

文章编号: 1002-1582(2026)01-0086-08

融合滑动窗口与改进SGBM的高分辨率 目标检测与测距研究

丁建军, 汪刚林, 陈正超, 张超, 夏礼恒, 夏骏麒, 孙超*

(1. 江汉大学 专用装备智能测控湖北省工程研究中心, 湖北 武汉 430056)

(2. 江汉大学 智能制造学院, 湖北 武汉 430056)

摘 要: 随着成像技术的发展, 高分辨率图像广泛应用于多个领域, 但在复杂天气、多尺度变化和立体匹配复杂度高等条件下仍面临挑战。提出一种结合滑动窗口的目标检测与测距方法, 在处理高分辨率图像时, 采用滑动窗口对图像进行分割并保留20%的重叠区域, 随后将各子图输入至YOLOv8模型进行检测。为提升模型性能, 在backbone中引入GCANet模块以增强雾霾环境下的图像质量, 并加入HorBlock以提升多尺度特征表征能力。检测完成后, 合并各子图结果, 并采用引入了CIoU匹配机制的非极大值抑制算法(NMS)进行后处理, 以缓解密集目标场景中的过抑制问题。最后, 设计了基于检测框约束的局部匹配算法, 结合加权最小二乘滤波(WLS)补全视差空洞。实验结果表明, 在自制的雾天场景Fog-A数据集与公开的无人机视角VisDrone2019数据集上, 方法相较于原始YOLOv8n模型, mAP分别提升9.8%和6.1%; 匹配时间缩短至原SGBM的三分之一, 精度提升0.7%。

关 键 词: 成像光学; 高分辨率; 滑动窗口; 目标检测; YOLOv8n; SGBM

中图分类号: TP394.1; TH691.9

文献标识码: A

High-resolution object detection and ranging via sliding window integration and enhanced SGBM

DING Jianjun, WANG Ganglin, CHEN Zhengchao, ZHANG Chao,
XIA Liheng, XIA Junqi, SUN Chao*

(1. Hubei Engineering Research Center for Intelligent Detection and Control of Specialized Equipment,
Jiangnan University, Wuhan 430056, China)

(2. College of intelligent manufacturing, Jiangnan University, Wuhan 430056, China)

Abstract: With the rapid advancement of imaging technologies, high-resolution images have found widespread applications across diverse fields. Nevertheless, challenges persist under complex weather conditions, multi-scale variations, and the high computational burden of stereo matching. A high-resolution object detection and distance measurement framework that integrates a sliding window strategy is proposed. Specifically, input images are partitioned into sub-images with 20% overlap and subsequently processed by the YOLOv8 model for detection. To enhance model robustness, a GCANet (Gated Context Aggregation Network) module is incorporated into the backbone to enhance feature robustness against atmospheric degradation, while a HorBlock (High-order Interaction Block) module is introduced to strengthen multi-scale feature representation. After de-

收稿日期: 2025-05-30; 收到修改稿日期: 2025-08-24

基金项目: 国家重点研发计划基金资助项目(2018YFD1100104); 湖北省教育厅教学研究项目(2022277) 武汉市教育局产学研项目(CXY202210)

作者简介: 丁建军(1969—), 男, 教授, 博士, 从事化学传感器、智能检测、智能制造、图像处理、机电控制研究。10107093@jhun.edu.cn

孙超(1984—), 男, 副教授, 博士, 从事传感器设计、深度学习算法、计算机视觉研究。

* 通讯作者: sycsc115@163.com

tection, results from the sub-images are merged, and a CIoU-based NMS is applied to alleviate over-suppression in dense object scenarios. Furthermore, a detection box-constrained local stereo matching algorithm is designed, which is then refined with Weighted Least Squares (WLS) filtering to fill disparity holes and improve depth estimation. Experimental results on a self-constructed foggy dataset (Fog-A) and the public VisDrone2019 dataset show that the proposed method achieves mAP improvements of 9.8% and 6.1%, respectively, over the baseline YOLOv8n model. Moreover, the proposed local matching algorithm reduces computation time to one-third of that of the original SGBM and improves matching accuracy by 0.7%.

Keywords: Imaging optics; high resolution; sliding window; target detection; YOLOv8n; SGBM

0 引言

随着成像与计算技术的发展, 高分辨率图像在智能交通、无人机感知及安防监控等领域应用广泛。然而, 恶劣天气干扰、多尺度特征提取困难、密集目标遮挡以及双目立体匹配计算量大等问题, 仍严重制约着检测与测距的精度与效率。

在针对高分辨率图像的目标检测领域中, 研究者们提出了多种方法, Liu^[1]等提出DZNet (Dual-Zone Network) 双阶段检测模型, 利用候选区域动态筛选与局部特征强化实现小目标亚像素级识别, 但该方法依赖低分辨率缩略图, 可能丢失高分辨率信息细节。Zhou^[2]等构建高分辨率特征保持网络, 并通过下采样率调整与马赛克混合增强提升复杂背景适应性, 但在极端天气条件下鲁棒性仍有限。Nguyen Hoanh^[3]提出了一种融合多层目标聚焦网络与Transformer编码器-解码器结构的高分辨率目标检测方法, 然而, 该方法在高分辨率场景下仍面临计算资源消耗较大的问题。IA-YOLO (Image-Adaptive YOLO)^[4]引入可微分图像处理模块, 实现了图像自适应增强, 从而改善恶劣天气下的检测性能; MSAR-DefogNet (Multi-Scale Attention Residual Defogging Network)^[5]结合多尺度注意力机制与残差网络结构, 提高了雾霾场景下图像去雾效果与检测准确性; HSF-YOLO (Hierarchical Scale Fusion YOLO)^[6]则通过多层次特征融合与轻量化设计, 提升了小目标检测能力并降低了模型复杂度。这些方法在特定场景下均取得了良好效果, 但仍存在过度依赖图像增强、高分辨率条件下效率不足, 或对多尺度目标适应性不强等问题。

为应对高分辨率图像^[7,8]中目标检测与测距在效率、鲁棒性及多尺度适应性方面的挑战, 本文提出一种基于滑动窗口的高效检测与测距方法。首先, 采用滑动窗口策略对高分辨率图像进行裁切, 解决小目标细节丢失问题。其次, 在目标检测环节, 提出改进的YOLOv8n算法: 在Mosaic模块融入GCANet (Gated Context Aggregation Network) 以增强雾天特征保留, 并在特征金字塔中引入HorBlock (High-order Interaction Block) 模块以兼

顾多尺度交互效率; 同时, 结合CIoU-NMS策略缓解密集目标过抑制现象。最后, 针对传统SGBM算法视差空洞与耗时高的问题, 提出WY-SGBM (Weighted Least Squares YOLO-SGBM) 局部匹配算法, 利用检测框约束与加权最小二乘滤波, 实现快速且精确的深度估计。

为验证所提方法的有效性, 本文设计了多层次的实验方案: 在自制Fog-A雾天数据集上评估检测改进模块的有效性; 在VisDrone2019^[9]无人机航拍公共数据集上验证滑动窗口模型在目标检测任务中的性能; 在小车图像数据集上测试改进后的立体匹配效果。进一步地, 设计了一个综合实验, 将所有改进方案进行整体应用与测试, 以评估改进方法在多尺度、低能见度场景下的检测精度和距离估计能力。

1 基于滑动窗口的分割方法

YOLOv8接收的图像尺寸多为 640×640 、 416×416 等, 高分辨率图像在送入模型之前会被压缩, 会丢失大量细节信息, 从而导致目标检测性能下降, 同时可能出现检测框漂移或NMS^[10]过抑制问题。

综合以上问题考虑, 本文采用滑动窗口方法对高分辨率图像进行分割。为避免额外的缩放失真并保持与训练阶段一致, 滑动窗口的裁剪尺寸 640×640 与YOLO检测模型的输入尺寸保持一致。同时, 为防止图像边缘处的目标被截断或漏检, 本文在相邻窗口之间设置了20%的重叠区域。既保证目标至少在一个子图中完整出现, 又避免了过多冗余计算。分割结果如图1所示, 其中深色阴影部分为重叠部分。



图1 图像切割法

裁剪完成后, 将裁切后的 640×640 图片送入YOLO检测算法, 对所有生成的子图进行目标检测并融合所有图片的检测结果, 但是此时在滑动窗口

的重叠处会出现重复的检测目标,因此需要使用 NMS 算法对多余的检测框进行抑制。传统 NMS 算法存在过抑制和检测框漂移问题,交并比作为 NMS 算法的相似性度量指标,具有非负性、对称性、三角不等式及尺度不变等特性。但 IoU 有两个明显缺陷:一是检测框与预测框无重叠时恒为零,无法有效区分远距离目标的匹配关系;二是不同空间分布的检测对可能有相同值,易造成误匹配。

为了解决这个问题,本文将 IoU (Intersection over Union) 替换为 CIoU (Complete Intersection over Union)^[11],CIoU 考虑了目标边界框的中心点之间、宽高之间以及目标边界框的对角线之间的距离。相比之下,DIoU (Distance Intersection over Union)^[12]仅利用中心点之间的欧氏距离来优化位置关系,而 GIoU (Generalized Intersection over Union)^[13]的惩罚项并不包含中心点距离和长宽比的几何约束。本文所使用的 CIoU 综合考量这些,提供了更为全面的距离度量方式,能更精准地反映目标边界框间的相似度。

此外,为了在非极大值抑制过程中更准确地衡量预测框与真实框的几何重叠度,CIoU 在考虑重叠面积和中心点距离的基础上,额外引入了长宽比的惩罚项,重新定义了度量标准。使得模型在度量预测框与真实框的匹配程度时更加精确。计算公式为

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

式中, b 是预测边界框, b^{gt} 表示真实边界框, $\rho^2(b, b^{gt})$ 表示预测框中心点与真实框中心点的欧氏距离平方; c^2 表示包围预测框与真实框的最小闭包对角线长度平方; α 是权重函数;而 v 用来度量长宽比的相似性。

CIoU 通过引入中心点归一化距离与长宽比惩罚项,有效克服了上述局限性,从而在各种重叠情形下均能提供更为鲁棒且准确的匹配度量。

2 基于目标检测的改进 YOLOv8

2.1 复杂场景图像增强算法

在雨雾场景下,图像饱和度与对比度的降低严重阻碍了检测模型的特征提取,导致准确率显著下降。传统去雾方法多基于物理模型,利用人工先验假设估计透射率与大气光。然而,此类方法对先验条件依赖性强,在复杂雨雾环境中难以准确建模,导致图像增强效果受限。

本研究在 YOLOv8n 网络中引入 GCANet^[14]模块以提升复杂场景下的目标检测性能。GCANet 模块位于 Backbone 输入端,对雾霾或雨雾图像进行端到端增强,直接建模雨雾图像到清晰图像的映射关系。其模型结构如图 2 所示。

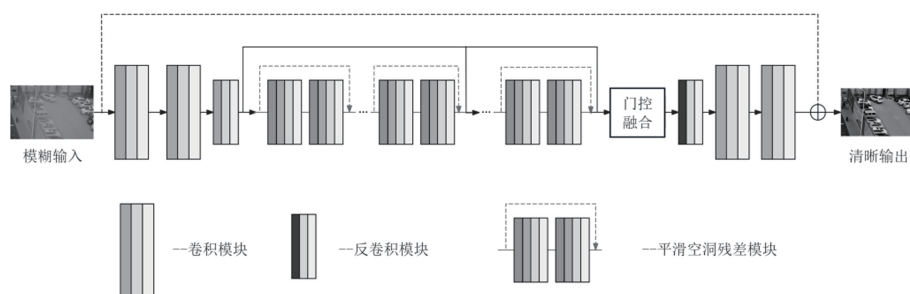


图2 GCANet算法结构

GCANet 的整体流程包括输入输出、特征提取、特征融合、解码和残差添加五个阶段:裁剪后的雾霾图像首先经卷积块提取特征,再通过平滑空洞残差块 SDC 捕捉多尺度信息并利用门控机制自适应融合,随后在解码阶段通过反卷积上采样与跳跃连接恢复细节,最后将残差增强图与原图叠加生成清晰图像。增强后的图像直接传递至 YOLOv8n 首层 Conv 模块,替代原始雾霾图像作为检测网络的输入,有效的提升了后续检测头在雾霾环境下的鲁棒性与准确性。

2.2 YOLOv8n 多尺度优化

在目标检测场景中,由于目标距离差异和尺度变化,检测网络往往面临严重的多尺度建模问

题。传统的特征金字塔结构在捕捉长程依赖和全局上下文时存在不足,容易导致小目标或远距离目标的特征表达不充分。针对这一问题,本文引入 HorNet^[15]的高阶空间思想,在特征金字塔中加入 HorBlock 模块以优化多尺度特征处理。

HorNet 通过递归门控卷积构建骨干网络,能够实现对输入特征的自适应长程高阶空间交互,同时有效降低自注意力机制所带来的高计算复杂度。HorBlock 由 LayerNorm 和 g^nConv ^[16]组成,其中 LayerNorm 保证跨尺度特征融合时的分布稳定性与网络鲁棒性;而 g^nConv 通过递归门控卷积实现任意阶的空间交互建模,突破了传统卷积仅限于局部感受野的限制。由此,HorBlock 不仅继承

了卷积的平移不变性,还克服了局部注意力的空间不对称问题,在小目标与远距离目标的建模上表现尤为突出。HorNet结构如图3所示,图示左侧为HorNet整体结构图,右侧为 g^nConv 结构图。

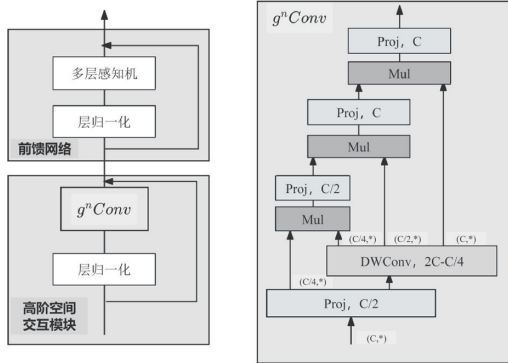


图3 HorNet算法结构

在传入特征图后,采用LayerNorm进行归一化,以保证不同尺度和不同层级的特征在分布上的一致性。其次,利用 g^nConv 对特征进行处理,该操作能够在卷积的局部感受野基础上,引入高阶空间交互建模机制,实现对全局上下文与多尺度空间关系的建模。与自注意力不同, g^nConv 保留了卷积的平移不变性,并且计算复杂度更低。最后,通过残差连接将输入特征与变换后的特征进行融合,以提升网络的梯度流动性和稳定性。

在YOLOv8n中,HorBlock被嵌入到特征金字塔结构中。HorBlock的高阶空间交互能力使得这些特征在融合过程中能够更好地传递跨尺度信息。这样一方面增强了网络对小目标和远距离目标的表达能力,另一方面也提升了模型在复杂天气条件下的鲁棒性,能够有效的弥补YOLOv8n在多尺度特征交互上的不足,从而提升整体检测性能。

3 立体匹配算法设计

双目匹配与测距在高分辨率图像立体匹配的应用中存在较大难度。无论是全局匹配还是半全局匹配,在计算视差时,计算量都非常庞大,难以实现实时测距的效果。在全图SGBM中,每一对像素都需要计算代价和聚合,这在高分辨率下会带来巨大的计算开销。

因此,本章基于前文的方法,设计了一种基于目标检测框的局部匹配算法,并融合WLS滤波来平滑深度图,从而在保证视差图质量的同时有效提高匹配效率。双目相机使用ZED2i,并采用MATLAB完成标定^[17]。

针对传统SGBM^[18,19]算法在高分辨率场景下计算量过大、视差图中常出现空洞的问题,本文在

目标检测任务中仅关注检测框内的区域,结合局部立体匹配显著减少了计算量。同时,引入WLS^[20]滤波对视差进行优化,有效提升了视差图的连续性和边缘质量。其整体流程如图4所示。

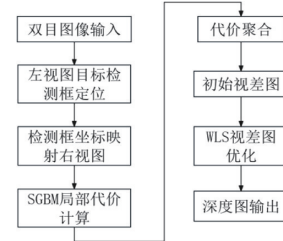


图4 WY-SGBM立体匹配流程图

本文提出的WY-SGBM立体匹配算法以双目相机的左右视图作为输入。首先,对左视图进行目标检测,获得车辆边界框坐标,并对检测框区域进行适度扩展,以保留车辆周围的上下文信息。针对边界框扩展比例,经实验验证,5%扩展会导致边缘纹理缺失,而20%则引入过多背景噪声,故最终选取10%以平衡上下文信息的完整性与局部匹配的稳定性。随后,将扩展后的检测框坐标映射至右视图,在左右视图对应的局部区域进行立体匹配。SGBM算法在该局部区域内计算像素间匹配代价,并通过路径聚合等方法降低匹配歧义。基于聚合后的代价生成视差图,并进一步通过加权最小二乘滤波对视差进行平滑处理,以优化车辆边缘信息并抑制噪声,从而得到最终的车辆深度图。

加权最小二乘(Weighted Least Squares, WLS)滤波是一种非线性全局滤波方法,通过赋予不同像素不同的权重,在保留图像的边缘细节的情况下使图像更加平滑。具体的WLS滤波公式如下

$$\sum_p \{ (\mu_p - g_p)^2 + \lambda \sum_p \left[\alpha_x p(g) \left(\frac{\partial \mu}{\partial x} \right)_p^2 + \alpha_y p(g) \left(\frac{\partial \mu}{\partial y} \right)_p^2 \right] \} \quad (2)$$

式中, p 表示图像像素坐标, g_p 和 u_p 分别表示图像 g 和 u 中的在像素点 p 处的取值,式中第一项 $(u_p - g_p)^2$,用于约束输出图像与输入图像之间的差异,保证滤波结果能够保留输入图像的主要结构信息;第二项通过约束梯度项来实现平滑,从而在全局范围内抑制噪声与不连续性。平滑强度由引导图像 g 的系数 α_x, α_y 决定,根据局部内容自适应地调整滤波强度,从而在保留边缘细节的同时提升整体的平滑效果。

4 实验结果与分析

4.1 去雾检测模型

由于目前目标检测领域缺乏公开的雾天小目

标检测数据集，本文构建了Fog-A数据集。该数据集基于1000张涵盖不同能见度的航拍图像，经几何变换与合成雾等增强操作扩充至4000张，并对车辆目标进行了精确的人工标注。数据集按7:2:1划分，兼顾了雾浓度与目标特性的多样性。实验部分采用了Fog-A与VisDrone2019数据集验证改进模型的检测性能，并利用RESIDE-OTS数据集训练去雾网络，以提升其对室外场景的适应能力。

4.1.1 性能评估指标

为了综合比较所提出的去雾检测算法与原算法，选取mAP（Mean Average Precision）、FLOPs（Floating Point Operations）、FPS（Frames Per Second）和Params（Parameters）作为本实验结果的评价指标。

（1）mAP@0.5：表示在交并比（IoU）阈值为0.5时所有类别平均精度的均值。

（2）FPS：反映模型推理速度的指标。

表 1 消融实验结果对比

编号	GCAvNet	HorBlock	模型大小/MB	计算量FLOPs/G	帧率/FPS	平均精度 mAP/%
1	—	—	6.2	8.1	103.0	85.3
2	√	—	6.2	8.1	95.3	92.2
3	—	√	6.6	8.8	94.3	86.6
4	√	√	7.2	12.7	92.2	95.1

本研究的实验对比中，第一组为未改进的YOLOv8n模型，其在测试集上的mAP为85.3%。在此基础上，第二组在网络中加入GCAvNet去雾模块后，mAP提升至92.2%，表明GCAvNet能够通过学习雾图与清晰图之间的残差，有效去除图像中的雾气，从而显著提升了低能见度场景下的检测精度。第三组实验将Backbone中间部分的C2f模块替换为HorBlock模块，结果mAP提升至86.6%，验证了HorBlock借助递归门控卷积实现高阶空间交互，从而增强多尺度特征处理能力的优势。最后，在第四组中同时融合GCAvNet与HorBlock两种改进策略，模型mAP达到95.1%，相较于原始YOLOv8n提升了9.8个百分点，充分证明了所提出改进方法在保持检测速度的同时，能够显著提升整体检测性能。

4.1.4 不同数据集泛化能力实验

为了验证本文提出的YOLOv8n-ours改进模型在复杂场景下目标检测任务中的泛化能力，实验分别在自制的Fog-A雾天数据集和VisDrone2019航拍数据集上进行测试。通过对比原始YOLOv8n与改进后的YOLOv8n-ours模型在两个数据集上的mAP等指标，评估改进方法在不同场景下的检测性能和鲁棒性。实验结果如表2所示。

（3）Model Size：衡量模型的大小。

（4）FLOPs：评估一个网络结构的计算复杂度的指标。

4.1.2 实验平台与训练参数

本文实验均在Ubuntu 22.04系统下进行，硬件配备NVIDIA RTX 4090 GPU，软件环境基于PyTorch1.12.0与CUDA 11.3。训练参数设置如下：输入尺寸640×640，Batch-size设为16，迭代300轮；采用SGD优化器（动量0.937，初始学习率0.01，权重衰减0.0005）。此外，设定模型缩放因子（depth=0.33，width=0.25）并结合常规数据增强策略，以在优化显存占用的同时提升模型的泛化能力。

4.1.3 消融实验

为验证每种改进点对网络模型的提升，在自制数据集（Fog-A数据集）进行4组消融实验对比。环境参数设置均一致。“√”表示使用对应的改进点，实验结果如表1所示。

实验结果表明，YOLOv8n-ours模型在VisDrone2019和Fog-A两个复杂场景数据集上均表现出较强的泛化能力。在VisDrone2019数据集上，YOLOv8n-ours相较于原始YOLOv8n在mAP50指标上提升了6.1个百分点；在Fog-A数据集上，精度提升幅度更为显著，达到了9.8个百分点。虽然改进模型在参数规模和计算量方面相对增加，推理速度有所下降，但其在复杂环境下的检测精度提升更为明显，体现出较强的鲁棒性和实用价值。

表 2 不同数据集泛化实验结果对比

数据集	模型	模型大小/MB	计算量FLOPs/G	帧率/FPS	平均精度 mAP@0.5
Vis-Drone2019	YOLOv8n	6.2	8.1	103	32.3
	YOLOv8n-ours	7.2	12.7	92.2	38.4
Fog-A	YOLOv8n	6.2	8.1	103	85.3
	YOLOv8n-ours	7.2	12.7	92.2	95.1

4.1.5 不同算法对比实验

为了验证本文所提出改进YOLOv8n（下称YOLOv8n-ours）算法相对于其他算法的有效性以及优越性，本文将其与YOLOv3、Faster R-CNN、SSD、Mask R-CNN、ConvNext、YOLOv10n和

YOLOv11n 等模型在 Fog-A 数据集上进行对比实验，实验结果如表 3 所示。

表 3 不同算法实验结果对比

Algorithms	模型大小/MB	帧率/FPS	平均精度 mAP/%
Faster—RCNN	81.0	22.4	65.8
SSD	52.2	65.3	61.5
YOLOv3	123.5	64.3	73.0
YOLOv4s	18.6	76.5	74.5
Mask R—CNN	74.2	60.4	72.2
YOLOv5n	3.8	81.5	84.0
ConvNext	135.4	75.6	82.1
YOLOv8n	6.2	103	85.3
YOLOv10n	5.8	78	89.2
YOLOv11n	5.5	112	88.4
YOLOv8n—ours	7.2	92.2	95.1

如表 3 所示，本文提出的 YOLOv8n-ours 在 mAP 指标上表现最佳，达到 95.1%，显著优于 Faster R-CNN、Mask R-CNN 及 YOLOv11n 等对比算法。在模型复杂度方面，尽管引入改进模块使得参数量略高于原始 YOLO 系列，但仍远低于 Faster R-CNN 等传统模型，保持了轻量化与易部署的优势。在推理效率上，该算法达到了 92.2FPS。虽然相比原始 YOLOv8n 略有下降，但该计算负载的增加对整体效率影响有限，仍能满足实时检测需求。首先，GCANet 去雾模块通过学习雾图与清晰图的残差，有效改善了低能见度下的图像质量，增强了目标的边缘与纹理信息；其次，HorBlock 多尺度融合模块提升了特征表达能力，解决了密集与大尺度差异目标的检测难题。两者的协同作用显著增强了模型在复杂雾天场景下对小目标及弱对比度目标的检测鲁棒性。

4.2 滑动窗口的图像分割模型

为验证滑动窗口分割检测模型性能，本次实验所用数据集为 VisDrone2019，数据集共计 10209 张静态图像。本文将按照 7:2:1 划分为训练集、验证集与测试集。

为验证改进点对网络模型的提升，对自制高分辨航拍数据集进行 4 组消融实验对比。环境参数设置均一致。其中 NMS-C 是修改原 IoU 为 CIoU 的 NMS 算法。消融实验的结果见表 4，“√”表示使用对应的改进点。

表 4 消融实验结果对比

编号	改进 YOLOv8n	NMS—C	平均精度 mAP/%
1	—	—	52.3
2	√	—	72.7
3	—	√	68.6
4(本文)	√	√	76.7

如表 4 所示，以原始 YOLOv8n (mAP 52.3%) 为基准，改进后的 YOLOv8n 检测模型将 mAP 提升至 72.7%，验证了其对于低能见度小目标的识别增强能力；NMS-C 策略的引入使 mAP 提升了 16.3 个百分点，有效解决了密集目标的误抑制问题。最终，两者融合使得 mAP 较基准提升 24.4 个百分点，充分证明了特征提取优化与高效抑制策略之间的协同增益。

为了直观展示各改进对检测性能的影响，本文将实验结果进行了可视化呈现。图 5 展示了原始 YOLOv8n 在测试图像上的检测结果；图 6 展示了先使用本文提出的滑动窗口方法对图像进行分割，再使用改进 YOLOv8n 进行检测后的结果，图 7 为局部放大对比图。



图 5 YOLOv8n 的检测结果



图 6 改进后的检测结果



图 7 检测结果对比图

如图 7 所示，7 (a) 中原始 YOLOv8n 因局部捕捉能力不足导致在大幅图像中产生误检，引入滑动窗口策略后有效降低了背景混淆；7 (b) 中针对原始模型漏检的微小行人，改进后模型通过增强多尺度特征表征实现了精准识别；7 (c) 中面对树枝遮挡场景，改进模型凭借更鲁棒的特征融合与上下文利用机制，有效解决了漏检问题，提升了复杂环境下的检测稳定性。

为了更全面地评估改进方法的有效性，将 YOLOv8n 改进模型与 DETR^[21]进行了对比实验。DETR 与改进模型的局部检测效果对比图如图 8 所示。



图 8 检测结果对比图

结果表明，DETR 在小目标检测与复杂背景判别上存在显著局限，如图 8 (a) 所示，由于全局特征表征较弱，微小行人目标被漏检；图 8 (b) 中将绿化带误判为车辆，则暴露了其局部判别力

的不足。这是因为 DETR 主干网络的多层下采样导致小目标信息稀疏，且全局自注意力机制在处理高分辨率图像时难以兼顾局部细节。相比之下，本文方法通过引入滑动窗口策略有效保留了局部高分辨率特征，并结合多尺度表示优化，显著增强了小目标在特征图中的显著性。同时，改进后的上下文利用机制有效抑制了背景干扰，从而在复杂场景下实现了更优的检测准确性与鲁棒性。

4.3 立体匹配实验

为验证改进算法的优越性，实验在距目标车辆 6m 处采集十组双目数据计算平均测距值，并对比了 SGBM、W-SGBM（引入 WLS 滤波）、Y-SGBM（基于检测框局部匹配）以及本文提出的 WY-SGBM 四种算法。图 9 展示了实验场景（左目图像），图 10 与图 11 分别呈现了 Y-SGBM 与 WY-SGBM 的局部匹配深度图效果。



图 9 左目图像

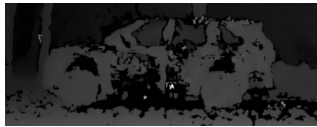


图 10 Y-SGBM 局部匹配深度图



图 11 WY-SGBM 局部立体匹配深度图

由图 10~11 实验效果图可以看出，经过 WLS 滤波后的深度图比原来的 SGBM 算法明显的要平滑很多，视差图中的空洞与无效像素得到了有效填充。而只计算局部区域的深度图则明显计算量会减少很多，具体实验数据如表 5 所示。

表 5 不同立体匹配算法性能对比

算法	匹配时间/s	测量距离/m	实际距离/m	误差/%
SGBM	1.12	5.85	6.0	2.5
W-SGBM	2.49	5.89	6.0	1.8
Y-SGBM	0.15	5.84	6.0	2.7
WY-SGBM	0.33	5.89	6.0	1.8

如表 5 所示，在 2208×1242 分辨率下，各算法性能差异显著。传统 SGBM 耗时 1.12s（误差 2.5%），实时性不足；W-SGBM 引入 WLS 滤波虽将误差降至 1.8%，但因计算开销增加，耗时延长至 2.49s。

相比之下，Y-SGBM 利用检测框剔除冗余区域，耗时仅 0.15s，大幅提升了计算效率。本文提出的 WY-SGBM 算法结合了两者的优势，在检测框约束下引入滤波优化，不仅将匹配时间缩短至原始 SGBM 的约 1/3，且精度优于 Y-SGBM 与传统 SGBM，实现了速度与精度的最佳平衡。

4.4 综合实验

为了验证本文方法的有效性，设计了一项整合滑动窗口、改进 YOLOv8n 及 WL-SGBM 算法的综合实验。该实验旨在评估多模块协同下，各改进点在复杂场景中对检测与深度估计精度及鲁棒性的提升效果。检测结果如图 12、13 所示，测距结果表 6 所示。

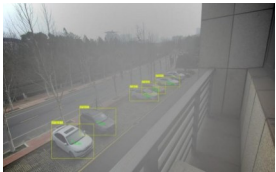


图 12 YOLOv8+SGBM 检测效果图



图 13 改进后 YOLOv8+WY-SGBM 检测效果图

表 6 测距实验结果对比

车辆编号	真实距离/m	改进前 测量距离/m	改进后 测量距离/m	相对 误差/%
1	13.9	14.5	14.3	2.8
2	15.2	15.8	15.6	2.6
3	18.4	19.6	19.2	4.3
4	19.7	21.0	20.7	5.0
5	22.1	23.2	23.1	4.5
6	23.0	—	24.2	5.2
7	23.8	—	25.2	5.8
8	24.5	—	26.5	8.1

从实验结果图可以看出在改进前的实验中，针对图像中的远处车辆和部分遮挡的车辆，模型未能有效检测，部分目标出现漏检。经过改进后的 YOLOv8n 联合 WY-SGBM 方法处理后，这些原本漏检的远处车辆及复杂场景下的目标均被成功检测，且整体检测置信度明显提高。

从表 6 中可以看出，改进前模型对于近处车辆，测量距离存在一定偏差，相对误差在 2.6~5.0% 之间。经过本文提出的改进方法，所有车辆均被成功测距，测量结果更加准确，相对误差控制在 2.6~8.1% 之间。远处车辆相对误差略高，主要由于距离增加导致深度图像细节减少，特征提

取与匹配难度加大,但改进方法通过检测框约束和WLS滤波增强了远处目标的可测性和匹配精度,从而显著提升整体测量效果。

5 总 结

本文针对高分辨率图像的目标检测与测距挑战,提出了融合改进YOLOv8n与WY-SGBM的解决方案。通过引入GCANet与HorBlock模块,算法在Fog-A和VisDrone2019数据集上的mAP分别提升9.8%和6.1%;结合滑动窗口策略与局部匹配算法,有效保留了图像细节,并将立体匹配耗时缩短至原算法的1/3,精度提升0.7%,实现了精度与实时性的良好平衡。

尽管效果显著,但滑动窗口策略增加了计算开销,且模型在极端光照或严重遮挡场景下的鲁棒性仍有不足。未来工作将集中在两方面:一是通过TensorRT加速、量化与剪枝优化计算效率;二是引入Transformer全局建模及自监督时序信息,以进一步提升复杂动态环境下的适应能力。

参考文献:

- [1] 刘家晨. 高分辨率图像中的小目标识别算法研究与实现[D]. 武汉:华中科技大学,2020.
Liu Jiachen. Research and implementation of small object detection algorithms in high-resolution images[D]. Wuhan:Huazhong University of Science and Technology,2020.
- [2] 周璇,葛琦,邵文泽. 高分辨率特征增强的无人机航拍小目标检测[J]. 数据采集与处理,2024,39(04):908—921.
Zhou Xuan, Ge Qi, Shao Wenze. Small object detection in UAV aerial images with high-resolution feature enhancement [J]. Data Collection and Processing, 2024, 39(04):908—921.
- [3] Hoanh N, Pham T V. Focus-attention approach in optimizing detr for object detection from high-resolution images[J]. Knowledge-Based Systems, 2024, 296:111939.
- [4] 尚柯言. 基于深度学习的无人艇海面场景目标检测方法研究[D]. 长沙:湖南大学,2023.
Shang Keyan. Research on target detection methods for unmanned surface vehicle maritime scenes based on deep learning [D]. Changsha:Hunan University,2023.
- [5] Zhou Y, Jing W, Wang J, et al. MSAR-DefogNet: Lightweight cloud removal network for high resolution remote sensing images based on multi scale convolution[J]. IET Image Processing, 2022, 16(3):659—668.
- [6] Wang F, Wang X. HSF-YOLO: a multi-scale and gradient-aware network for small object detection in remote sensing images[J]. Sensors (Basel, Switzerland), 2025, 25(14):4369.
- [7] 原铭,李凡,李华锋,等. 基于多级注意力密集残差网络的单幅图像超分辨率重建[J]. 光学技术,2022,48(03):357—363.
Yuan Ming, Li Fan, Li Huafeng, et al. Single-image super-resolution reconstruction based on multi-level attention dense residual network[J]. Optical Technique, 2022, 48(03):357-363.
- [8] 李新宇. 基于深度学习的高分遥感图像建筑物变化检测研究[D]. 大连:大连交通大学,2025.
- Li Xinyu. Research on building change detection in high-resolution remote sensing images based on deep learning[D]. Dalian:Dalian Jiaotong University,2025.
- [9] Du D, Zhu P, Wen L, et al. VisDroneDET2019:The vision meets drone object detection in image challenge results[C]//Proceedings of The IEEE/CVF International Conference on Computer Vision Workshops. Long Beach, USA:IEEE,2019:213—216.
- [10] Bodla N, Singh B, Chellappa R, et al. Soft-NMS-improving object detection with one line of code[C]//Proceedings of The IEEE International Conference on computer Vision. Venice, Italy:IEEE,2017:5561—5569.
- [11] Du S, Zhang B, Zhang P, et al. An improved bounding box regression loss function based on CIOU loss for multi-scale object detection[C]//2021 IEEE 2nd International Conference on Pattern Recognition and Machine Learning (PRML). Chengdu, China:IEEE,2021:92—98.
- [12] Zheng Z, Wang P, Liu W, et al. Distance-IoU loss:Faster and better learning for bounding box regression[C]//Proceedings of The AAAI Conference on Artificial Intelligence. New York, USA:AAAI,2020:12993—13000.
- [13] Zhai H, Cheng J, Wang M. Rethink the IoU-based loss functions for bounding box regression[C]//2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC). Chongqing, China:IEEE,2020:1522—1528.
- [14] Das S, Islam M S, Amin M R. Gca-net:utilizing gated context attention for improving image forgery localization and detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA:IEEE,2022:81—90.
- [15] Rao Y, Zhao W, Tang Y, et al. Hornet:Efficient high-order spatial interactions with recursive gated convolutions[J]. Advances in Neural Information Processing Systems, 2022, 35:10353—10366.
- [16] Wang J, Hu X. Convolutional neural networks with gated recurrent connections[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(7):3421—3435.
- [17] 许鸿萍. 基于机器视觉的工业机器人相机标定方法研究[J]. 现代制造技术与装备,2024,60(12):209—211.
Xu Hongping. Research on camera calibration method of industrial robots based on machine vision[J]. Modern Manufacturing Technology and Equipment, 2024, 60(12):209—211.
- [18] 吕钦,周诗洋,万波为,等. 基于双目视觉的铝合金件三维点云重建研究[J]. 组合机床与自动化加工技术,2025,(02):63—68.
Lv Qin, Zhou Shiyang, Wan Bowei, et al. Research on 3D point cloud reconstruction of aluminum alloy parts based on binocular vision[J]. Modular Machine Tool & Automatic Manufacturing Technique, 2025, (02):63—68.
- [19] Liu Y, Huang W, Wang Y. A new stereo matching method for RAW image data based on improved SGBM[C]//Three-Dimensional Image Acquisition and Display Technology and Applications. Beijing, China:SPIE,2018,10845:208—214.
- [20] Cai Zhenhao, Yu Hongyi. Optimal design of hybrid filter bank based on weighted least squares regularization method[J]. Journal of Signal Processing, 2010, 26(7):1110—1114.
- [21] Zhu X, Su W, Lu L, et al. Deformable DETR: deformable transformers for end-to-end object detection[C]//International Conference on Learning Representations Vienna, Austria:Open-Review.net.2021:1—16.