

基于SODM-YOLOv8的无人机图像小目标检测算法

朱志鹏¹, 王 玫²

(1. 桂林理工大学 计算机科学与工程学院; 2. 桂林理工大学 物理与电子信息工程学院, 广西 桂林 541004)

摘要: 针对无人机航拍图像目标检测算法精度较低且模型相对复杂的问题, 提出小目标检测算法SODM-YOLOv8。首先, 设计小目标检测模块, 通过大尺寸特征图来检测小目标, 保留了小目标的细节特征, 提高了对小目标的关注度; 其次, 设计DC-C2f模块, 通过双流卷积结构对原有卷积进行改进, 以降低模型参数量; 再次, 在小目标检测模块中引入高效多尺度注意力机制EMA, 提升算法对多尺度特征的融合能力; 最后, 引入Inner-WIoU损失函数, 即采用Inner-IoU的思想改进WIoU损失函数, 以提高模型的泛化能力。在VisDrone2019数据集上的实验结果表明, SODM-YOLOv8算法的mAP@0.5达到43.8%, 相较其他目标检测算法均有一定提升, 其中比YOLOv8模型的mAP@0.5提高了5.2%, 参数量下降了28.3%。在DOTA1.0数据集上的实验结果表明, 所提算法的mAP@0.5达到72.3%, 比YOLOv8提高了1.4%。

关键词: 小目标检测; 无人机; YOLOv8; 特征融合; 注意力机制

DOI: 10.11907/jjdk.241539

开放科学(资源服务)标识码(OSID):

中图分类号: TP391

文献标识码: A

文章编号: 1672-7800(2025)008-0164-09



Unmanned Aerial Vehicle Image Small Target Detection Algorithm Based on SODM-YOLOv8

ZHU Zhipeng¹, WANG Mei²

(1. College of Computer Science and Engineering, Guilin University of Technology;

2. College of Physics and Electronic Information Engineering, Guilin University of Technology, Guilin 541004, China)

Abstract: To address the issue of low accuracy and relatively complex models in object detection algorithms for drone aerial images, the small object detection algorithm SODM-YOLOv8 is proposed. Firstly, design a small target detection module that detects small targets through large-sized feature maps, preserving the detailed features of small targets and increasing attention to them; Secondly, design the DC-C2f module to improve the original convolution through a dual stream convolution structure, in order to reduce the number of model parameters; Once again, an efficient multi-scale attention mechanism EMA is introduced in the small object detection module to enhance the algorithm's ability to fuse multi-scale features; Finally, the Inner IoU loss function is introduced, which adopts the idea of Inner IoU to improve the WIoU loss function and enhance the model's generalization ability. The experimental results on the VisDrone2019 dataset show that the SODM-YOLOv8 algorithm mAP@0.5 Reached 43.8%, showing a certain improvement compared to other object detection algorithms, among which the YOLOv8 model outperforms mAP@0.5 Improved by 5.2%, while the parameter count decreased by 28.3%. The experimental results on the DOTA1.0 dataset show that the proposed algorithm mAP@0.5 Reached 72.3%, an increase of 1.4% compared to YOLOv8.

Key Words: small object detection; unmanned aerial vehicle; YOLOv8; feature fusion; attention mechanism

0 引言

无人机是无需人员驾驶的飞行器, 能够远程控制或自

行执行飞行任务, 具有机动性强、视野开阔的优势。随着深度学习技术的快速发展, 无人机与目标检测任务相结合已成为重要的发展趋势, 极大提升了工作效率与作业精度, 在安全防护、无人驾驶、工业制造、道路交通等领域发

收稿日期: 2024-06-18

扫描二维码阅读全文:

基金项目: 国家自然科学基金项目(62071135)

作者简介: 朱志鹏(1998-), 男, 桂林理工大学计算机科学与工程学院硕士研究生, 研究方向为目标检测; 王玫(1963-), 女, 博士, 桂林理工大学物理与电子信息工程学院教授、博士生导师, 研究方向为多媒体信息感知与处理。本文通讯作者: 朱志鹏。



挥着越来越重要的作用^[1-4]。

目前,基于深度学习的目标检测算法可分为一阶段和二阶段两种。一阶段目标检测算法包括YOLO(You Only Look Once)系列^[5-9]、SSD(Single Shot Multibox Detector)系列算法^[10]、RetinaNet算法^[11]以及Cornernet算法^[12]等。该类算法仅需一次前向传播即可获得目标类别和位置信息,能够在检测精度与速度之间取得很好的平衡。二阶段目标检测算法包括R-CNN(Region-based Convolutional Neural Network)^[13]、Fast R-CNN^[14]、Faster R-CNN^[15]、Mask R-CNN^[16]等。该类算法首先使用区域提议技术生成目标潜在区域,再进行定位与分类,虽然检测精度有所提高,但实时性较差。

YOLO系列算法的核心优势在于其实时性和效率,尤其适合需要快速响应的场景。该系列模型通过持续迭代与模型优化,精度在不断提高,同时保持了速度优势。近年来,国内外学者对YOLO系列目标检测算法进行了大量研究,从轻量化、多尺度学习、注意力机制、数据增强等多个角度对其进行了改进。例如,于傲泽等^[17]设计了分块复合注意力模块(Patch-Wise Co-Attention, PWCA),通过特征融合与注意力机制相结合,有效提高了检测精度,但是对计算资源要求较高,并且实时性较差。Wang等^[18]通过优化YOLOv5提出轻量化检测模型MFP-YOLO,该模型设计了倒置残差块,引入了卷积块注意力模块(Convolutional Block Attention Module, CBAM),提高了复杂环境下的检测效果和实时性,但是检测精度未能得到较好提升。Wang等^[19]通过引入BiFormer注意力机制来优化YOLOv8骨干网络,提高了模型对关键信息的提取能力;同时设计FFNB(Focal FasterNet block)特征处理模块,使浅层网络与深层网络得到充分融合,降低了漏检率。然而改进模型训练时间较长,不适用于实时任务。Du等^[20]从轻量化模型的角度设计了稀疏卷积来优化检测头,减少了计算量,但在检测精度方面存在不足。Min等^[21]将全局残差和局部融合机制整合到YOLO-DCTI模型中,融合了空间上下文信息和通道特征,能有效处理复杂背景的目标检测任务,具有较好的鲁棒性,但模型较为复杂,需要消耗更多计算资源。Safaldin等^[22]通过引入Bi-FPN模块提高YOLOv8模型的多尺度特征融合能力,改善了小目标的误检和漏检问题;同时设计Ghostblock模块优化骨干网络,用于捕捉远程特征信息,提高模型的泛化能力,但检测精度仍需提高。Cai等^[23]通过引入GSConv降低YOLOv8模型的参数量,设计GD(Gather-and-Distribute)模块融合全局多尺度特征,并将全局信息分布到深层网络,在不增加时延的情况下提高了模型的信息融合能力,但是计算复杂度较高。Wang等^[24]采用结构重参数技术改进YOLOv8的骨干网络,将定位能力差的层重新输入到网络中,以确保在特征提取过程中全面保留上下文信息,但是增加了额外的资源消耗。

虽然现有研究通过加强YOLO系列模型的特征融合

能力来获取小目标的上下文信息,但是多次下采样会导致微小特征逐渐丢失,在复杂环境下仍然难以精确定位小目标的位置。无人机设备拍摄的小目标尺寸很小,且分辨率也较低,在数据集上呈现密集,导致标注面积小,更加难以准确定位目标区域。无人机目标检测模型的检测速度在60帧/s以上时可被认为具有实时性,现有方法虽然在检测精度上有所提升,但检测速度却有所降低,没有在检测精度与速度之间达到很好的平衡。

为此,本文以YOLOv8为基线模型,提出无人机图像小目标检测算法SODM-YOLOv8(Small Object Detection Module-YOLOv8),具体贡献点为:①考虑到YOLOv8在下采样后会丢失小目标的特征信息这一问题,设计一个额外的P2小目标检测模块;②考虑到复杂环境下YOLOv8出现的漏检和误检问题,引入高效多尺度注意力(Efficient Multi-scale Attention, EMA)机制;③考虑到YOLOv8模型相对复杂,在颈部网络设计DC-C2f模块替换原有的C2f模块;④考虑到小目标的边界框重叠程度较高以及低质量样本对检测精度的不良影响,结合Inner-IoU思想和WIoU损失函数对CIoU损失函数进行改进。

1 YOLOv8模型概述

如图1所示,YOLOv8模型结构包含输入端(Input)、骨干网络(Backbone)、颈部网络(Neck)和预测头(Head)4个部分。其中,输入端采用Mosaic数据增强处理输入图像来提高模型的泛化能力;骨干网络通过快速空间金字塔池化模块(Spatial Pyramid Pooling-Fast, SPPF)来提高检测速度,通过C2f模块来提高特征提取能力;颈部网络采用特征金字塔网络和路径聚合网络来提高多尺度特征融合能力;预测头通过解耦头(Decoupled Head)的方式将分类头与检测头分离,通过分类损失函数对目标进行分类,同时借助回归损失函数来定位目标位置。

2 SODM-YOLOv8模型

如图2所示,SODM-YOLOv8模型首先改进了特征融合方式,通过残差连接的方式提高了多尺度特征融合能力;其次,设计小目标检测模块,将160×160×128尺寸的大特征图残差连接到颈部网络进行拼接,同时结合EMA机制来捕捉小目标的空间信息和通道信息,进一步提高模型对小目标的定位与预测能力;再次,在颈部网络设计DC-C2f模块,通过组卷积的方式排列卷积核,降低了模型参数量;最后,采用Inner-IoU的思想改进WIoU损失函数,用于替换原有的CIoU损失函数,同时在边界框内部增设额外的辅助框,赋予低质量样本较小的梯度增益,有效提高了小目标的检测精度。

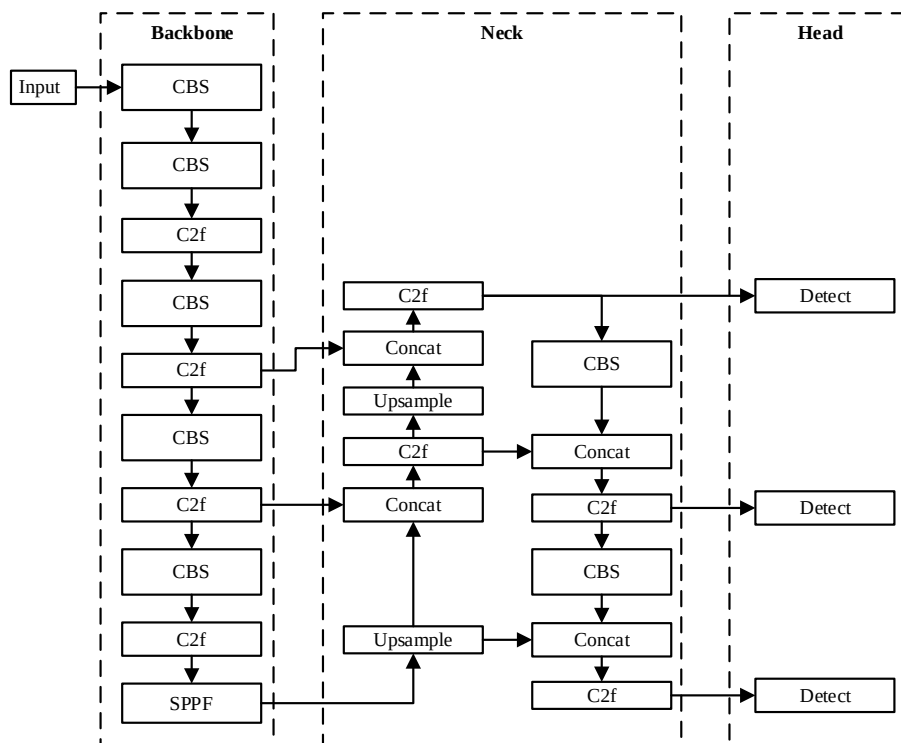


Fig. 1 Structure of YOLOv8 model

图1 YOLOv8模型结构

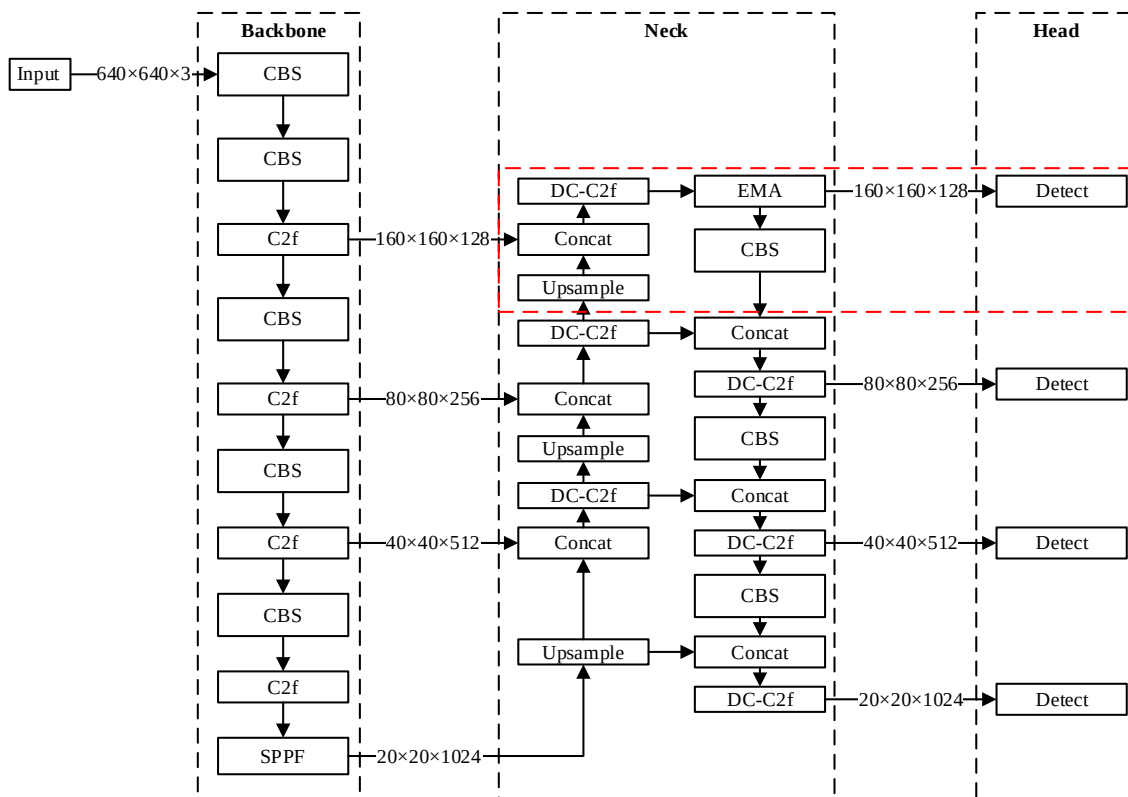


Fig. 2 Structure of SODM-YOLOv8 model

图2 SODM-YOLOv8模型结构

2.1 特征融合方式

如图3(彩图扫OSID码可见下同)所示,紫色代表大目标检测头,蓝色代表中目标检测头,黄色代表小目标检测头,红色代表微小目标检测头。YOLOv8网络的特征融合

信息包含浅层和深层信息,目标位置信息主要存在于浅层网络。在骨干网络进行特征提取与下采样操作时,小目标的位置信息会逐渐被丢失,卷积操作也会导致目标细节特征丢失,最终使小目标检测精度较差。为保留更多小目标

特征信息,避免过多卷积操作带来的不良影响,本文将骨干网络的第二层网络残差连接到颈部网络,然后与骨干网络卷积后的特征图拼接。改进后的特征融合方式有效保留了小目标的空间信息,提高了对小目标的检测能力。

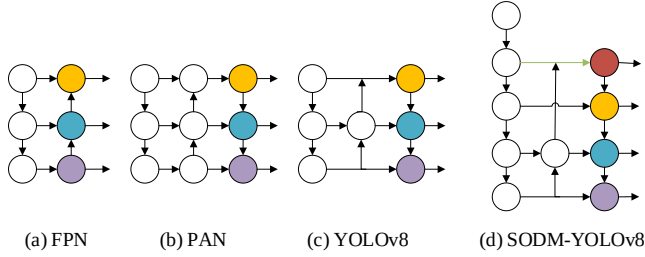


Fig. 3 Feature fusion method of SODM-YOLOv8

图 3 SODM-YOLOv8 的特征融合方式

2.2 小目标检测层

本文将输入图像分辨率统一为 640×640 。原始 YOLOv8 模型对输入图像进行下采样操作,生成 20×20 、 40×40 、 80×80 的特征图,分别检测 32×32 、 16×16 、 8×8 尺寸的目标。然而,无人机通常飞行高度较高,拍摄的小目标尺寸较小,当小目标尺寸低于 8×8 时,YOLOv8 模型的检测效果较差。此外,YOLOv8 模型进行的多次下采样操作丢失了很多小目标特征信息,进一步降低了对小目标的检测效果。因此,本文设计一个小目标检测模块,在原有特征图的基础上引入 160×160 的大特征图检测小目标,该分辨率可以保留更多小目标的细节特征与位置信息,具体改进方法为:首先,在骨干模块的第二层与颈部模块的第十七层使用残差连接,将骨干网络输出的 $160 \times 160 \times 128$ 大特征图与上采样后的 $160 \times 160 \times 256$ 特征图整合;其次,将堆叠后的特征图通过 EMA 机制对通道维度进行融合,提高了模型的多尺度特征融合能力;再次,通过 Upsample 上采样方法将特征图恢复到原始大小,通过内容信息和位置信息堆叠来获得完整的小目标特征信息;最后,通过额外的 P2 检测头对小目标进行检测。

2.3 EMA 机制

复杂环境下的目标检测任务由于受到图像背景的干扰,容易出现小目标漏检问题。为降低漏检率,本文采用 EMA 机制提升小目标的检测能力。该机制主要由特征分组、并行子网、跨域学习模块构成^[25]。具体结构如图 4 所示。

首先,EMA 机制通过特征分组模块将输入通道划分为批量维度。对于输入特征图 $X \in \mathbb{R}^{C \times H \times W}$, H 和 W 分别表示特征图的高度和宽度, C 表示输入通道数量。将 X 划分为 G 个子特征来学习多尺度特征信息,通过更新注意力权重来加强子特征表示。

其次,通过并行子网模块,采用 3 个并行结构处理多尺度特征信息。两条 1×1 的并行分支分别从垂直和水平方向对空间编码,公式为式(1)和式(2)。将两个编码特征沿着高度方向拼接,并使用 1×1 的共享卷积和非线性的

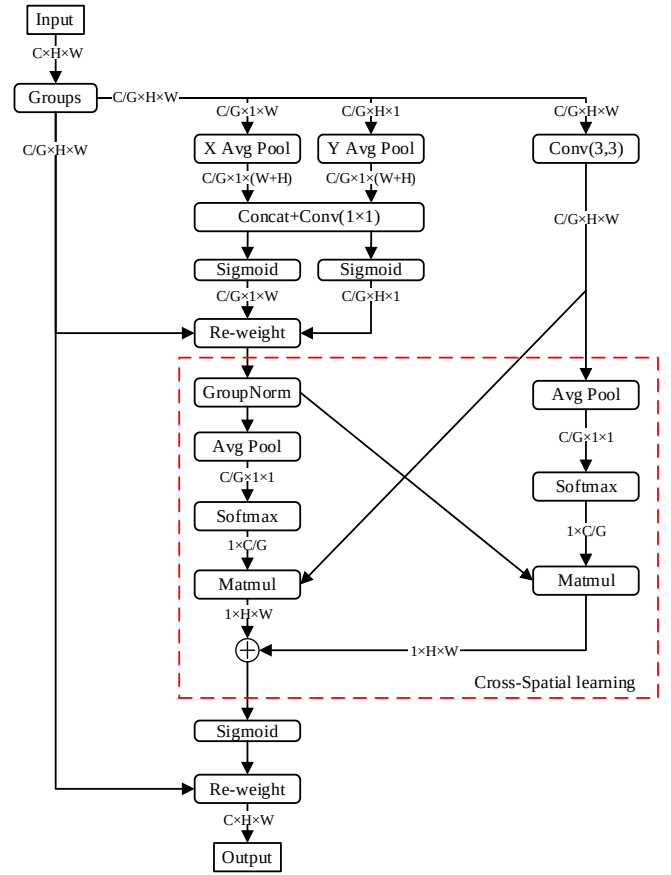


Fig. 4 Structure of EMA mechanism

图 4 EMA 机制结构

Sigmoid 函数分解为两个特征向量,维度分别为 $C/G \times 1 \times W$ 和 $C/G \times H \times 1$ 。另一条并行支路通过 3×3 的卷积得到一个特征矩阵,维度为 $C/G \times H \times W$ 。3 条并行分支将空间信息保留在通道内,用于捕捉多尺度特征信息。

$$Z_c^H(H) = \frac{1}{W} \sum_{0 \leq j \leq W} x_c(H, j) \quad (1)$$

$$Z_c^W(W) = \frac{1}{H} \sum_{0 \leq i \leq H} x_c(i, W) \quad (2)$$

最后,为收集来自不同空间的信息,通过跨域学习模块将 3×3 分支和 1×1 分支得到的特征信息聚合,再通过 $2D$ 全局平均池化对两个分支空间信息编码,公式为式(3)。编码后生成两个空间注意力图并对其注意力加权,最终经过 Sigmoid 函数得到输出特征图。跨域的信息聚合可更准确地获得图像中目标的特征信息,减少干扰因素;同时可处理短程和长程依赖性,捕捉了更多上下文信息。

$$Z_c = \frac{1}{H \times W} \sum_j \sum_i x_c(i, j) \quad (3)$$

2.4 DC-C2f 模块

普通卷积涉及到深层次的卷积运算,会导致模型计算量和参数量大幅上升,性能下降。如图 5 所示,本文采用 DualConv 结构对 YOLOv8 模型 C2f 模块中的卷积层进行改进,设计 DC-C2f 模块,以降低模型参数量^[26]。

如图 6 所示, DualConv 通过结合局部的 3×3 和全局的

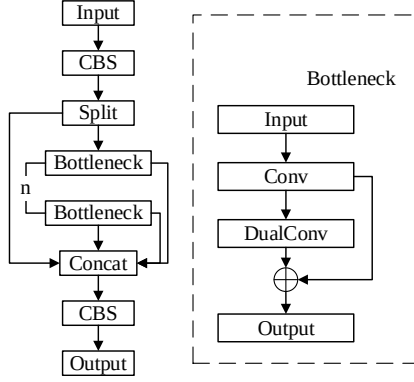


Fig. 5 Structure of DC-C2f

图5 DC-C2f结构

1×1 卷积来实现组内与组外特征的信息交互。将输入特征首先通过 3×3 的卷积核进行分组卷积,以降低模型参数量和计算量;其次经过 1×1 的卷积核进行点卷积,实现组外特征交互,以提高模型的特征提取能力。

将 Bottleneck 中的 3×3 卷积改为两个部分:第一部分采用分组卷积,分组数为 4;第二部分采用点卷积,步长设为 1,并通过组卷积排列滤波器,在保证模型检测能力的同时降低其参数量和计算量。滤波器排列结构如图 7 所示。

输入特征图大小为 $160 \times 160 \times 128$,输出特征图大小为 $160 \times 160 \times 128$,卷积核为 3×3 ,那么原有卷积的参数量为 $128 \times 3 \times 3 \times 128$,即 147 456。本文采用分组数为 4 的分组卷积,参数量为 $4 \times 32 \times 3 \times 3 \times 32$,即 36 864。可以看出,改进后

的参数量仅为原来的 1/4。

假设输出特征图大小为 $D_1 \times D_2 \times N$, D_1 和 D_2 分别为输出特征图的高度和宽度。输入特征映射到 N 个 $K \times K \times M$ 的卷积滤波器, M 为输入特征的通道数, $K \times K$ 为卷积核大小。则未改进的卷积计算量为:

$$F_{sc} = D \times D \times K \times K \times M \times N \quad (4)$$

DualConv 卷积的计算量为:

$$FL_{DC} = D^2 \times K^2 \times M \times N/G + D^2 \times M \times N \quad (5)$$

式(6)为两者计算量的比较公式。假设 $K=3$ 且 G 分组数足够大, DualConv 的计算量会比普通卷积少 9 倍。本文取 $G=4$,降低了约 2.7 倍的计算量。

$$R = \frac{FL_{DC}}{FL_{sc}} = \frac{1}{G} + \frac{1}{K^2} \quad (6)$$

2.5 损失函数

原始 YOLOv8 模型采用纵横比惩罚的损失函数 CIOU。该损失函数虽然引入了纵横比惩罚机制,但对低质量样本的学习能力较差。WIoU 通过动态聚焦机制对不同尺度的目标采用不同的梯度增益,有效提高了模型对低质量样本的学习能力^[27]。WIoU_{v3} 在 WiOU_{v1} 上附加了动态非单调聚焦机制,降低了低质量样本对梯度的不良影响。各损失函数的公式为:

$$\mathcal{L}_{IoU} = 1 - IoU \quad (7)$$

$$\mathcal{L}_{WIoU_{v1}} = \mathcal{R}_{WIoU} \mathcal{L}_{IoU} \quad (8)$$

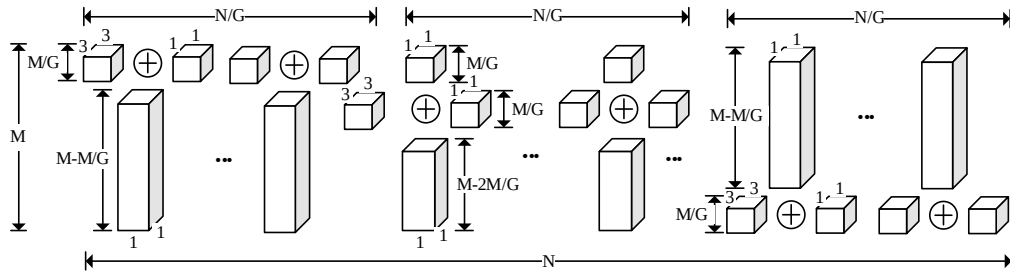


Fig. 6 Structure of DualConv

图6 DualConv结构

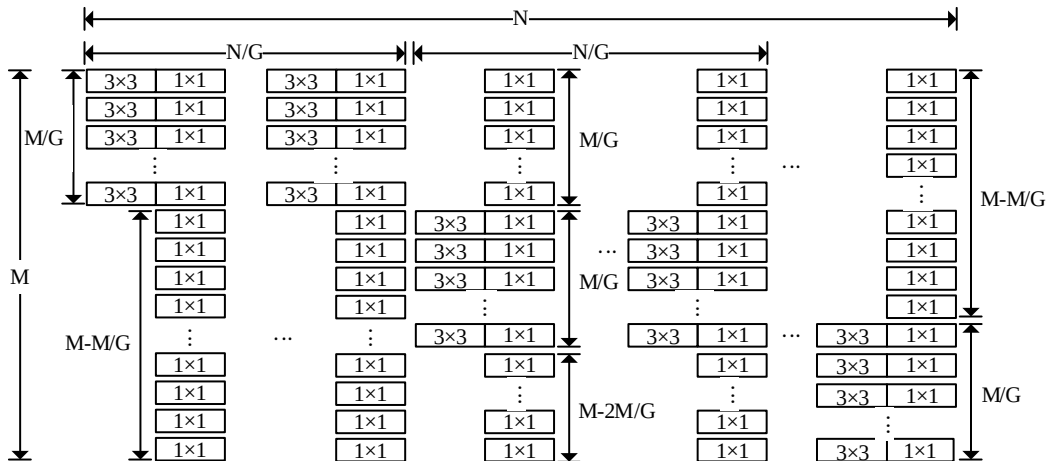


Fig. 7 Structure of filter arrangement

图7 滤波器排列结构

$$\mathcal{R}_{WIoU} = \exp \left(\frac{(x_{pred} - x_{gt})^2 + (y_{pred} - y_{gt})^2}{(W_g^2 + H_g^2)^*} \right) \quad (9)$$

$$\mathcal{L}_{WIoU_3} = r \mathcal{L}_{WIoU_1}, r = \frac{\beta}{\delta \alpha^{\beta - \delta}} \quad (10)$$

$$\beta = \frac{\mathcal{L}_{IoU}^*}{\mathcal{L}_{IoU}} \in [0, +\infty) \quad (11)$$

式中: IoU 为预测框与真实框的交并比; 离群度 β 用于描述样本质量; W_g 和 H_g 分别表示最小闭包区域的宽度和高度; (x_{pred}, y_{pred}) 和 (x_{gt}, y_{gt}) 分别表示预测框和真实框坐标。

然而, 上述损失函数均未考虑到 IoU 内部的限制。Inner- IoU 通过在边界框内部使用更小的辅助框来精确定位小目标, 提升了高度重叠的边界框的检测效果^[28]。本文采用 Inner- IoU 的思想改进 $WIoU$ 损失函数, 假设边界框的中心坐标为 (x_c, y_c) , 通过缩放比例参数 $ratio$ 控制辅助框大小并得到辅助框的 4 个顶点 k_a, k_b, k_c, k_d 。表示为:

$$k_a = x_c - \frac{w \times ratio}{2}, k_b = x_c + \frac{w \times ratio}{2} \quad (12)$$

$$k_c = y_c - \frac{h \times ratio}{2}, k_d = y_c + \frac{h \times ratio}{2} \quad (13)$$

Inner- IoU 的计算公式为:

$$Inter = \left(\min(k_b^{pred}, k_b^{gt}) - \max(k_a^{pred}, k_a^{gt}) \right) \times \left(\min(k_d^{pred}, k_d^{gt}) - \max(k_c^{pred}, k_c^{gt}) \right) \quad (14)$$

$$Union = (w^{gt} \times h^{gt}) \times ratio^2 + (w^{pred} \times h^{pred}) \times ratio^2 - Inter \quad (15)$$

$$IoU^{Inner} = \frac{Inter}{Union} \quad (16)$$

式中: k^{pred}, k^{gt} 分别表示预测框和真实框的计算值; w^{gt}, h^{gt} 分别表示真实框的宽和高; w^{pred}, h^{pred} 分别表示预测框的宽和高。

Inner- $WIoU$ 的计算公式为:

$$\mathcal{L}_{WIoU_3}^{Inner} = r(1 - IoU^{Inner}) \mathcal{R}_{WIoU}, r = \frac{\beta}{\delta \alpha^{\beta - \delta}} \quad (17)$$

式中: α, δ 为超参数, 分别设为 1.5 和 4.5。 $WIoU$ 通过将小梯度增益分配给低质量样本来降低有害梯度的影响。Inner- IoU 通过调整 $ratio$ 的值来使高度重叠的边界框使用较小的辅助框, 而重叠程度低的边界框使用较大的辅助框, 这种自适应动态调整方式能对重叠区域进行更精确的评估, 其中 $ratio$ 的调整范围为 $[0.5, 1.5]$ 。

3 实验方法与结果分析

3.1 实验环境与训练参数配置

硬件设备为 20 核 L20 处理器, 100 G 内存, 40 G 显存。软件环境为 Ubuntu 20.04 操作系统、Python3.8、CUDA11.3 以及 Pytorch1.11.0。训练参数设置见表 1。

3.2 实验数据集

采用 VisDrone2019 数据集, 包含黑暗、强光和密集等

Table 1 Training parameter settings

表 1 训练参数设置

参数名称	参数值
输入分辨率	640×640
训练轮数	250
批量大小	16
初始学习率	0.01
优化器	SGD
动量	0.937
权重衰减系数	0.000 5
Patience	50
Mosaic	1.0
图像缩放	0.5
图像平移	0.1
图像翻转	0.5

复杂场景下的 10 种类别图片, 其中训练集 6 471 张, 验证集 548 张, 测试集 1 610 张图片; 极小目标 (16×16 像素面积) 占比 12.05%, 小目标 (32×32 像素面积) 占比 44.70%。

为验证模型的泛化性, 在公开数据集 DOTA1.0 上进行比较实验。该数据集包含 15 种类别图片, 其中训练集 1 403 张, 验证集 468 张, 测试集 935 张。

3.3 评价指标

(1) 召回率 (Recall) 用于衡量模型找到真正例的能力。计算公式为:

$$Recall = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad (18)$$

式中: N_{TP} 表示真正例, N_{FN} 表示假负例。

(2) 精确率 (Precision) 用于衡量模型预测正样本的准确度。计算公式为:

$$Precision = \frac{N_{TP}}{N_{TP} + N_{FP}} \quad (19)$$

式中: N_{FP} 表示假正例。

(3) 平均精度 (mAP) 用于衡量模型的性能强弱, 表示不同类别精度的均值, 能准确描述模型的整体性能。计算公式为:

$$mAP = \frac{1}{c} \times \sum_{i=1}^c AP_i \quad (20)$$

式中: c 表示目标种类数; AP_i 表示第 i 个目标对象的精度。

(4) 每秒传输帧率 (Frames Per Second, FPS) 为描述模型检测速度的指标, 其值越高表示模型可以更快地检测和识别目标。计算公式为:

$$FPS = \frac{frameCount}{elapsedTime} \quad (21)$$

式中: $frameCount$ 表示总帧数; $elapsedTime$ 表示总时间。

(5) 参数量 (Params) 用于衡量模型的复杂度。

3.4 分类实验

如表 2 所示, SODM-YOLOv8 模型对 VisDrone2019 数据集中行人、汽车、货车、公交车、摩托车 5 类目标的平均检测精度高于基线模型 YOLOv8, 其中对于汽车等较大目

标的检测效果有稳定改进,对于人、行人等小目标的平均检测精度增长幅度最高,达到 10.7%,说明本文模型对小目标检测有一定的针对性。

Table 2 Experimental results on the VisDrone2019 dataset

表 2 在 VisDrone2019 数据集上的实验结果 (%)

类别	YOLOv8	SODM-YOLOv8
行人	40.7	51.4
人	32.0	42.0
自行车	12.0	16.1
汽车	79.3	83.7
货车	44.2	48.0
卡车	36.2	37.9
三轮车	27.2	29.8
带篷三轮车	15.5	18.5
公交车	55.1	59.3
摩托车	43.8	51.6
mAP@0.5	38.6	43.8

3.5 比较实验

为考察本文算法在检测速度与精度方面的平衡性,将其与一些经典的小目标检测算法进行比较,包括 Faster-RNN^[7]、RetinaNet^[15]、CornerNet^[16]、YOLO-PWCA^[17]、YOLOv5^[29]、Bi-YOLOv8s^[30]和 SC-YOLO^[31]。如表 3 所示, SODM-YOLOv8 算法的 mAP@0.5 值为 43.8%, mAP@0.5:0.95 值为 26.0%,均优于其他对照模型,虽然在检测速度上稍有逊色,但是仍然具有实时性,在检测精度与速度之间达到了平衡。

Table 3 Performance comparison of various small target detection algorithms

表 3 各小目标检测算法性能比较

算法	mAP@0.5/%	mAP@0.5:0.95/%	FPS/(帧/s)
RetinaNet	28.7	11.8	23
Faster-RNN	33.2	17.0	15
CornerNet	34.1	17.4	33
YOLOv5	35.2	19.5	89
YOLO-PWCA	40.6	22.9	181
Bi-YOLOv8s	40.7	/	135
SC-YOLO	43.1	25.7	/
SODM-YOLOv8	43.8	26.0	80

3.6 消融实验

为验证本文所提改进模块的有效性,进行消融实验,其中 S 表示小目标检测模块, E 表示 EMA 注意力机制, D 表示 DC-C2f 模块, I 表示 Inner-WIoU 损失函数。消融实验结果如表 4 所示。可以看出, YOLOv8-S 模型的 mAP@0.5 相较 YOLOv8 提高了 4.3%, mAP@0.5:0.95 提高了 2.1%,验证了小目标检测模块的有效性。加入 DC-C2f 模块后,模型的参数量降低了 28.3%,验证了该模块的有效性。将损失函数替换为 Inner-WIoU 损失函数后,模型的 mAP@0.5 提高了 0.5%, FPS 提高了 3 帧/s,验证了改进损失函数的有效性。综上所述,改进模型的检测性能相较原始 YOLOv8 模型大幅提升,与理论分析结果相对应,验证了所提各模块的可行性和有效性。

Table 4 Ablation experiment result

表 4 消融实验结果

算法	mAP@0.5	mAP@0.5:0.95	Precision	Recall	FPS	Params
	/%	/%	/%	/%	/(帧/s)	/M
YOLOv8	38.6	22.8	49.7	37.8	93	10.6
YOLOv8-S	42.9	24.9	52.2	41.4	71	10.9
YOLOv8-S-E	43.1	25.3	52.4	41.6	65	11.1
YOLOv8-S-E-D	43.3	25.5	52.9	41.7	84	7.6
YOLOv8-S-E-D-I	43.8	26.0	53.1	41.9	87	7.6

为验证 EMA 机制的有效性,将其与 CA (Coordinate Attention) 和 CBAM 机制进行比较,结果见表 5。可以看出,结合 EMA 机制的模型平均精度最高,这是由于 VisDrone2019 数据集的小目标实例较多, CA 和 CBAM 机制虽然能够在一定程度上提升特征表示能力,但对于小目标不够敏感。EMA 机制能通过多尺度特征融合和跨域学习模块有效捕捉小目标,提高了其检测精度。

Table 5 Experimental results of attentional mechanism ablation

表 5 注意力机制消融实验结果

模型	mAP@0.5/%	mAP@0.5:0.95/%	FPS/(帧/s)
CA	41.8	23.8	89
CBAM	42.1	25.1	75
EMA	43.1	25.9	81

3.7 验证实验

在强光条件下进行检测, YOLOv8、SODM-YOLOv8 算法的可视化检测结果如图 8 和图 9 所示。图中阳光强烈,车辆与树木之间相互遮挡,难以分辨小目标的细节特征。与 YOLOv8 相比, SODM-YOLOv8 算法能检测出行人、汽车等不同大小的目标,并在红色矩形框处检测出更多目标,降低了漏检率,在蓝色矩形框处将 YOLOv8 误检的轿车识别为货车,有效降低了误检率,验证了本文算法在强光环境下的小目标检测能力。



Fig. 8 Detection result visualization of YOLOv8 under strong light conditions

图 8 YOLOv8 在强光条件下的检测结果可视化

在黑暗条件下进行检测, YOLOv8、SODM-YOLOv8 算法的可视化检测结果如图 10 和图 11 所示。图中光线黑暗,难以分辨目标位置,在红色矩形框处,本文算法依然能检测出更多的行人、摩托车等小目标,而这些目标被 YOLOv8 算法漏检,验证了本文算法在黑暗环境下的小目标检测能力。



Fig. 9 Detection result visualization of SODM-YOLOv8 under strong light conditions

图 9 SODM-YOLOv8 在强光条件下的检测结果可视化



Fig. 10 Detection results visualization of YOLOv8 under dark conditions

图 10 YOLOv8 在黑暗条件下的检测结果可视化



Fig. 11 Detection results visualization of SODM-YOLOv8 under dark conditions

图 11 SODM-YOLOv8 在黑暗条件下的检测结果可视化

如图 12 和图 13 所示, 车辆与车辆之间、车辆与房屋相互遮挡, 拍摄角度导致小目标位于较远的位置, 红色矩形框标记漏检目标, 蓝色矩形框标记误检目标。可以看出, 本文算法依然能检测出汽车、公交车等被 YOLOv8 漏检、误检的目标, 验证了其在遮挡条件下的小目标检测能力。

3.8 泛化性实验

为验证本文算法的泛化性, 在 DOTA1.0 数据集上进行比较实验。结果见表 6。可以看出, 本文算法在 DOTA1.0 数据集上的平均精度有一定提升, 在小型车辆、轮船、篮球场等目标上的 mAP@0.5 值超过总体平均精度, mAP@0.5 较 YOLOv8 模型提升了 1.4%, 验证了本文算法具有较好的泛化性。



Fig. 12 Detection results visualization of YOLOv8 under occlusion

图 12 YOLOv8 在遮挡条件下的检测结果可视化



Fig. 13 Detection results visualization of SODM-YOLOv8 under occlusion

图 13 SODM-YOLOv8 在遮挡条件下的检测结果可视化

Table 6 Experimental results on DOTA1.0 dataset

表 6 在 DOTA1.0 数据集上的实验结果 (%)

类别	YOLOv8	SODM-YOLOv8
小型车辆	74.3	78.4
大型车辆	62.0	63.2
储罐	92.5	93.7
棒球场	72.5	74.0
桥梁	94.7	95.1
篮球场	85.0	86.6
地面田径场	61.4	63.7
港口	54.7	55.3
网球场	89.3	90.1
轮船	87.0	89.4
飞机	68.6	70.2
直升机	57.1	58.3
环形交叉路口	64.3	65.0
足球场	44.2	44.9
游泳池	55.7	56.2
mAP@0.5	70.9	72.3

4 结语

针对无人机图像小目标检测算法精度较低和模型较为复杂的问题, 本文提出 SODM-YOLOv8 算法。该算法首先在特征融合网络中设计了小目标检测模块, 同时结合 EMA 机制, 在提升对小目标检测能力的同时亦能稳定提升对大目标的检测能力; 其次引入 Inner-WIoU 损失函数, 改善低质量样本和 IoU 内部高度重叠的不良影响; 最后设计了 DC-C2f 模块, 有效降低了模型参数数量和计算量, 使得其

在检测速度与精度之间达到良好的均衡,能有效完成实时目标检测任务。未来将通过超分辨率重建等方法进一步提高复杂环境下的无人机图像小目标检测效果。同时改进骨干网络结构,使模型在保证检测精度的同时达到更快的检测速度。

参考文献:

- [1] WU H Y, LEI J S, CHEN L F, et al. Application of improved target detection algorithm in substation safety control[J]. Journal of Computer Engineering & Applications, 2022, 58(24): 313-320.
吴宏毅,雷景生,陈林锋,等.改进目标检测算法在变电站内安全管控的应用[J].计算机工程与应用,2022,58(24):313-320.
- [2] JI Y M, CHEN Z Y, TIAN P H, et al. Review on 3D target detection methods in unmanned vehicles[J]. Journal of Nanjing University of Posts and Telecommunications: Natural Science Edition, 2019, 39(4): 72-79.
季一木,陈治宇,田鹏浩,等.无人驾驶中3D目标检测方法研究综述[J].南京邮电大学学报:自然科学版,2019,39(4):72-79.
- [3] ZHANG G S, GE G Y, ZHU R H, et al. Gear defect detection based on improved YOLOv3 network[J]. Laser & Optoelectronics Progress, 2020, 57(12):153-161.
张广世,葛广英,朱荣华,等.基于改进YOLOv3网络的齿轮缺陷检测[J].激光与光电子学进展,2020,57(12):153-161.
- [4] WANG P F, HUANG H M, WANG M Q. Improved complex road target detection algorithm of YOLOv5[J]. Journal of Computer Engineering & Applications, 2022, 58(17): 81-92.
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.
- [5] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271.
- [6] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: optimal speed and accuracy of object detection [DB/OL]. <https://arxiv.org/abs/2004.10934>.
- [7] LI C, LI L, JIANG H, et al. YOLOv6: a single-stage object detection framework for industrial applications [DB/OL]. <https://arxiv.org/abs/2209.02976>.
- [8] 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.
- [9] 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.
- [10] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C]//European Conference on Computer Vision, 2016: 21-37.
- [11] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980-2988.
- [12] LAW H, DENG J. Cornernet: detecting objects as paired keypoints [C]//Proceedings of the European Conference on Computer Vision, 2018: 734-750.
- [13] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//Columbus: 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014.
- [14] GIRSHICK R. Fast R-CNN [C]//Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [15] REN S, HE K, 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, 2016, 39(6): 1137-1149.
- [16] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961-2969.
- [17] YU A Z, WEI W W, WANG P, et al. Small target detection algorithm for UAV based on block compound attention [J/OL]. Journal of Aircraft, 2023-09-15. <http://kns.cnki.net/kcms/detail/11.1929.V.20230913.1505.014.html>.
于傲泽,魏维伟,王平,等.基于分块复合注意力的无人机小目标检测算法[J/OL].航空学报,2023-09-15. <http://kns.cnki.net/kcms/detail/11.1929.V.20230913.1505.014.html>.
- [18] WANG J, ZHANG F, ZHANG Y, et al. Lightweight object detection algorithm for UAV aerial imagery [J]. Sensors, 2023, 23(13): 5786.
- [19] WANG G, CHEN Y, AN P, et al. UAV-YOLOv8: a small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios [J]. Sensors, 2023, 23(16): 7190.
- [20] DU B, HUANG Y, CHEN J, et al. Adaptive sparse convolutional networks with global context enhancement for faster object detection on drone images [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 13435-13444.
- [21] MIN L, FAN Z, LV Q, et al. YOLO-DCTI: small object detection in remote sensing base on contextual transformer enhancement [J]. Remote Sensing, 2023, 15(16): 3970.
- [22] SAFALDIN M, ZAGHDEN N, MEJDOUB M. An improved YOLOv8 to detect moving objects [J]. IEEE Access, 2024, 12: 59782-59806.
- [23] CAI Y, ZHAO Y, WEN S, et al. Improved YOLOv8 SAR image aircraft object detection method [C]//2024 7th International Symposium on Autonomous Systems, 2024: 1-6.
- [24] WANG H, LIU C, CAI Y, et al. YOLOv8-QSD: an improved small object detection algorithm for autonomous vehicles based on YOLOv8 [J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 1-16.
- [25] OUYANG D, HE S, ZHANG G, et al. Efficient multi-scale attention module with cross-spatial learning [C]//ICASSP 2023 IEEE International Conference on Acoustics, Speech and Signal Processing, 2023: 1-5.
- [26] ZHONG J, CHEN J, MIAN A. DualConv: dual convolutional kernels for lightweight deep neural networks [J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 34(11): 9528-9535.
- [27] TONG Z, CHEN Y, XU Z, et al. Wise-IoU: bounding box regression loss with dynamic focusing mechanism [DB/OL]. <https://arxiv.org/abs/2301.10051>.
- [28] ZHANG H, XU C, ZHANG S. Inner-IoU: more effective intersection over union loss with auxiliary bounding box [DB/OL]. <https://arxiv.org/abs/2311.02877>.
- [29] ZHAO J D, ZHEN G Y, CHU C Q. UAV image target detection algorithm based on YOLOv8 [J]. Computer Engineering, 2024, 50(4): 113-120.
赵继达,甄国涌,储成群.基于YOLOv8的无人机图像目标检测算法[J].计算机工程,2024,50(4):113-120.
- [30] CHENG H X, QIAO Q Y, LUO X L, et al. Object detection algorithm of UAV aerial photography image based on improved YOLOv8 [J]. Radio Engineering, 2024, 54(4): 871-881.
程焕新,乔庆元,骆晓玲,等.基于改进YOLOv8的无人机航拍图像目标检测算法[J].无线电工程,2024,54(4):871-881.
- [31] LI Y C, SHI W Y, FENG C. Lightweight YOLOv8 detection algorithm for small target detection in UAV aerial photography [J/OL]. Computer Engineering and Application, 2024-06-06. <http://kns.cnki.net/kcms/detail/11.2127.TP.20240605.1426.006.html>.
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/OL].计算机工程与应用,2024-06-06. <http://kns.cnki.net/kcms/detail/11.2127.TP.20240605.1426.006.html>.

(责任编辑:尹晨茹)