WEMD–YOLOv11n：融合频域增强和动态感知的小目标检测算法

闫建红[1] 王嘉乐[1]
YAN Jianhong WANG Jiale

摘 要 随着无人机、固定摄像头等多源监测设备普及，采集图像分辨率差异显著增大，其中小目标物体的监测常面临背景噪声、纹理模糊、漏检率高等困扰。因此，文章提出一种改进的 WEMD-YOLOv11n 小目标检测算法。首先，在 C3k2 模块中引入小波卷积，利用频域多频段分解强化不同分辨率下的小目标特征；其次，结合 EMA 注意力机制引导跨尺度跨层次特征融合，构建更高分辨率的三层检测输出，减少低分辨率状态下小目标细节信息丢失；最后，采用动态检测头提升不同分辨率下小目标的多维度感知及动态适应性。实验结果表明，改进的模型召回率、平均检测精度较 YOLOv11n 分别提升 0.8% 和 3.8%，参数量减少 32.6%。为跨分辨率场景下小目标监测需求提供了一种高精度、低资源消耗的解决方案。

关键词 YOLOv11；小波卷积；EMA 注意力；多尺度特征融合；动态检测头

doi：10.3969/j.issn.1672-9528.2026.02.001

0 引言

近年来，无人机、智能监控等监测设备的普及，全方位数字化监管技术得以广泛应用，但同时设备差异也造成了图像分辨率波动，如无人机高清影像与低分辨率监控画面，分辨率差异对其中的小目标检测带来了极大挑战。传统目标检测方法依赖手工特征提取，如尺度不变特征变换匹配算法 SIFT、方向梯度直方图 HOG 等，通过滑动窗口和非极大值抑制 NMS 筛选检测框改善检测性能，但泛化性差、效率低下。而基于深度学习的检测方法 [1-2] 因其强大的特征学习能力和适应性成为当前主流方法，具体又分为两个阶段（如 Fast R-CNN[3]、Faster R-CNN[4]）与单阶段算法，前者通过区域建议机制提升精度，但计算量大；后者如 SSD[5] 和 YOLO 系列 [6-10] 可以直接从输入图像中预测目标位置与类别，在实时性与精度间更加均衡，其中 YOLO 的结构相对简单，易于实现与训练，且是一个端到端的系统，其中 YOLOv11 采用了多项先进技术特点，如 Mosaic 数据增强、动态锚点调整等，进一步提升模型的检测精度和泛化能力，基于此，本文以 YOLOv11 算法为基础展开研究。

在多源设备监测场景下，小目标检测存在纹理模糊、语义信息易丢失、漏检率高等问题，已有研究提出了多种改进方法。李彬等 [11] 针对高分辨率影像中小目标特征不显著问题，通过改进 C3k2 模块扩大感受野、设计膨胀金字塔卷积以及增加小目标检测层来增强特征信息的融合能力和对不同尺寸目标及位置特征的关注度，较基线模型很好地平衡了精度和模型计算量。卢彬等 [12] 通过多尺度注意力 MSCA 模块以及通道 - 空间注意力机制抑制背景噪声干扰，在低分辨率图像下有助于突出小目标特征，但同时也带来了模型结构的复杂度和计算量上涨。Soudeep 等 [13] 在监控画面检测中通过 YOLOv11 和动态图神经网络的结合，聚合邻居节点的信息来建模目标物体的时空依赖关系，有效处理遮挡、轨迹重叠等复杂情况。但图构建和更新的复杂性以及对极端天气下检测的泛化性仍有待提高。Huang 等 [14] 在遥感图像下的船舶检测中通过引入轻量化卷积模块 GSConv 以及多尺度滑动窗口膨胀注意力机制解决了小目标特征稀疏性问题，但模型结构复杂度增加。

综合上述研究，现有改进方案多聚焦单一维度优化，尤其在跨设备、跨尺度场景下的小目标检测性能仍存提升空间。因此，面向跨分辨率图像检测，本文提出改进模型 WEMD-YOLOv11n 小目标检测算法。

1 YOLOv11 网络架构

YOLOv11 的网络架构主要分为三个部分：主干网络 Backbone、颈部网络 Neck 和检测头 Head，其网络结构如

1. 太原师范学院计算机科学与技术学院 山西晋中 030619
[基金项目] 山西省科技战略研究专项重点项目（202304031401011）；山西省重点研发计划项目（202102010101008）；山西省研究生精品教学案例项目（2024AL27）

图 1 所示。Backbone 采用 C3k2 模块优化特征信息的提取和传递，使用空间金字塔快速池化 SPPF 模块以多尺度汇集特征，并采用 C2PSA 模块，通过空间注意力机制提高模型对图像中重要区域的关注度。Neck 则将 FPN（feature pyramid networks）和 PAN（path aggregation network）结构进行特征融合，增强浅层特征的表达能力和不同层次之间信息交流效率。Head 采用深度可分离方法减少冗余计算，并使用多尺度预测头检测不同大小的物体。

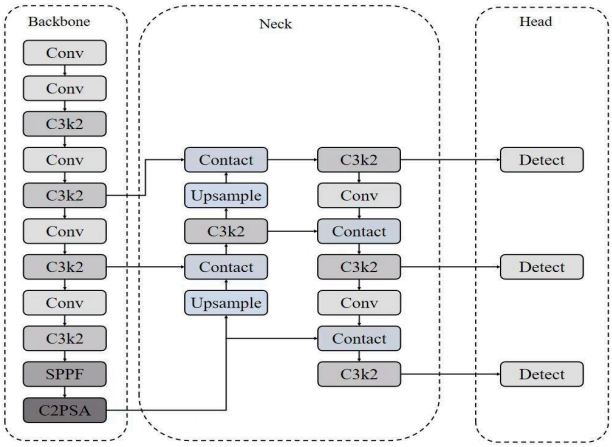

图 1 YOLOv11 网络结构图

2 WEMD-YOLOv11n 网络架构

YOLOv11n 在跨分辨率小目标检测中，面临传统卷积对不同分辨率背景噪声干扰、特征融合中低分辨率小目标细节丢失以及检测头对分辨率差异的敏感性等问题。因此，本文利用小波卷积抑制分辨率差异引起的背景干扰，结合高效多尺度注意力提升低分辨率图像中小目标细节保留能力，采用动态检测头增加对多分辨率场景下小目标的多维度动态感知。结构图如图 2 所示。

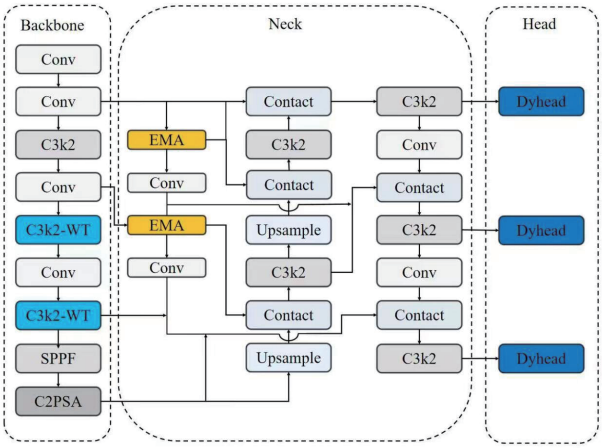

图 2 WEMD-YOLOv11 模型结构图

2.1 C3k2-WT：融合频域增强特征提取

在 YOLOv11 中，C3k2 模块受限于卷积核的大小，难以有效捕捉全局上下文信息，而往往不同成像分辨率造成的复杂背景容易对小目标检测带来干扰，为此，引入小波卷积层（wavelet transform convolution, WTConv）[15]，利用 Haar 小波变换，通过多频率响应扩展卷积感受野，捕捉不同分辨率下的低频信息。具体实现过程如图 3 所示。

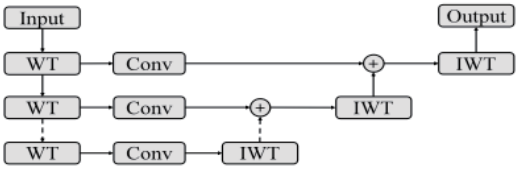

图 3 WTConv 结构图

首先，使用小波变换对输入图像进行滤波和下采样，然后再对生成不同频率图进行小核深度卷积之后，最后使用逆小波变换来构建输出。该过程可表示为：

$$Y = \mathrm{IWT}\left(\mathrm{Conv}\left(\boldsymbol{W},\mathrm{WT}\left(X\right)\right)\right) \tag{1}$$

式中：X 表示给定的输入；IWT 表示逆小波变换用来构建输出；$\boldsymbol{W}$ 表示一个 $k\times k$ 的深度卷积核的权重张量，其输入通道数是 X 的四倍；WT 表示对输入 X 进行一级 Haar 小波变换，通过深度卷积和标准的 2 倍下采样算子实现，执行过程中利用 1 个低通滤波器、3 个高通滤波器进行深度卷积，随后采用级联原理扩展并累计相加作为最终输出。

将小波卷积 WTConv 引入 Bottleneck 模块中得到改进的 C3k-WT 模块，如图 4 所示。小波卷积的引入不仅能使检测过程中更有效地结合不同图像分辨率下的局部和全局特征信息，而且通过对输入不同频率带的处理解决了传统卷积在实现大感受野时遇到的过度参数化问题，提高模型效率。

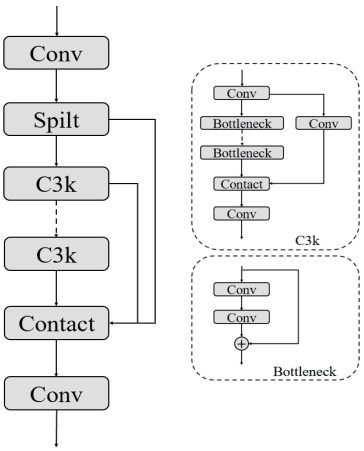

图 4 C3k2-WTConv 模型结构图

2.2 改进 Neck

在 YOLOv11 的特征融合网络结构中，虽然实现了多尺度、跨层级的特征交互，但面对跨分辨率图像时，对底层细节信息的利用不足，多次上下采样操作容易导致关键信息和细节的损失，这对于小目标检测及精确定位尤为关键。本文

对 Neck 进行改进，如图 5 所示，图 5（a）表示 Neck 改进前的 YOLOv11 网络结构；图 5（b）表示对 Neck 改进后的 YOLOv11 网络结构。

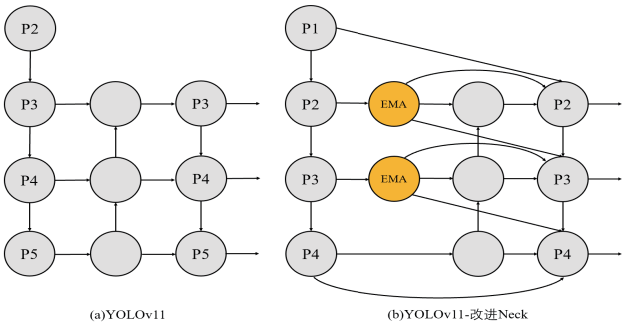

图 5 改进前后特征融合网络对比

首先，将特征提取网络中 P3~P5 层的输入调整为 P2~P4 层，同时删除主干网络 P5 层减少模型参数量；随后，利用高效多尺度注意力 EMA[16] 增强底层特征信息表达，通过对空间方向上的通道信息以及多尺度特征的捕获所组成的并行子网络来增加特征交互，增强小目标信息的表达能力；最后，引导 P2、P3 层进行跳跃连接和跨层聚合。

在此改进下可以减少深层网络特征信息传递，使得更多底层特征信息与部分深层特征进行级联融合，这样的多尺度融合思想与注意力机制的深度耦合减少了传统采样过程中信息的丢失同时进一步降低模型大小，实现了性能与效率的双重提升。

2.3 Dyhead：动态检测头

为了应对不同分辨率下的小目标特征差异，将传统检测头替换为动态检测头（dynamic head, Dyhead）[17]，如图 6 所示。模块将目标检测头的输入视为一个具有层级、空间、通道三个维度的三维张量，并在三个特定维度上分别部署注意力机制。

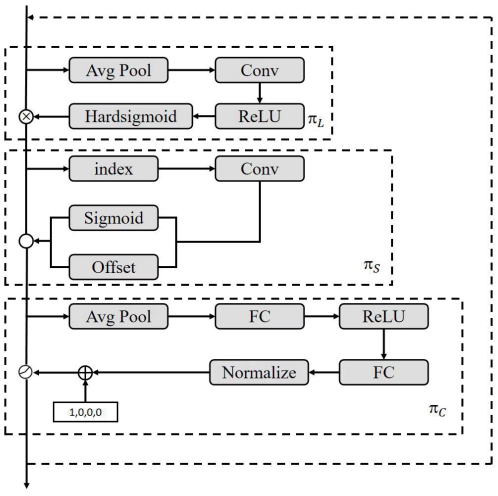

图 6 动态检测头模块结构图

首先是层级感知注意力，使用平均池化、1×1 卷积、ReLU 激活函数和硬 Sigmoid 函数，将特征金字塔缩放到相同的尺度来形成一个三维张量 $\boldsymbol{F}\in\mathbf{R}^{L\times S\times C}$ 作为动态头的输入。其次是空间感知注意力，包括偏移量学习和 3×3 卷积，专注于空间位置，通过聚合特征来关注图像中的区别性区域。最后是任务感知注意力模块，通过全连接层、ReLU 激活函数，以及归一化操作来处理，专注于通道，动态开启或关闭特征通道以支持不同的任务。具体注意力函数计算公式为：

$$\boldsymbol{W}(\boldsymbol{F})=\pi_C\left(\pi_S\left(\pi_L(\boldsymbol{F})\times\boldsymbol{F}\right)\times\boldsymbol{F}\right) \tag{2}$$

式中：$\pi_L(\cdot)$、$\pi_S(\cdot)$、$\pi_C(\cdot)$ 分别是应用于维度 L（金字塔层数）、S（$H\times W$）和 C（通道数）的 3 个不同注意力模块，三者串联构成单个动态检测头结构。

3 实验及结果分析

3.1 数据集

本文实验采用 MH-SODD 数据集验证模型改进后的有效性。多源异构小目标检测数据集（multi-source heterogeneous small object detection dataset, MH-SODD）是通过网络搜集与中国水利水电科学研究院公开数据集 [18] 构建，包含 3 000 张图像，图像采集包括手机、相机、无人机等设备，涵盖了 1 920 px×1 080 px、3 840 px×2 160 px、5 280 px×2 970 px 等多种图像分辨率，目标类别为 floater，数据集按 7:2:1 的比例划分为训练集、验证集、测试集用于实验。

3.2 实验环境

实验使用 GPU 为 NVIDIA GeForce RTX 4090 24 GB，操作系统为 Ubuntu 22.04，基于 Python3.10.14，内存为 60 GB，开发框架为 PyTorch2.2.0-Ubuntu 22.04，训练轮次为 200，batch-size 为 32，选用 SGD 优化器，初始学习率为 0.01，momentum 为 0.937。

3.3 模型评价指标

文中使用精度（precision, P）、召回率（recall, R）、平均精度均值（mean average precision, mAP）评价模型的检测效果，Params 评价模型的参数大小，FLOPs（floating point operations）是模型在执行一次前向传播过程中所需的浮点操作次数。相关计算公式为：

精确率表示模型预测正确的正样本准确度，即：

$$P=\frac{\mathrm{TP}}{\mathrm{TP+FP}} \tag{3}$$

召回率表示预测正确的正例数量占总体正例数量的比例，即：

$$R=\frac{\mathrm{TP}}{\mathrm{TP+FN}} \tag{4}$$

mAP 是对所有类别的平均精确率（average precision,

AP）取平均值，即：

$$\mathrm{mAP}=\frac{1}{n}\sum_{i=1}^{n}\int_{0}^{1}P(R)\mathrm{d}R \tag{5}$$

3.4 实验结果与分析

3.4.1 消融实验

为了验证改进对模型性能提升效果，进行递进式消融实验，实验结果如表 1 所示。

表 1 消融实验

Method	*P*/%	*R*/%	mAP50/%	mAP50-95/%	Para/10^6
YOLOv11n	84.1	65.6	75.3	50.8	2.58
YOLOv11n +C3k2-WT	81.6	67.6	76.6	51.6	2.52
YOLOv11n + 改进 Neck	79.5	67.7	77.0	53.1	1.19
YOLOv11n +Dyhead	80.9	67.6	76.5	51.7	3.09
YOLOv11n +C3k2-WT + 改进 Neck	81.0	67.7	77.8	52.8	1.13
YOLOv11n +C3k2-WT +Dyhead	82.6	67.8	77.6	52.8	3.04
YOLOv11n + 改进 Neck +Dyhead	82.5	66.4	77.9	53.7	1.76
Ours	83.8	66.4	79.1	54.6	1.74

C3k2-WT 和 Dyhead 的结合最大提升了模型对漏检目标的捕捉能力，召回率较基准模型提升 2.2%，虽然牺牲了部分误检率，但表明三层注意力机制的加入增强了模型不同分辨率下低频分量的捕捉能力，更关注漏检抑制，C3k2-WT 和 Neck 改进的结合使得模型参数量达到最低，下降 56.2%，此时平均检测精度 mAP50 较基准模型依旧增长 2.5%，并且优于单个改进的平均检测精度，频域增强策略和高效多尺度融合机制的结合在实现模型轻量化的同时保持检测精度稳步上升，在总的改进实验中 WEMD-YOLOv11n 的平均检测精度达到了最高，较基准模型 YOLOv11n 提升 3.8%，此时参数量下降 32.6%，表明文中的改进算法在降低模型大小的同时，实现了漏检率的降低和平均检测精度提升的平衡。

3.4.2 模型可视化

为了更直观地观察模型改进的提升，选取四个场景下检测效果进行对比，场景一至四分别来自手机、固定监控、高速相机、无人机设备采集，图像分辨率依次从低到高排列，如图 7 所示。

无论是低分辨率图像还是高分辨率图像，改进模型整体目标检测置信度都有了很大提升，并且在固定监控和无人机监测画面中，可以明显看到漏检率的降低。可视化结果表明改进模型在不同分辨率图像的检测任务中，漏检率的减少和特征融合能力及空间定位精度方面均表现出显著优势。

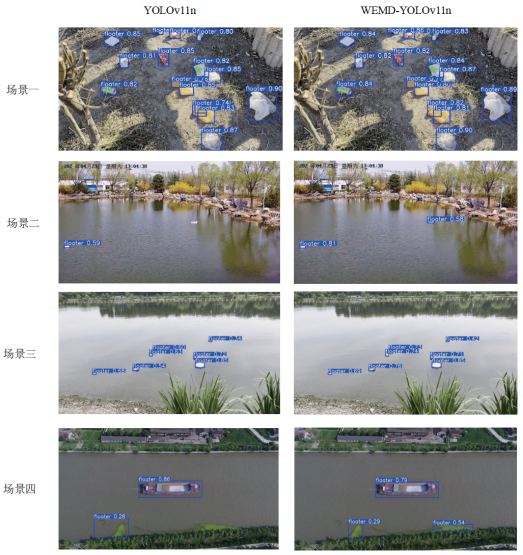

图 7 各类场景下检测效果对比

3.4.3 对比实验

为了验证改进模型的综合性能将其与其他主流算法进行了对比，结果如表 2 所示。

表 2 对比试验

Method	mAP50/%	Para/10^6	FLOPs/10^9
Faster R-CNN	63.7	136.6	401.7
SSD	49.9	23.61	60.75
YOLOv5s	76.7	9.12	23.8
YOLOv8n	76.1	3.01	8.1
YOLOv11s	76.2	9.41	21.3
YOLOv11n	75.3	2.58	6.3
Ours	79.1	1.74	17.2

相较于传统两阶段算法 Faster R-CNN，文中改进模型在 mAP50 上提升 15.4%，模型整体复杂度急剧下降，表明改进模型在精度与效率上均显著优于传统方法。与单阶段算法 SSD 相比，模型 mAP50 提升 29.2%，模型参数量和计算复杂度依旧是最低。针对 YOLO 系列算法，改进模型同样展现出显著优势，相较于 YOLOv5s，mAP50 提升 2.4%，参数量减少 80.9%，与轻量化模型 YOLOv8n 相比，mAP50 提升 3%，参数量下降 42.2%。尽管改进模型的 FLOPs 较 YOLOv11n 有所增加，但相对于传统的两阶段算法 Faster R-CNN、单阶段算法 SSD，YOLOv5s、YOLOv11s 是最低的，并且参数量仅为 1.74×10^6。

4 结论

本文针对不同分辨率图像中的小目标检测难题，提出了一种融合频域增强与动态感知的 WEMD-YOLOv11n 模型。在特征提取深层网络中嵌入小波卷积，通过频域分解实现全局 - 局部特征协同表达；设计跨层特征融合机制，动态平衡浅层纹理与深层语义信息；采用 Dyhead 弱化分辨率差异带来的定位干扰。实验表明，改进模型平均检测精度 mAP50 达到 79.1%，参数量降低 32.6%。未来工作将探索模型在极端天气（如雨、雾）下的适应性，并进一步压缩计算开销以满足边缘设备部署需求。

参考文献：

[1] 郭庆梅 , 刘宁波 , 王中训 , 等 . 基于深度学习的目标检测算法综述 [J]. 探测与控制学报 , 2023,45(6):10-20.

[2]FAN B, SONG X N. A review of deep learning based object detection algorithms[J].Journal of engineering research and reports, 2024, 26(11): 88-99.

[3]GIRSHICK R. Fast R-CNN[C]//Proceedings of the 2015 IEEE International Conference on Computer Vision. Washington: IEEE Computer Society, Piscataway: IEEE, 2015: 1440-1448.

[4]REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towardsreal-time object detection with region proposal networks[J].IEEE transactions on pattern analysis and machine intelligence, 2017, 39(6): 1137-1149.

[5]LIU W, ANGUELOV D, ERHAN D, et al. SSD: single Shot multibox detector[C]// Proceedings of European Conference on Computer Vision. Berlin:Springer,2016: 21-37.

[6]REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C] //2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 779-788.

[7]REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL].(2018-04-08)[2025-12-30].https://pjreddie.com/media/files/papers/YOLOv3.pdf.

[8]LI C Y, LI L L, JIANG H L, et al. YOLOv6: a single-stage object detection framework for industrial applications[EB/OL].(2022-09-07)[2025-12-30].https://github.com/meituan/YOLOv6.

[9]WANG C Y, BOCHKOVSKIY A, LIAO H-Y M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for realtime object detectors[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition.Piscataway: IEEE, 2023: 7464-7475.

[10]REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE,2017: 7263-7271.

[11] 李彬 , 李生林 . 改进 YOLOv11n 的无人机小目标检测算法 [J]. 计算机工程与应用 ,2025,61(7):96-104.

[12] 卢彬 , 王俊 . 基于 YOLOv11 改进的无人机图像小目标检测模型 [J]. 信息与电脑 (理论版),2024,36(22):150-152.

[13]SOUDEEP S, MRIDHA M F, JAHIN M A, et al. DGNN-YOLO: dynamic graph neural networks with YOLO11 for small object detection and tracking in traffic surveillance[EB/OL]. (2024-11-26)[2025-12-30].https://arxiv.org/html/2411.17251v1.

[14]HUANG J W, WANG K B, HOU Y, et al. LW-YOLO11: a lightweight arbitrary-oriented ship detection method based on improved YOLO11[J]. Sensors, 2024, 25(1): 65.

[15]FINDER S E, AMOYAL R, TREISTER E, et al. Wavelet convolutions for large receptive fields[C]//European Conference on Computer Vision. Berlin:Springer, 2024: 363-380.

[16]OU-YANG D L, HE S, ZHANG G Z, et al. Efficient multi-scale attention module with cross-spatial learning[C]// ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE, 2023: 1-5.

[17]DAI X Y, CHEN Y P, 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. Piscataway: IEEE, 2021: 7373-7382.

[18] 杨明祥 , 乔广超 , 王浩 , 等 . 水面漂浮物数据集 (IWHR_AI_Lable_Floater_V1)[R]. 北京 : 中国水利水电科学研究院 , 2023.

【作者简介】

闫建红（1972—），女，山西太原人，博士，教授，研究方向：机器学习、计算机视觉，email: yan_jian_hong@163.com。

王嘉乐（1999—），男，山西长治人，硕士研究生，研究方向：目标检测，email:2540747073@qq.com。

（收稿日期：2025-07-24 修回日期：2026-01-30）