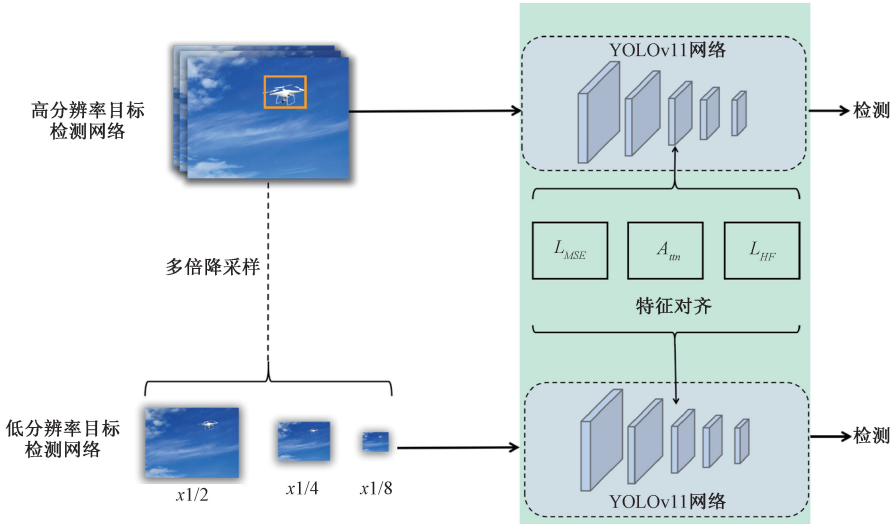


图2 双通道目标检测网络结构

Fig. 2 Dual-channel object detection network structure

1.3 基于 CAPC 的改进 YOLOv11 算法结构

受 Han 等^[23]改进 YOLOv5 的检测算法的启发,本文也在 YOLOv11 基础上进行改进。由于传统卷积方法对小

目标检测中的空间特征捕捉不足,且难以高效地处理不同尺度的目标。受近期提出的风车形卷积(partial convolution, PConv)^[24]和上下文聚合网络(Container)^[25]

思想的启发,提出了一种新的卷积模块,称为上下文聚合风车卷积。

PConv 采用非对称填充来为图像的不同区域创建水平和垂直卷积核,此方法可使 PConv 更好地符合微弱小目标的像素高斯空间分布,增强特征提取,显著增加感受野,且只引入最小的参数增加^[24]。受顾桂梅等^[26]引入坐标注意力机制的启发,本文引入了上下文聚合注意力机制。上下文聚合网络,一个用于多头上下文聚合的通用构建块,它可以利用上下文聚合网络的长期交互作用,同时利用局部卷积操作的诱导偏差,导致更快的收敛速度^[24]。

PConv 的本质是提升原始网络卷积对空间方向的感知,能够更好的学习小目标在不同特殊环境下的状态特

征,上下文聚合网络的本质是通过整合关联信息,让场景理解更完整,更好的联系小目标周围环境。

本文中,将两者结合,提出上下文聚合风车卷积,具体结合方式如下:将上下文聚合注意力机制加入 PConv 中,然后采用非对称填充来为图像的不同区域创建水平和垂直卷积核。这些卷积核向外扩散,其中 h_1 、 w_1 和 c_1 分别表示输入张量 $\mathbf{X}(h_1, w_1, c_1)$ 的高度、宽度和通道大小。为了增强训练的稳定性 and 速度,在每次卷积后应用批归一化 (BN) 和 SiLU 激活函数。PConv 的第 1 层执行并行卷积,具体改进后的 YOLOv11 整体网络如图 3 所示,将 YOLOv11 原本网络的部分卷积 (convolution, Conv) 替换为本文提出的 CAPC 模块。

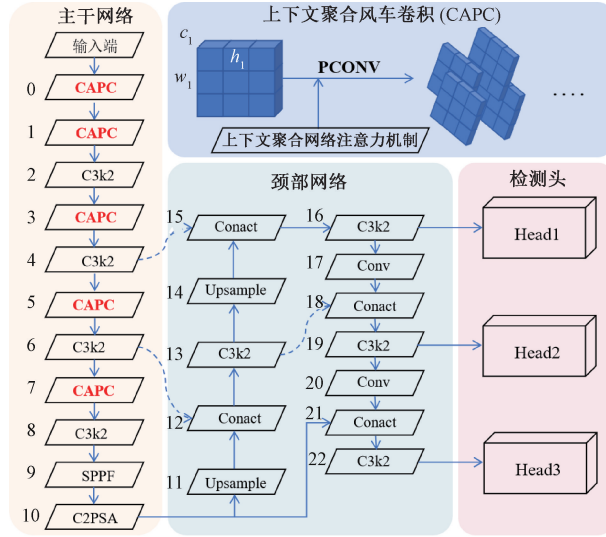


图 3 改进 YOLOv11 网络结构

Fig. 3 Improved YOLOv11 network structure

1.4 多尺度特征对齐损失函数

本文提出了一种新颖的损失函数:多尺度特征对齐损失。通过融合多维度约束机制实现特征图在不同分辨率下的精准对齐。该损失函数突破传统单一损失的局限性,创新性地将特征图均方误差损失、注意力加权损失与高频特征保持损失有机结合,形成层次化的优化体系,MSFAL 由以下 3 部分组成:

1) 特征图均方误差损失 (feature map MSE loss): 衡量高分辨率网络和低分辨率网络对应层特征图的整体差异,如式 (6) 所示。

$$L_{MSE} = \frac{1}{N} \sum_{i=1}^N \|F_{HR}^{(i)} - F_{LR}^{(i)}\|_2^2 \quad (6)$$

其中, $F_{HR}^{(i)}$ 和 $F_{LR}^{(i)}$ 分别表示高分辨率和低分辨率网络的第 i 层特征图,这些特征图是卷积神经网络提取的中间层输出,蕴含了图像的语义、纹理和结构信息。 N 是特征图总数。 $\|\cdot\|_2^2$ 为 L_2 范数的平方,即对特征图中每个对应位置的元素差值进行平方求和,确保所有差异均以非负形式累加,避免正负误差相互抵消,从而精确反映特征图间

的整体差异。

2) 注意力加权损失 (attention-weighted loss): 核心在于动态权重分配,通过注意力权重生成函数 $A(\cdot)$ 计算特征图中每个区域的重要性,为小目标区域赋予更高权重,使得模型在计算损失时能优先关注并优化这些易被忽视的部分。如式 (7) 所示。

$$L_{Attm} = \frac{1}{N} \sum_{i=1}^N \|A(F_{HR}^{(i)}) \odot (F_{HR}^{(i)} - F_{LR}^{(i)})\|_1 \quad (7)$$

其中, $F_{HR}^{(i)}$ 、 $F_{LR}^{(i)}$ 和分别代表高分辨率网络和低分辨率网络的第 i 层特征图; $A(\cdot)$ 是注意力权重生成函数; $\odot$ 表示逐元素乘积, $\|\cdot\|_1$ 为 L_1 范数。

3) 高频特征保持损失 (high-frequency preservation loss): 当低分辨率特征图向高分辨率对齐时,传统损失函数易导致边缘模糊、纹理丢失等问题,而高频特征保持损失通过拉普拉斯滤波精准捕捉图像中的高频信息,确保特征图在跨分辨率转换过程中,边缘、纹理等关键细节得以完整保留,为模型输出的清晰度与真实感提供保障。该损失通过最小化高频成分差异,强制模型在特征对齐时优先

保留细节信息,如式(8)所示。

$$L_{HF} = \frac{1}{N} \sum_{i=1}^N \|L(F_{HR}^{(i)}) - L(F_{LR}^{(i)})\|_2^2 \quad (8)$$

其中, $F_{HR}^{(i)}$, $F_{LR}^{(i)}$ 和分别代表高分辨率网络和低分辨率网络的第 i 层特征图; $L(\cdot)$ 为拉普拉斯滤波操作,用于提取特征图中的高频成分; $\|\cdot\|_2^2$ 是 L_2 范数的平方,用于量化高低分辨率特征图高频成分的差异; N 为参与计算的特征图总数。

综合以上 3 部分,最终的特征对齐损失函数如式(9)所示。

$$L_{Align} = \lambda_1 L_{MSE} + \lambda_2 L_{Attn} + \lambda_3 L_{HF} \quad (9)$$

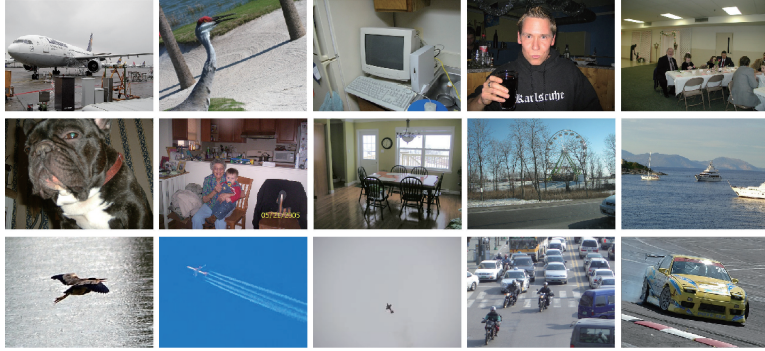


图4 VOC 2012 数据集(部分)

Fig. 4 VOC 2012 dataset(partial)

针对低分辨率图像的小目标检测任务,传统的高分辨率目标检测方法往往难以实现高性能,因此将数据集原始尺寸多尺度降低至原图的 $1/2$ 、 $1/4$ 和 $1/8$,使之适应更为挑战性的检测环境。

同时,为验证该方法在复杂环境下的泛化能力和现实场景中地分辨率目标的检测能力,选择 Visdrone2019 数据集开展小目标检测方法验证。如图 5 所示,该数据集采集自真实

其中, λ_1 、 λ_2 、 λ_3 是权重超参数,用于平衡不同损失项的影响,经过多次实验将其参数设置为 $1.0:0.1:0.5$ 效果最佳。

2 实验与结果分析

2.1 数据集

为充分验证该小目标检测方法的性能,采用 PASCAL VOC 2012 数据集^[27]进行模型训练及测试。如图 4 所示,该数据集包括 20 个前景类别和 1 个背景类别。该数据集分为 3 个部分:训练集(1 464 张图像)、验证集(1 449 张图像)和测试集(1 456 张图像)。

交通监控场景,图像中密集分布着尺度差异显著的目标(如远处车辆与近景行人),尤其包含大量像素占比低于 3% 的小目标实例,这种贴近实际应用场景的复杂性,能有效检验模型在拥挤场景与小目标分辨率下的检测鲁棒性。其划分的训练集(6 471 张)、验证集(548 张)与测试集(7 518 张),不仅覆盖晴天、雨天等多元天气条件,还包含不同拍摄视角的动态场景,为模型在复杂环境下的泛化能力提供了系统性的验证基准。

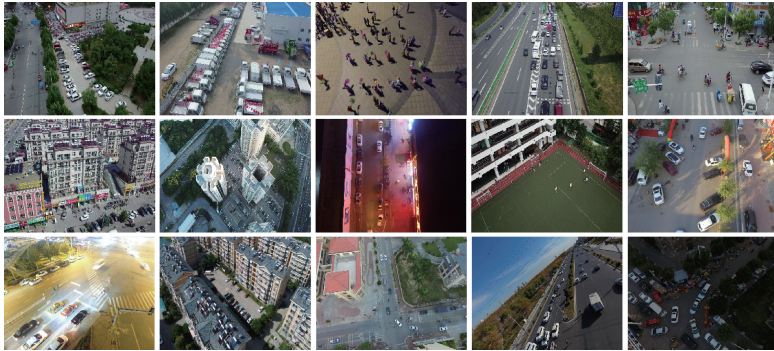


图5 Visdrone 2019 数据集(部分)

Fig. 5 Visdrone 2019 dataset(partial)

2.2 实验环境及实验步骤

本文所有的实验环境采用 Win11 系统,具体实验环境如表 1 所示。

为了验证双流网络主干、基于 CAPC 的改进、特征对齐损失对模型检测性能高低的影响,本文实验流程分为以

下 5 个步骤:

1)准备 VOC2012、Visdrone2019 数据集及其经过多倍降采样得到的数据集 VOC2012_down, Visdrone2019_down。

2)按照先后顺序对模块进行改进:首先,为保证训练

表 1 实验环境配置信息

Table 1 Experimental environment configuration information

训练环境	环境配置
CPU	i9-13900KF
GPU	RTX 4090(24 G)
深度学习框架	Pytorch2.1.0
CUDA	CUDA12.1
集成开发环境	Pycharm

时数据的公平性,将步骤 1)中的原始数据集和其降采样数据集做为整体用 YOLOv11n 进行训练;接着将基于 CAPC 改进的 YOLOv11n 进行训练;然后将没有改进卷积的 YOLOv11n 双通道网络进行训练;最后将本文方法,即增加双流网络并用特征对齐损失约束模型中间层,数据集分

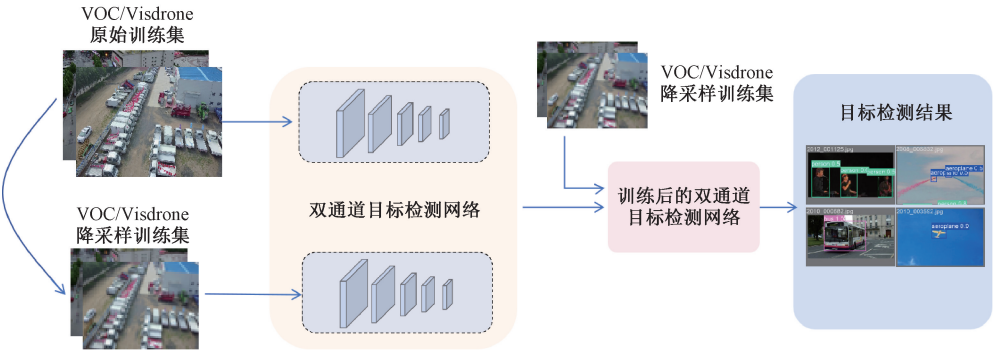


图 6 双通道目标检测网络训练测试流程

Fig. 6 Dual-channel object detection network training and testing

2.3 消融实验

本文研究通过引入改进的 CAPC 模块、在特征对齐损失中加入不同的损失函数(L_{MSE} 、 L_{Attn} 、 L_{HF})和双通道目标检测网络方法,逐步优化 YOLOv11n 模型,并评估其在小目标检测中的表现,如表 2、3 所示。以在 VOC1012 数据集上的消融实验结果为例,加入 CAPC 模块后,YOLOv11n 网络检测能力 mAP 提升约 1.2%。随后增加双通道网络,特征对齐损失仅

为原始和降采样分别放入高分辨率目标检测网络和低分辨率目标检测网络中进行训练。

3)设置学习率为 0.01,Epoch 为 300,Batchsize 为 32,Patience 为 50,先后对 YOLOv11、增加改进的上下文聚合风车卷积 YOLOv11 网络、高分辨率特征引导的双通道网络进行消融实验,训练数据集集 loss 达到最大迭代次数或没有明显改善以提前停止训练的时期,得到最佳的训练输出模型。

4)利用最佳训练输出模型对验证集及测试集进行测试,对不同框架同种场景下检测性能及效果进行讨论。

5)评价模型从验证集及测试集中获得性能指标,得到多种类别的平均精确度指标(mAP50 及 mAP50:95)、大中小目标预测平均精确度指标 mAP50。具体双通道目标检测网络训练测试流程如图 6 所示。

使用 L_{MSE} ,mAP 再次增加约 1%。分别加入 L_{Attn} 、 L_{HF} ,两种方式对整体网络效果提升有相近的效果。最终,本方法使用双通道网络和完整特征对齐损失函数,使得整体 mAP 和小目标检测精度达到了最佳效果。相比于原始 YOLOv11n,改进后的模型在低分辨率小目标检测中 mAP50%提升约 1%,整体检测能力 mAP50%提升约 4.1%,充分展现了该方法在提升精度和鲁棒性方面的显著优势。

表 2 在 VOC1012 数据集上的消融实验结果

Table 2 Ablation experiment results on the VOC1012 dataset

网络结构	参数/ 10^6	计算量/GFLOPs	mAP50/%	mAP50:95/%	小目标 mAP50/%	检测速度/ms
YOLOv11n	2.593 740	3.23	61.663	44.862	16.73	6.65
YOLOv11n+CAPC	2.615 484	3.45	62.915	46.620	15.89	10.26
双通道(L_{MSE})	2.645 484	3.55	63.795	47.970	16.23	10.38
双通道($L_{MSE} + L_{Attn}$)	2.645 484	3.55	64.964	48.250	16.94	10.55
双通道($L_{MSE} + L_{HF}$)	2.645 484	3.55	64.875	48.030	16.83	10.49
双通道(MSFAL)	2.645 484	3.55	65.776	49.114	17.50	10.82

表 2、3 中,在评价模型性能指标时,本次实验选取准确率 P (定义为模型识别图片中正样本被正确预测的图片占比)、召回率 R (指被正确识别的图片在测试集全

部图片中的占比)、平均精度均值 mAP50(当 IoU 阈值设为 0.5 时,所有类别平均精度 AP 的均值)、检测速度(模型每秒可处理的最大图片数量)、参数量大小及权重

表 3 在 Visdrone2019 数据集上的消融实验结果

Table 3 Ablation experiment results on the Visdrone2019 dataset

网络结构	参数/ 10^6	计算量/GFLOPs	mAP50/%	mAP50:95/%	小目标 mAP50/%	检测速度/ms
YOLOv11n	2.591 790	3.23	30.994	17.691	12.91	5.99
YOLOv11n+CAPC	2.613 534	3.45	31.224	17.972	12.64	10.19
双通道(L_{MSE})	2.643 534	3.55	32.034	18.173	13.01	10.30
双通道($L_{MSE}+L_{Attn}$)	2.643 534	3.55	32.263	18.675	13.22	10.43
双通道($L_{MSE}+L_{HF}$)	2.643 534	3.55	32.337	18.711	13.27	10.46
双通道(MSFAL)	2.643 534	3.55	33.231	19.188	13.84	10.69

大小作为评价指标。各评价指标的计算式如式(10)~(13)所示。

$$P = \frac{TP}{TP + FP} \tag{10}$$

$$R = \frac{TP}{TP + FN} \tag{11}$$

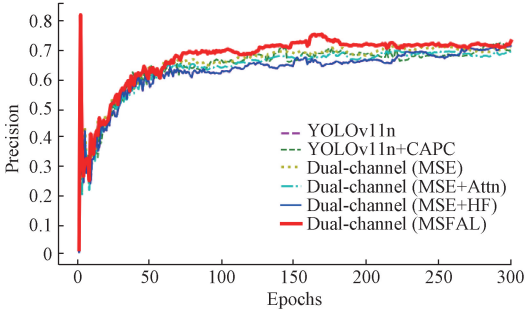
$$AP = \int_0^1 P(t) dt \tag{12}$$

$$mAP = \frac{\sum_{n=1}^N AP_n}{N} \tag{13}$$

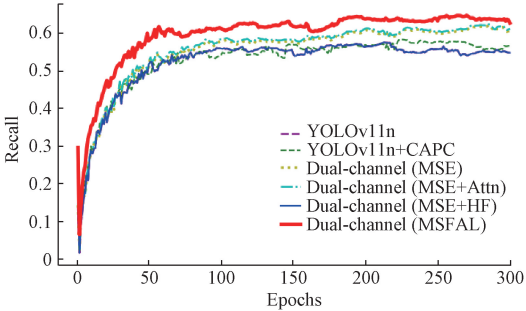
其中, TP 为模型正确预测为鸟类的数量(真正例);

FP 为模型将背景误判为鸟类的数量(假正例); FN 为模型将鸟类误判为背景的数量(假反例); AP 为平均精度, 其中 $P(t)$ 表示交并比阈值为 t 时的精确率; N 为目标检测任务中的目标类别数量, 本次实验仅包含鸟类单一类别, 故 $N=1$; AP_n 为第 n 个类别的平均精度。

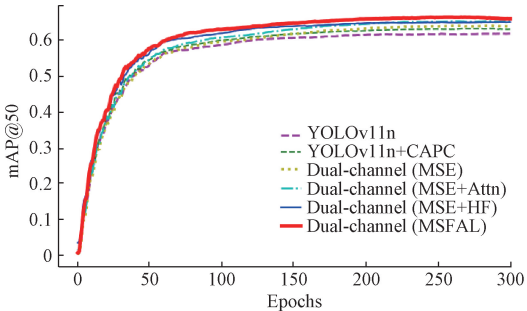
在 VOC2012 数据集、Visdrone2019 数据集下, 6 项消融实验的训练过程预测回归精度、mAP 变化分别如图 7、8 所示。在前 100 轮训练中, 模型整体曲线震荡较为明显, 100 轮后逐渐趋于平缓稳定。在训练后期, 双通道网络模型 mAP 一直高于其对比模型, 模型训练过程中没有出现明显的过拟合和欠拟合现象。



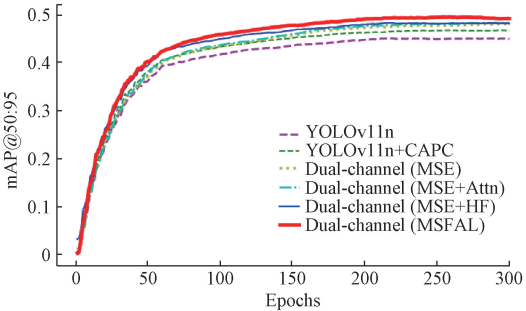
(a) Relationship curve between training iterations and prediction accuracy



(b) Relationship curve between training iterations and recall rate



(c) Relationship curve between training iterations and mAP@50



(d) Relationship curve between training iterations and mAP@50:95

图 7 VOC2012 数据集消融实验训练结果对比

Fig. 7 Comparison of ablation experiment training results on the VOC2012 dataset

2.4 对比试验

将双通道目标检测网络与其他目标检测模型进行对

比试验, 结果如表 4、5 所示。改进 YOLOv11n 的双通道目标检测网络在所有模型中的 mAP50 最高, 为 65.8%。此

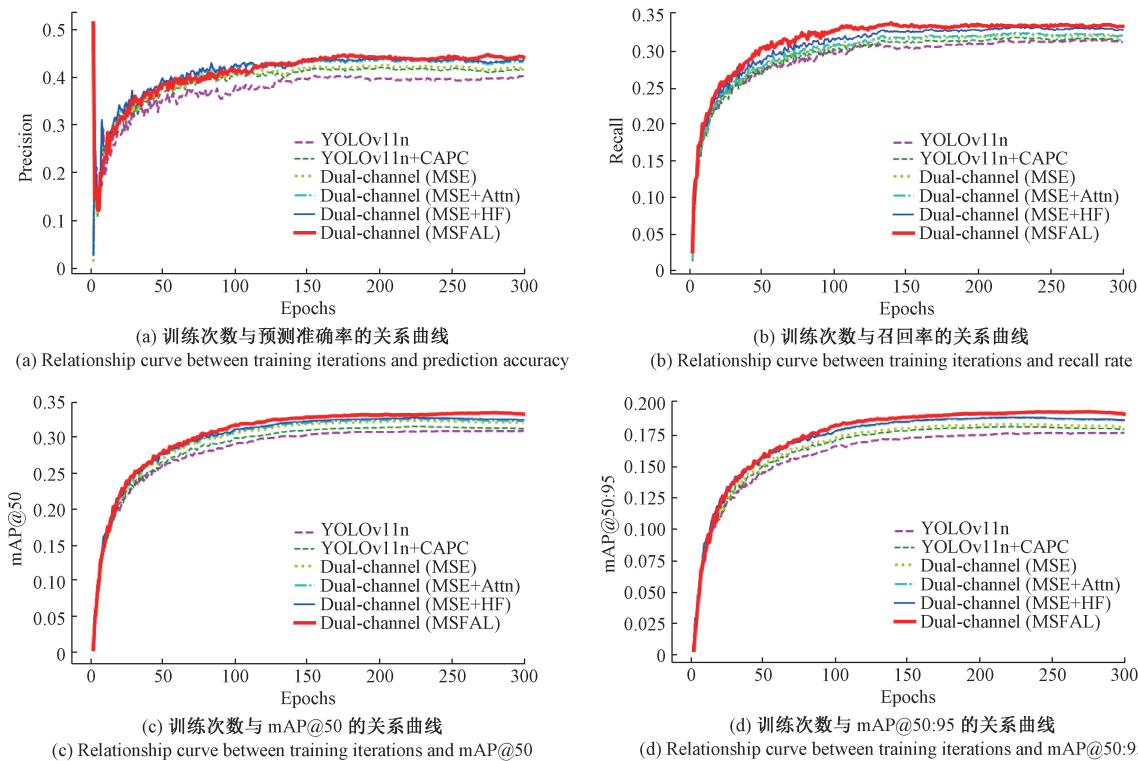


图 8 Visdrone2019 数据集消融实验训练结果对比

Fig. 8 Comparison of ablation experiment training results on the Visdrone2019 dataset

外,本实验还计算了相应模型 mAP50:95 和每一种模型对小目标检测的 mAP50,其中本文方法对低分辨率下的小目标检测性能最好,mAP50 达 17.5%。检测速度最快的是 YOLOv5n,单张图片完成训练流程的时间为 5.71 ms,但其 mAP50 值比本文低约 4.6%,小目标检测 mAP50 值低了 3%。相比于其他模型的参数量,本文双通道目标检测网络也取得不错结果,控制在 2.6×10^6 左右。此外,本文还与最新的相关研究进行了对比,实验结果显示,相较于部分最新研究,本方法在不依赖大量数据和计算量的情况

下,仍在小目标检测精度上取得了少量提升。这一成果主要得益于高分辨率特征引导学习在小目标检测中的先进性以及本文数据集的特殊性。在实验中,最新的研究方法在包含多倍下采样的低分辨率小目标数据集上表现并不突出,而高分辨率特征引导学习成功解决了这一问题,显著提升了低分辨率小目标的检测效果。

综上,基于高分辨率特征引导学习的小目标检测方法在各模型中的性能表现维持在较高水平,该模型能够较好的运用在低分辨率下的小目标检测系统中。

表 4 VOC2012 对比实验结果

Table 4 Comparative experimental results on the VOC2012 dataset

网络结构	参数/ 10^6	计算量/GFLOPs	mAP50/%	mAP50:95/%	小目标 mAP50/%	检测速度/ms
YOLOv8n	3.011 043	8.72	63.345	41.123	13.81	8.70
YOLOv5n	1.973 569	4.45	63.197	41.090	13.50	5.71
YOLOv7-tiny	6.014 988	13.21	60.23	39.901	13.89	7.23
Faster R-CNN	42.304 43	180.42	58.789	32.012	12.41	35.64
SSD	26.013 23	95.31	59.567	32.890	12.78	40.12
CenterNet	34.602 32	60.43	52.123	34.345	12.12	12.34
PE-YOLO ^[28]	30.869 35	110.78	61.345	47.234	16.95	12.82
DMF-YOLOv11 ^[29]	3.196 490	5.82	62.981	47.910	16.61	8.63
Relation-DETR ^[30]	36.747 97	97.94	65.475	48.963	17.24	14.32
本文	2.615 484	3.55	65.776	49.114	17.50	10.42

表 5 Visdrone2019 对比实验结果

Table 5 Comparative experimental results on the Visdrone2019 dataset

网络结构	参数/10 ⁶	计算量/GFLOPs	mAP50/%	mAP50:95/%	小目标 mAP50/%	检测速度/ms
YOLOv8n	3.011 043	8.72	30.694	17.491	12.41	6.04
YOLOv5n	1.843 835	4.43	30.525	17.259	12.26	9.02
YOLOv7-tiny	6.014 988	13.52	30.680	17.321	12.39	8.75
Faster R-CNN	42.014 42	180.10	28.342	16.420	12.24	32.06
SSD	26.023 41	95.42	28.949	17.009	12.51	28.44
CenterNet	34.602 71	90.55	26.315	16.003	11.89	31.23
PE-YOLO ^[28]	30.069 34	110.42	31.134	18.21	12.95	12.53
DMF-YOLOv11 ^[29]	3.196 490	546.8	30.355	18.02	12.43	5.83
Relation-DETR ^[30]	37.747 97	120.23	33.021	18.994	13.50	12.66
本文	2.613 534	3.55	33.231	19.188	13.84	10.69

同时实验对 VOC2012 和 Visdrone2019 数据集检测结果做了可视化分析,结果如图 9、10 所示,其中图 9(a)、图 10(a)为 YOLOv11n 预测结果(部分)可视化,图 9(b)、图 10(b)为本研究方法的预测可视化。



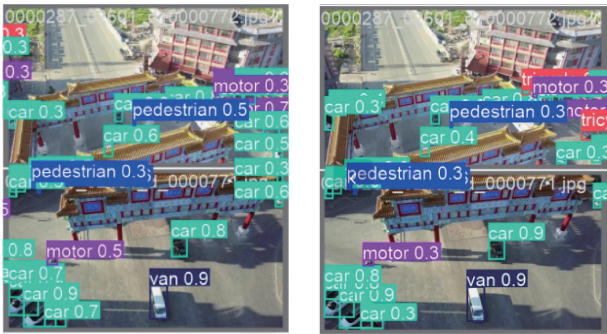
(a) YOLOv11n 预测结果(部分)
(a) YOLOv11n prediction results(partial)



(b) 双通道目标检测网络预测结果(部分)
(b) Dual-channel object detection network prediction results(partial)

图 9 VOC 2012 数据集检测结果可视化

Fig. 9 Visualization of detection results on the VOC 2012 dataset



(a) YOLOv11n 预测结果(部分)
(a) YOLOv11n prediction results (partial)



(b) 双通道目标检测网络预测结果(部分)
(b) Dual-channel object detection network prediction results(partial)

图 10 Visdrone 2019 数据集检测结果可视化

Fig. 10 Visualization of detection results on the Visdrone 2019 dataset

3 结 论

本文针对低分辨率图像下的小目标检测,提出了一种基于高分辨率特征引导学习的小目标检测方法。该框架包括加入上下文聚合风车卷积的双通道 YOLOv11 网络和双通道中间层引入的多尺度特征对齐损失,以解决低分辨率下小目标检测无教师网络引导,上下文信息不足的问题,同时将原始数据集进行下采样作为检测目标,提高了网络对低分辨率下小目标的适应能力。实验结果显示,本文的算法在低分辨率图像的小目标检测任务中表现显著优于 YOLOv11 基准网络,不仅将小目标检测的平均精度提升了 1%,更充分证明了所提算法在该场景下的有效性与实用性。

参考文献

- [1] 秦学斌,薛宇强,景宁波,等. 露天煤矿自动驾驶矿卡前障碍物检测算法研究[J]. 金属矿山,2025(2):145-151.
QIN X B,XUE Y Q,JIN N B,et al. Study on obstacle detection algorithm in front of autonomous mining truck in open-pit coal mine[J]. Metal Mine,2025(2): 145-151.
- [2] LIU Y,SUN P,WERGELES N,et al. A survey and performance evaluation of deep learning methods for small object detection [J]. Expert Systems with Applications,2021,172:114602.
- [3] 江松,孔若男,李鹏程,等. 融合 Swin Transformer 与 CNN 的露天矿车前障碍物智能检测算法[J]. 金属矿山,2023(5):228-236.
JIANG S,KONG R N,LI P CH,et al. Intelligent detection algorithm of obstacles in front of open-pit mine cars based on swin transformer and CNN[J]. Metal Mine,2023(5):228-236.
- [4] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]. 2017 Conference, Advances in Neural Information Processing Systems 30,2017:5999-6009.
- [5] TU X G, HE ZH, HUANG Y, et al. An overview of large AI models and their applications [J]. Visual Intelligence, 2024, 2(1): 34.
- [6] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [7] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [C]. 2015 Conference, Advances in Neural Information Processing Systems 28,2015:91-99.
- [8] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]. Computer Vision-ECCV 2016;14th European Conference,2016:21-37.
- [9] REDMON J,DIVVALA S,GIRSHICK R,et al. You only look once:Unified,real-time object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition,2016:779-788.
- [10] LAW H, DENG J. Cornernet: Detecting objects as paired keypoints [C]. European Conference on Computer Vision(ECCV),2018:734-750.
- [11] DUAN K, BAI S, XIE L, et al. Centernet: Keypoint triplets for object detection [C]. IEEE/CVF International Conference on Computer Vision, 2019: 6569-6578.
- [12] 安梦军,吴朝明,邓承志. MEC-YOLOv11n:水面小目标漂浮物的检测算法[J]. 电子测量技术,2025, 48(14):185-197.
AN M J, WU ZH M, DENG CH ZH. MEC-YOLOv11n:Detection algorithm for floating objects of small targets on the water surface [J]. Electronic Measurement Technology,2025,48(14):185-197.
- [13] 王浩,张其猛,龚德成. 复杂场景下的交通标志小目标检测算法[J]. 电子测量技术,2025,48(2):158-169.
WANG H, ZHANG Q M, GONG D CH. Small target detection algorithm for traffic signs in complex scenes[J]. Electronic Measurement Technology,2025,48(2):158-169.
- [14] 刘洋,任旭虎,刘宝弟,等. 基于 LDF-YOLO 的小目标检测方法[J]. 电子测量技术,2025,48(12):156-165.
LIU Y, REN X H, LIU B D, et al. Small object detection method based on LDF-YOLO[J]. Electronic Measurement Technology,2025,48(12):156-165.
- [15] 王迎龙,孙备,丁冰,等. BG-YOLO:复杂大视场下低慢小无人机目标检测方法[J]. 仪器仪表学报,2025, 46(2):255-266.
WANG Y L, SUN B, DING B, et al. BG-YOLO: A low-altitude slow-moving small UAV targets detection method in complex large field of view[J]. Chinese Journal of Scientific Instrument,2025,46(2):255-266.
- [16] 张莹,邓华宣,王耀南,等. 基于多通道特征融合学习的印制电路板小目标缺陷检测[J]. 仪器仪表学报, 2024,45(5):10-19.
ZHANG Y,DENG H X,WANG Y N,et al. Small defects detection of PCB based on multi-channel feature fusion learning [J]. Chinese Journal of Scientific Instrument,2024,45(5):10-19.
- [17] ZHANG H, ZHANG H, MEI A, et al. SO-DETR: Leveraging dual-domain features and knowledge distillation for small object detection [J]. ArXiv preprint arXiv:2504.11470,2025.
- [18] MA Y C, MA X, HAO T R, et al. Knowledge distillation via hierarchical matching for small object detection [J]. Journal of Computer Science and Technology,2024,39(4):798-810.
- [19] ZHU Y, ZHOU Q, LIU N, et al. Scalekd: Distilling scale-aware knowledge in small object detector[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition,2023:19723-19733.
- [20] JOBAER S, TANG X, ZHANG Y, et al. A novel knowledge distillation framework for enhancing small object detection in blurry environments with unmanned aerial vehicle-assisted images [J]. Complex &

- Intelligent Systems, 2025, 11(1): 63.
- [21] 向雷, 蒋文波. 改进 YOLOv8 的城市行车道路障碍物检测算法研究[J]. 电子测量技术, 2025, 48(1): 29-38. XIANG L, JIANG W B. Research on improved YOLOv8 urban driving road obstacle detection algorithm[J]. Electronic Measurement Technology, 2025, 48(1): 29-38.
- [22] 冉宁, 施高朗, 张少康, 等. 基于 YOLOv8 的遥感小目标检测算法[J]. 电子测量与仪器学报, 2025, 39(5): 197-207. RAN N, SHI G L, ZHANG SH K, et al. Remote sensing small target detection algorithm based on YOLOv8[J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(5): 197-207.
- [23] HAN X, ZHOU X. Two-wheeler helmet wearing detection algorithm based on improved YOLOv5[J]. Instrumentation, 2025, 12(1): 11-29.
- [24] YANG J, LIU S, WU J, et al. Pinwheel-shaped convolution and scale-based dynamic loss for infrared small target detection [C]. AAAI Conference on Artificial Intelligence, 2025, 39(9): 9202-9210.
- [25] GAO P, LU J, LI H, et al. Container: Context aggregation network [J]. ArXiv preprint arXiv: 2106.01401, 2021.
- [26] 顾桂梅, 王小亮. 深度学习的接触网小目标缺陷识别研究[J]. 电子测量与仪器学报, 2024, 38(4): 151-160. GU G M, WANG X L. Identification of the catenary small target defects in deep learning[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(4): 151-160.
- [27] EVERINGHAM M, ALI ESLAMI S M, GOOL L, et al. The pascal visual object classes challenge: A retrospective[J]. International Journal of Computer Vision, 2015, 111(1): 98-136.
- [28] 景婷婷, 曹玉东, 陈鑫, 等. 改进 YOLOv11 的无人机航拍小目标检测算法[J]. 计算机工程与应用, 2026, 62(2): 138-148. JING T T, CAO Y D, CHEN X, et al. Improved YOLOv11 algorithm for aerial small target detection in UAVs [J]. Computer Engineering and Applications, 2026, 62(2): 138-148.
- [29] 贺智轩, 陈里里, 王翔, 等. DMF-YOLOv11: 基于改进 YOLOv11n 的无人机航拍图像目标检测算法[J]. 计算机工程与应用, 2025, 61(14): 88-100. HE ZH X, CHEN L L, WANG X, et al. DMF-YOLOv11: Target detection algorithm for UAV images based on improved YOLOv11n[J]. Computer Engineering and Applications, 2025, 61(14): 88-100.
- [30] HOU X, LIU M, ZHANG S, et al. Relation detr: Exploring explicit position relation prior for object detection [C]. European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2024: 89-105.

作者简介

涂晓光(通信作者), 博士, 副教授, 主要研究方向为人工智能、图像处理。

E-mail: xguangtu@outlook.com

李卓骏, 硕士研究生, 主要研究方向为人工智能、图像处理。

E-mail: a215715001@outlook.com

刘建华, 博士, 副教授, 主要研究方向为信息安全、云计算、图像处理。

E-mail: ljh2583265@163.com

杨明, 博士, 主要研究方向为室温红外技术及应用, 包括红外无损检测和目标检测。

E-mail: yangming932@163.com

魏麟, 硕士, 主要研究方向为航空电子、民机智能驾驶。

E-mail: weilin@cafuc.edu.cn