中图法分类号:TP18;TP391.4　　**文献标识码**:A　　**文章编号**:1006-8961(2026)10-3691-13

论文引用格式:Guo S H, Lu X G and Wang Y. 2026. Multi-scale frequency-domain adaptive network for aerial small object detection. Journal of Image and Graphics, 31(10):3691-3703(郭宋浩, 卢晓光, 王阳. 2026. 航拍小目标检测的多尺度频域自适应网络. 中国图象图形学报, 31(10):3691-3703)[DOI:10.11834/jig.250462]

航拍小目标检测的多尺度频域自适应网络

郭宋浩[1],卢晓光[1*],王阳[2]

1. 天津市智能信号与图像处理重点实验室,天津　300300; 2. 昆山鲲鹏无人机科技有限公司,苏州　215300

摘　要:**目的**　针对无人机航拍图像中目标检测面临复杂背景、微小目标和障碍物等挑战,提出了一种多尺度频域自适应网络(multi-scale frequency adaptive you only look once,MSFA-YOLO)。**方法**　首先,构建的多尺度边缘特征提取器(multi-scale edge feature extractor,MEF)将图像多尺度边缘特征信息与原始特征跨通道融合增强局部细节的捕捉能力;其次,频域空间自适应模块(frequency spatial adaptive,FSA)结合空间域和频率域的联合表示,进一步丰富小目标特征信息;最后,轻量级特征增强模块(lightweight feature fusion enhancement,LFFE)采用多分支结构和可重参数化卷积技术,在保持特征表示能力的同时显著降低计算复杂度。**结果**　在VisDrone2019-DET(vision meets drones 2019 for detection for validation)和UAVDT(unmanned aerial vehicle benchmark: object detection and tracking)公开数据集上的实验结果表明,MSFA-YOLO与基线相比mAP_{50}(mean average precision)分别提高了4.3%和2.1%;与Hpyer-YOLO、Mamba-YOLO等先进算法方法相比,在参数量更低的同时mAP_{50}分别提高了4.4%、1.7%和1.4%、0.6%,进一步验证了所提出模块的有效性和泛化能力。**结论**　本文提出的MSFA-YOLO算法能够在复杂背景的无人机航拍小目标检测场景,相较于其他参与对比的模型,取得更加良好的性能。

关键词:小目标检测;航拍图像;YOLO11;多尺度边缘特征;特征增强

Multi-scale frequency-domain adaptive network for aerial small object detection

Guo Songhao[1], Lu Xiaoguang[1*], Wang Yang[2]

1. Tianjin Key Laboratory of Intelligent Signal and Image Processing, Tianjin　300300, China;

2. Kunshan Kunpeng UAV Technology Co., Ltd., Suzhou　215300, China

Abstract: **Objective** Small-object detection in unmanned aerial vehicle (UAV) imagery remains one of the most challenging tasks in computer vision and is particularly important for surveillance, search and rescue, traffic monitoring, and precision agriculture. UAV images present several intrinsic difficulties: small targets often occupy fewer than 32 × 32 pixels in high-resolution images, exhibit low contrast against complex backgrounds, suffer from motion blur and varying illumination, and show extreme scale variation because of changing flight altitude. Traditional object detectors, which are primarily designed for ground-level imagery with relatively large objects, do not address these challenges well, because conventional convolutional neural networks progressively lose fine-grained spatial information through pooling and strided convolutions. **Method** To overcome these limitations, we propose multi-scale frequency adaptive-you only look once(MSFA-YOLO), an

收稿日期:2025-09-23;**修回日期**:2026-03-18;**预印本日期**:2026-03-25

*通信作者:卢晓光　xglu@cauc.edu.cn

基金项目:国家重点研发计划资助(2023YFB3907003)

Supported by: National Key R&D Program of China (2023YFB3907003)

innovative detector that integrates multi-scale feature extraction, frequency-domain analysis, and efficient feature enhancement. The architecture includes three key components. First, the multi-scale edge feature extractor module preserves and enhances edge information that is crucial for small-object detection. It processes input features through parallel pathways at multiple scales, applies the Sobel operator for edge extraction, and then fuses the resulting edge features with the original features through an adaptive cross-channel attention mechanism that dynamically adjusts their relative contributions. Second, the frequency-spatial adaptive module exploits the complementary nature of spatial and frequency-domain representations. It applies the fast Fourier transform to decompose spatial features into frequency components and uses a dual-branch design in which the spatial branch preserves location-specific information through depthwise separable convolutions, while the frequency branch applies learnable filters to selectively enhance relevant frequency bands. An adaptive gating mechanism then combines the two branches. Third, the lightweight feature fusion enhancement module addresses computational efficiency through a multi-branch structure with different kernel sizes and re-parameterizable convolutions, which are merged during inference without performance loss. Group convolution and channel shuffling are further introduced to reduce parameter count while maintaining representational capacity. **Result** Comprehensive experiments were conducted on two challenging UAV datasets: VisDrone2019-DET, which contains 10 209 images with 54 200 annotated instances from ten object categories captured under diverse conditions, and UAVDT, which includes 80 000 frames with three vehicle categories and exhibits significant scale variation and occlusion. On VisDrone2019-DET, MSFA-YOLO achieved an mAP_{50} of 38.1% with only 3.08 M parameters, representing a 4.3% improvement over the YOLO11n baseline (33.8%) and outperforming state-of-the-art lightweight methods including LWUAVDet (37.4% with 9.7 M parameters), Mamba-YOLO (36.4% with 5.66 M parameters), LUD-YOLOn (35.2% with 2.81 M parameters), and Hyper-YOLO (33.7% with 2.68 M parameters). These results demonstrate a superior balance between efficiency and accuracy, with the highest detection accuracy among all compared methods while maintaining a compact model size. On UAVDT, MSFA-YOLO achieved an mAP_{50} of 32.3% and an $mAP_{50\text{-}95}$ of 19.3%, surpassing the YOLO11n baseline (30.2% mAP_{50} and 17.8% $mAP_{50\text{-}95}$) by 2.1% and 1.5%, respectively, while also outperforming Mamba-YOLO (31.7% mAP_{50} with 5.66 M parameters), Hyper-YOLO (30.9% mAP_{50}), and YOLO12n (30.6% mAP_{50}). Precision reached 45.1% and recall reached 31.2%, indicating robust detection performance across diverse scenarios. The proposed method achieves these results with only 3.08 M parameters, substantially fewer than Mamba-YOLO's 5.66 M and comparable to the most efficient baseline models, validating the effectiveness of the overall design. Ablation studies confirmed the contribution of each component: the MEF module provided clear gains through edge enhancement, the FSA module improved representation through frequency-spatial fusion, and the LFFE module optimized the efficiency-accuracy trade-off while reducing computational overhead. Cross-dataset evaluation further demonstrated strong generalization, as models trained on VisDrone2019-DET achieved robust performance on UAVDT without fine-tuning, suggesting effective feature learning and domain adaptation. **Conclusion** This work presents MSFA-YOLO as a novel detection framework that addresses the core challenges of small-object detection in UAV imagery through the synergistic integration of multi-scale edge features, frequency-spatial adaptive processing, and efficient architectural design. The approach significantly improves performance over existing methods while maintaining the computational efficiency required for real-world UAV deployment. Its effectiveness stems from three main insights: preserving and enhancing edge information through multi-scale Sobel processing is crucial for targets with very limited pixel support; frequency-domain analysis provides complementary fine-grained cues that are often lost in purely spatial processing; and careful architecture design can maintain a compact model size without sacrificing detection accuracy. Future work will focus on improving robustness under extreme conditions such as low illumination and adverse weather, extending the framework to video-based detection with temporal consistency constraints, exploring self-supervised pre-training for domain adaptation, and deploying the system on edge devices with hardware-specific optimizations. Overall, the proposed method represents a meaningful step toward practical, deployable UAV-based detection systems for challenging real-world environments.

Key words: small object detection; aerial imagery; YOLO11; multi-scale edge features; feature enhancement

0 引 言

近年来，无人机凭借其紧凑尺寸以及高机动性已成为设备巡检、灾害应急响应及军事侦察等复杂任务场景执行的关键手段。但是，无人机航拍图像因独特的空中视角特性，存在如目标尺寸小、遮挡以及复杂背景等特点，而现有目标检测方法依赖手工特征或浅层网络架构，难以达到高检测精度与资源利用之间的平衡。因此，针对此情形，设计解决跨尺度特征融合、轻量化与动态资源分配等问题的算法模型具有重要的理论意义和实际应用价值。

针对小目标的检测，已有大量研究开展，Zhu 等人（2020）提出融合目标感知非局部低秩建模与显著性滤波正则化的双约束框架，通过协同优化全局背景抑制与局部目标增强，显著提升复杂背景下小目标检测的鲁棒性与信杂比；Kou 等人（2022）结合改进的密度峰值全局搜索与人眼视觉局部对比机制，适用于单场景目标检测，但难以实现复杂背景下的目标检测。与上述传统的手工特征提取方法相比，基于深度学习的小目标检测技术已广泛使用。

金涛和胡配雨（2025）通过引入倒置残差级联模块（inverted residual cascade block，IRCB）构建轻量化主干网络，设计跨阶并行空洞融合网络结构（cross stage partial-parallel multi-atrous convolution，CSP-PMAC）以及采用小波下采样增强多尺度特征融合中的边界信息保留能力。郝川艳等人（2025）引入自适应切片辅助超推理（adaptive slicing-assisted hyper inference，ASAHI）框架，该框架根据图像分辨率动态调整切片数量，解决了在复杂的空中场景中检测小物体的关键挑战，有效平衡检测精度与计算效率。肖振久等人（2025）设计多尺度特征融合架构，使用渐进标准偏差的高斯核构建特征序列以保留小目标的语义信息，采用具有不同扩张率的级联扩张卷积来自适应捕获上下文信息，显著提升复杂遥感场景中的小目标检测性能。上述两阶段目标检测算法在精度的表现上较好，但是在需要实时检测的无人机环境下，模型小且检测速度快的 YOLO（you only look once）系列算法则更加适用。Peng 等人（2023）提出的 AMFLW（attention and multi-scale feature fusion lightweight）-YOLO 是一种轻量级遥感图像检测网络，通过采用深度可分离卷积和倒置残差线性瓶颈结构、引入坐标注意力机制、同时捕获通道间的方向感知和位置感知信息、实现高效的多尺度特征融合，提高遥感图像中不同尺度目标的检测效果。Zhang 等人（2024）提出的 FFCA（feature enhancement，fusion and context aware）-YOLO 结构，通过增强局部感知能力、优化多尺度特征融合和建立全局通道与空间关联，有效解决了遥感图像中小目标检测面临的特征表达不足和背景混淆问题。LSKF（large selective kernel feature fusion network）-YOLO（Shi 等，2024）利用上下文信息，设计多尺度特征对齐融合结构来融合低层语义信息并缓解深层网络特征模糊问题，引入 SPD（space to depth）-Conv 替代网络颈部后两层步卷积以增强小目标敏感性，提出 MPDIoU（minimum points distance intersection over union）损失函数优化定位效果。Hui 等人（2024）设计通过引入 SwinTransformer 模块与卷积神经网络（convolutional neural network，CNN）结合构建成新型卷积结构，设计轻量级分类器并将其整合到骨干网络中，与此同时采用无参数注意力机制 SimAM（simple parameter-free attention module）增强特征表示，最后使用新型检测头加强分类能力权重。Chen 等人（2025）设计在使用重参数化大核模块扩大网络的有效感受野并丰富了梯度流信息的同时，通过深度可分离卷积和特征提取模块减轻了大卷积核带来的计算复杂度，并引入 Priori Focal Loss 函数，动态调整损失权重来关注小目标和复杂样本。

虽然目前针对基于无人机小目标检测的方法已有很多，但仍然面临着特征表示不足、背景混淆等问题与挑战。针对这些挑战，本研究采用轻量级目标检测模型，设计多尺度边缘特征提取器（multi-scale edge feature extractor，MEF）、频域空间自适应模块（frequency spatial adaptive，FSA）和轻量级特征增强模块（lightweight feature fusion enhancement，LFFE），这些模块在稍微增加计算复杂度的情况下，充分利用局部和全局信息，进而有效地增强网络对小目标的感知。在具有挑战性的数据集上进行了广泛的实验，包括 VisDrone2019-DET（vision meets drones 2019 for detection for validation）、UAVDT（unmanned aerial vehicle benchmark：object detection and tracking），结果表明提出的方法 MSFA-YOLO（multi-scale frequency adaptive you only look once）与其他先进的方法相比具有非常具竞争性的性能。

本文主要贡献如下：1）提出多尺度边缘特征提取器MEF，该模块采用基于Sobel算子的梯度提取器生成多尺度边缘特征，并通过高效跨通道特征融合模块（cross-channel feature fusion block，EFB）实现边缘特征与卷积特征的整合，有效缓解了下采样过程中细粒度边缘信息的丢失问题。2）设计频域空间自适应模块FSA，该模块基于OmniKernel（Cui等，2024）架构并融合CSP（cross stage partial）思想，通过空间域和频率域的联合特征表示，结合多形态感受野设计，捕捉全面的特征信息，在保持计算效率的同时，有效提升对小目标的检测能力。3）构建轻量级特征增强模块LFFE，采用多分支并行处理架构和可重参数化卷积技术，在通过层级化特征提取增强小目标的特征表示的同时，应用通道尺度自适应机制优化计算资源分配。

1 多尺度频域自适应网络

在本节中，将详细介绍针对无人机小目标检测的MSFA-YOLO整体结构以及提出的3个改进模块：MEF、FSA和LFFE。

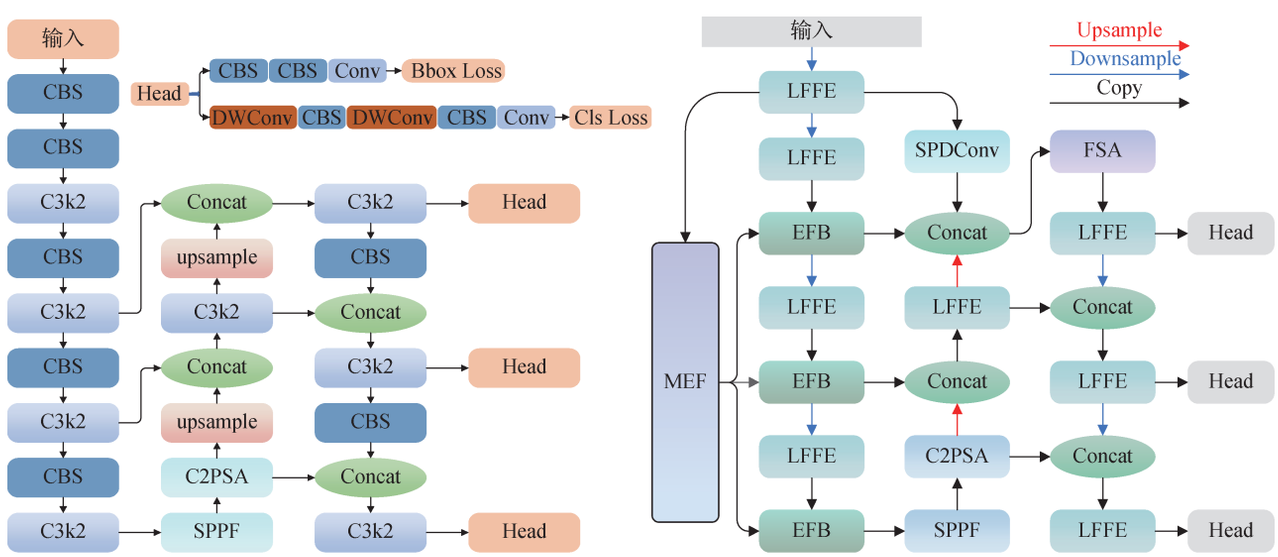
1.1 MSFA-YOLO结构

选择YOLO11作为本研究的基准框架，因为与YOLO12和YOLOv8等相比，它具有更小的参数量和计算量，并且在小目标检测任务中具有更高的准确性。基准模型YOLO11的框架图如图1（a）所示，改进模型MSFA-YOLO的框架如图1（b）所示。首先，MSFA-YOLO在原来YOLO11的主干上，在浅层P2添加MEF提取多尺度边缘信息，并通过EFB将边缘信息投放到主干的各个尺度中进行融合。其次，在颈部结构中使用P2特征层经过SPDConv（space to depth convolution）得到富含小目标信息的特征给到P3进行融合，融合后通过FSA进行小目标特征增强。最后，将主干和颈部的C3k2模块重构为轻量化的LFFE模块，该模块在保证精度不损失的同时，尽可能地减少了模型参数量。

1.2 多尺度边缘特征提取器MEF

物体框的精确定位高度依赖于物体的边缘信息，而传统的YOLO11通过连续的下采样操作逐步降低空间分辨率来提取特征，使得细粒度的细节和边缘信息会通过连续的下采样操作被系统性地稀释，最终导致对小目标识别至关重要的判别性特征丢失。为解决这些基本局限性，本文设计了多尺度边缘特征提取器MEF，其是一种新型主干架构增强方法，可以系统地重构特征提取管道，优先考虑边缘特征保存和多尺度信息融合。

由于原始图像中含有大量的背景信息，如果直接从原始图像上提取边缘特征信息传递到整个主干



（a）YOLO11框架结构图　　（b）MSFA-YOLO框架结构图

图1　模型结构图

Fig. 1　Model structure diagram ((a) YOLO11 structure diagram; (b) MSFA-YOLO structure diagram)

结构上，会导致给网络的学习带来过多的噪声，但是浅层的卷积层会过滤不必要的背景信息，因此MEF选择从浅层P2提取边缘特征信息，其结构图如图2所示。

首先，给输入的P2层特征图 $\boldsymbol{X} \in \mathbf{R}^{C \times W \times H}$ 进行边缘特征提取操作，具体为

$$\boldsymbol{E}(x) = \sqrt{\left(\boldsymbol{S}_x * \boldsymbol{X}\right)^2 + \left(\boldsymbol{S}_y * \boldsymbol{X}\right)^2} \tag{1}$$

式中，$\boldsymbol{S}_x$ 和 $\boldsymbol{S}_y$ 分别表示水平和垂直梯度提取的Sobel核，*表示卷积操作，而Sobel核定义为

$$\boldsymbol{S}_x = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}, \boldsymbol{S}_y = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \tag{2}$$

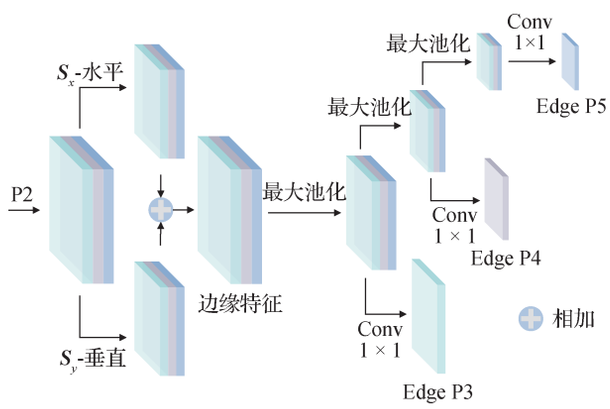


图2　MEF结构图

Fig. 2　MEF structure diagram

在得到边缘特征信息后，为生成多尺度的边缘特征表示，MEF对 $\boldsymbol{E}(x)$ 进行连续下采样处理，而为保留局部区域的最强特征，更好地体现边缘信息，采用最大池化（*Maxpool*）和1 × 1卷积的级联，因此对于输出的多尺度边缘特征 $\{\boldsymbol{E}_1, \boldsymbol{E}_2, \cdots, \boldsymbol{E}_n\}$ 为

$$\boldsymbol{E}_i = Conv_{1\times1}\left(Maxpool\left(\boldsymbol{E}_{i-1}(x)\right)\right) \tag{3}$$

式中，$Conv_{1\times1}$ 调整通道维度以匹配相应的特征图，*Avgpool* 更适用于需要平滑或均匀化特征的场景，但在保留细节和边缘信息方面的表现不如 *Maxpool*，并且通过级联递进式的结构选择，可以使得在保持计算效率的同时能够提取多尺度边缘信息。针对边缘特征与原始特征融合部分，设计了高效的跨通道特征融合模块EFB，其框架结构如图3所示，其数学表达式为

$$\boldsymbol{F}_{\text{fused}} = Conv_{1\times1}\left(Conv_{3\times3}\left(Conv_{1\times1}\left(\left[\boldsymbol{F} \oplus \boldsymbol{E}_i\right]\right)\right)\right) \tag{4}$$

式中，$\boldsymbol{F} \in \mathbf{R}^{C\times H\times W}$ 是原始特征图，⊕ 表示通道维度的拼接。

首先，使用 $Conv_{1\times1}$ 进行边缘特征信息与普通卷积特征的跨通道融合，帮助模型更好地整合不同来源的特征；然后采用 $Conv_{3\times3}$ 进一步提取融合后的特征，以增强模型对局部细节的捕捉能力；最后通过 $Conv_{1\times1}$ 调整输出特征维度，为后续的特征图传播做好铺垫。这种融合策略使网络能够利用标准卷积特征的语义信息和边缘特征的细粒度细节，从而获得更全面的特征表示，对小目标的检测有着显著的提升。

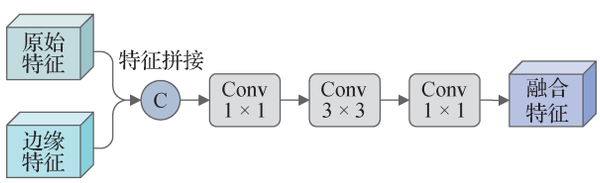


图3　EFB结构图

Fig. 3　EFB structure diagram

1.3　频率空间自适应模块FSA

小目标检测在标准特征金字塔网络的P3、P4、P5检测层上表现欠佳，传统方法通常通过引入P2检测层增强小目标的检测性能，然而此类方法不可避免地引入了计算复杂度增加和后处理延迟等一系列问题。相对于传统的添加P2检测层，本研究设计使用P2特征层经过SPDConv得到富含小目标信息的特征传到P3进行融合，再将融合后的特征经过频域空间自适应模块FSA进行增强，解决下采样过程中小目标关键信息丢失以及空间域特征提取忽略小目标独特的频率特性等问题，FSA的结构模型如图4所示，其中SFEM（spectral feature enhancement module）是频谱特征增强模块。

基于CSP思想对P3融合后得到的特征图 $\boldsymbol{X} \in \mathbf{R}^{C\times H\times W}$ 通过 $Conv_{1\times1}$ 卷积层处理，生成后经过 *Split* 分为两个分支的中间特征图，具体为

$$\left[\boldsymbol{X}_1, \boldsymbol{X}_2\right] = Split\left(Conv_{1\times1}(\boldsymbol{X})\right) \tag{5}$$

式中，$\boldsymbol{X}_1, \boldsymbol{X}_2 \in \mathbf{R}^{C_e\times H\times W}$ 表示分割后的特征图，每个特征图有 $C_e = e \cdot C$ 个通道，$e = 0.25$ 是控制隐藏通道维度的拓展比例。其中，一分支 $\boldsymbol{X}_2$ 保留原始特征信息并为梯度流创建直接路径；另一分支 $\boldsymbol{X}_1$ 通过 $Conv_{1\times1}$ 卷积层和GELU（Gaussian error linear unit）激活函数处理后进行一系列转换，可表示为

$$F\left(\boldsymbol{X}_1\right) = \boldsymbol{X}_1 + \sum_{i,j\in\{1,K\}} DW_{i\times j}Conv\left(\boldsymbol{X}_1\right) + F_{\text{SFEM}}\left(\boldsymbol{X}_1\right) \tag{6}$$

式中，$DW_{i\times j}Conv(\boldsymbol{X}_1)$表示形状为$i,j\in\{1,K\}$的深度可分离卷积操作，核大小$K=31$，$\boldsymbol{F}_{SFEM}(\boldsymbol{X}_1)$表示的级联注意力机制可以表示为复合函数的映射，具体为

$$\boldsymbol{F}_{SFEM}=\boldsymbol{F}_{FGM}\circ\boldsymbol{F}_{SCA}\circ\boldsymbol{F}_{FCA} \tag{7}$$

$$\boldsymbol{F}_{FCA}=\left|f^{-1}\left(f(\boldsymbol{X}_1)\cdot\sigma\left(\phi_{pool}(\boldsymbol{X}_1)\right)\right)\right| \tag{8}$$

$$\boldsymbol{F}_{SCA}=\boldsymbol{F}_{FCA}\cdot\sigma\left(\phi_{pool}(\boldsymbol{F}_{FCA})\right) \tag{9}$$

$$\boldsymbol{F}_{FGM}=\left|f^{-1}\left(f\left(DWConv_{1\times1}(\boldsymbol{F}_{SCA})\right)\times DWConv_{1\times1}(\boldsymbol{F}_{SCA})\right)\right| \tag{10}$$

式中，$\sigma\left(\phi_{pool}(\boldsymbol{X}_1)\right)$通过AdapativeAvgpool池化和$Conv_{1\times1}$卷积生成的通道注意力权重，$\sigma$表示激活函数，$f$和$f^{-1}$分别表示二维傅里叶变换和逆变换。在通过频域通道注意力得到$\boldsymbol{F}_{FCA}$后，特征图传入空间通道注意力得到$\boldsymbol{F}_{SCA}$，最后经过频率指导调制器得到F_{FGM}。级联注意力的设计使网络能够从频域和空间域两个维度重新校准特征通道，提升对小目标特征的表示能力。最终$\boldsymbol{F}_{FSA}$可表示为

$$\boldsymbol{F}_{FSA}(\boldsymbol{X})=Conv_{1\times1}\left(Conv_{1\times1}\left(RELU\left(\boldsymbol{F}(\boldsymbol{X}_1)\right)\right)+\boldsymbol{X}_2\right) \tag{11}$$

FSA模块结合空间域和频率域的联合特征表示，多形态感受野的设计以及梯度流分级特征增强的策略，都增强了模型在复杂场景中定位和识别小目标的能力，且通过CSP结构和深度可分离卷积的设计实现精度与计算资源之间的平衡。

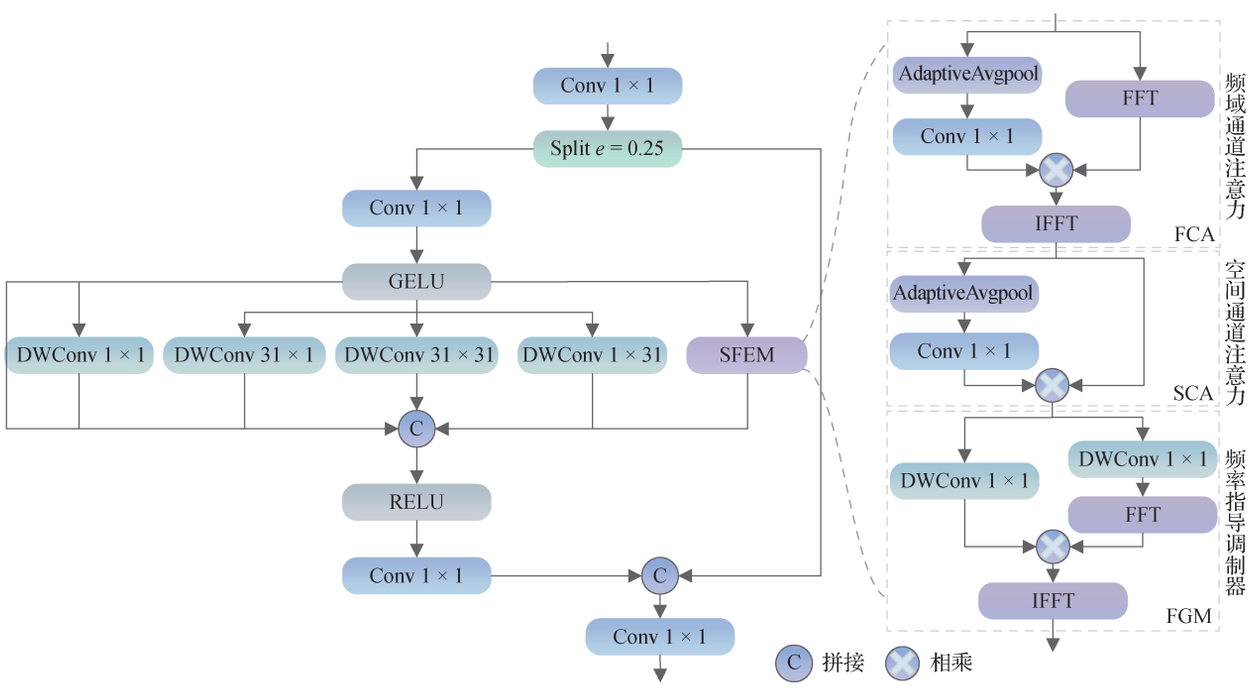


图4 FSA结构图

Fig. 4 FSA structure diagram

1.4 轻量级特征增强模块LFFE

传统的C3k2模块由于固定的感受野限制了其适应航空图像中常见的尺度变化的能力，且其特征提取中的计算冗余导致资源利用效率低下。为此设计轻量级特征增强模块LFFE替换全网络结构中的C3k2模块，其网络结构如图5所示。

LFFE模块的核心思想是通过多分支结构和可重参数化卷积增强特征表示能力，同时引入通道分割机制促进不同特征域之间的信息交流，且参考Ghost-Net的思想，以此降低计算量和参数量。首先，输入特征图$\boldsymbol{X}\in\mathbf{R}^{C_1\times H\times W}$通过$Conv_{1\times1}$卷积层，生成后经过$Split$分为两个分支的中间特征图，具体为

$$[\boldsymbol{X}_1,\boldsymbol{X}_2]=Split\left(Conv_{1\times1}(\boldsymbol{X})\right) \tag{12}$$

式中，$\boldsymbol{X}_1,\boldsymbol{X}_2\in\mathbf{R}^{C_e\times H\times W}$表示分割后的特征图，每个特征图有$C_e=e\cdot C_1$个通道，$e=0.5$是控制隐藏通道维度的拓展比例。第1分支$\boldsymbol{X}_1$保留原始特征信息并为梯度流创建直接路径，第2分支$\boldsymbol{X}_2$通过$RepConv$模块进行一系列转换，可表示为

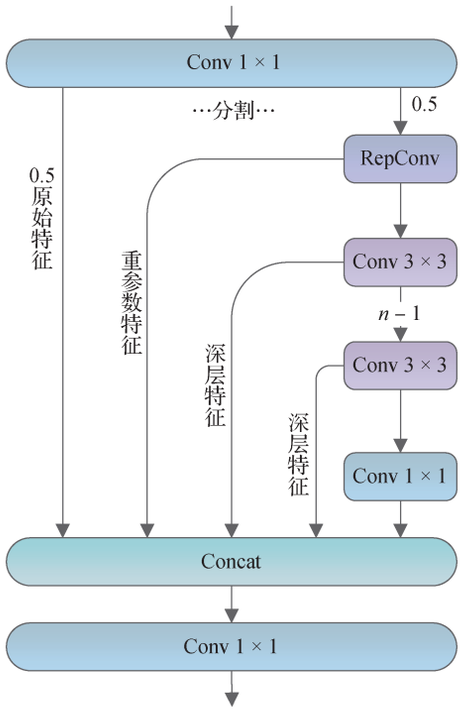


图 5　LFFE 结构图

Fig. 5　LFFE structure diagram

$$RepConv(\boldsymbol{X}_2)=\alpha\cdot Conv_{3\times3}(\boldsymbol{X}_2)+\beta\cdot Conv_{1\times1}(\boldsymbol{X}_2)+\gamma\cdot\boldsymbol{X}_2 \tag{13}$$

式中，α、β和γ是学习得到的权重参数。可重参数化卷积结合了不同核大小的卷积操作，能够同时捕获局部细节和更广泛的上下文信息，同时为了进一步提高计算效率，引入尺度因子s控制中间特征的通道数，具体为

$$C_{\text{mid}}=s\cdot C_e \tag{14}$$

式中，C_{mid}是*RepConv*和后续卷积层的输出通道数。在*RepConv*操作之后，依次应用一系列$n-1$个标准$Conv_{3\times3}$卷积以生成额外的特征图，并对最后一个特征图应用具有$Conv_{1\times1}$卷积的并行分支以生成互补特征。

$$\boldsymbol{X}_{i+1}=Conv_{3\times3}(\boldsymbol{X}_i),i\in\{1,2,\cdots,n-1\} \tag{15}$$

最后，将所有生成的特征图沿通道维度连接，并通过$Conv_{1\times1}$卷积处理，以产生输出特征图。具体为

$$\boldsymbol{F}(\boldsymbol{X})=Conv_{1\times1}\left(\boldsymbol{X}_1\oplus RepConv(\boldsymbol{X}_2)\oplus\bigcup_{i=1}^{n}Conv_{3\times3}(\boldsymbol{X}_i)\oplus Conv1\times1(\boldsymbol{X}_n)\right) \tag{16}$$

LFFE 可以根据尺度因子s和通道比例因子e的不同值灵活调整特征通道数，实现计算资源的高效分配，以此适应不同大小的模型；而从浅层到深层的渐进式特征提取的设计，使网络能够提取不同层次的特征，增强了小目标的检测精度。

2　实验与分析

在本节中，对所提出的模型 MSFA-YOLO 在两个代表性无人机图像数据集的小目标检测性能进行评估：VisDrone2019-DET 数据集、UAVDT 数据集，并将所提出方法与现有的目标检测算法进行比较。

2.1　数据集

VisDrone2019-DET 数据集是一个用于无人机视觉研究和算法评估的大规模数据集，其由天津大学 AISKYEYE 团队（Du 等，2019）收集。该数据集涵盖城市、乡村、高速公路和建筑工地等多种场景，包含 8 629 幅静态图像，其中训练集 6 471 幅、验证集 548 幅、测试集 1 610 幅；总共定义 10 个类别，分别为：行人、人、自行车、汽车、面包车、卡车、三轮车、遮阳三轮车、公共汽车和摩托车。数据集提供了详细的标注信息，包括对象边界框、对象类别、运动状态和遮挡等。

MSFA-YOLO 将应用于无人机基准物体检测与跟踪 UAVDT 数据集（Du 等，2018）来进一步验证其鲁棒性和实用性。UAVDT 数据集由无人机从城市区域获取，包含 23 258 幅训练图像和 15 069 幅测试图像，类别包含 3 类：汽车、卡车和公共汽车。

2.2　实验过程和性能指标

为评估所提出模型的检测效果，实验的软件环境为：Ubuntu 22.04、Python 3.10.14、PyTorch 2.2.2、CDUA 11.8；实验的硬件环境为：AMD EPYC 7H12（3.3 GHz CPU、60 GB）、NVIDIA GeForce RTX 4090Ti（24 GB）。本次网络训练的关键参数如表 1 所示。

为了加速模型收敛，在训练过程的最后 10 个 epoch 关闭了马赛克数据增强功能。为评估网络的有效性，选择精度（precision，P）、召回率（recall，R）、平均精度均值（mean average precision，mAP_{50}/$mAP_{50\text{-}95}$）、FPS（frames per second）、参数量（parameters/）、模型大小（model size）以及浮点运算次数（floating point operations，FLOPs）作为评价指标。

2.3　消融实验

为了逐步验证 MSFA-YOLO 中每个组件的效果，采用在基线模型中逐步应用 MEF、FSA 和 LFFE 模

表1 模型训练关键参数

Table 1 Key training parameters

参数	设置
训练时长	VisDrone2019-DET: 300 epochs UAVDT: 100 epochs
动量	0.937
初始学习率	0.01
权重衰减	0.000 5
批量大小	32 samples
输入图像大小	640×640像素
优化器	SGD
数据增强	Mosaic

块。表2和表3分别展示了在VisDrone2019-DET数据集的验证集和测试集上各模块对评估指标的影响。

1）MEF。加入MEF后，验证集和测试集上的mAP_{50}评估指标均有明显提升，分别从33.7%升至35.4%以及从27.2%提升到28.1%，这也证实了MEF使模型更容易区分小目标与背景。这是由于MEF提取多尺度边缘特征并与原始特征高效融合，该网络对复杂的背景展现出良好的抑制效果。

2）FSA。在单独加入FSA后可以改善所有评估指标，这表明该模块对于小目标的频域特征增强具有明显效果。且在主干网络加入MEF后引入FSA模块，所有评估指标得到显著的提升，在测试集中尤其是Precision和mAP_{50}分别从39.6%升至43.2%以及从28.1%升至30.8%，分别提升了3.6%和2.7%。经过主干结构中提取多尺度边缘特征并进行融合后，在颈部结构中对小目标特征进行频域特征增强，使得两个模块相互补充，得到更好的效果。

表2 VisDrone2019-DET验证集MSFA-YOLO消融实验结果

Table 2 MSFA-YOLO ablation results on VisDrone2019-DET validation set

MEF	FSA	LFFE	Precision/%	Recall/%	mAP_{50}/%	mAP_{50-95}/%	FPS/(帧/s)	参数量/M	模型大小/MB	FLOPs/G
×	×	×	44.8	34.1	33.7	19.6	142	2.58	5.2	6.3
√	×	×	46.7	35.3	35.4	21.2	135	2.60	5.3	7.7
×	√	×	45.8	35.1	35.1	20.8	131	2.94	6.0	7.6
×	×	√	45.5	34.6	34.9	20.1	**172**	**2.24**	**4.6**	**6.2**
√	×	√	48.1	35.8	37.2	22.3	147	2.72	5.6	9.2
×	√	√	48.4	36.1	36.9	22.1	136	3.06	6.2	9.1
√	√	×	48.3	36.5	37.6	22.5	129	3.42	6.9	10.5
√	√	√	**48.9**	**37.6**	**38.1**	**22.7**	133	3.08	6.1	10.4

注：加粗字体表示各列最优结果。“√”表示使用该模块，“×”表示未使用该模块。

3）LFFE。该模块的设计初衷是减少网络计算过程的冗余，实现计算资源的高效分配，在验证集上单独加入LFFE模块后，在评估指标都有所提升的情况下，参数量与模型大小分别减少了0.34 MB和0.6 MB，其中FPS提升效果明显，从142提升到172。在与前两个模块都同时加入基础模型时，得到最终模型MSFA-YOLO，其在参数量与模型大小都增加较少的同时，实现了Precision、Recall、mAP_{50}和mAP_{50-95} 4个评估指标全面提升，而且各模块之间较为稳定，相互没有冲突。在测试集上，与基线相比，最终模型的4个评估指标分别实现5%、3.8%、4.2%和2.6%的上涨。虽然由于模型的改进导致参数量、模型大小和计算量的小幅增加，但是对于FPS指标从142下降到133，只有少量的下降，说明在精度上涨较多的同时依旧保持了模型的实时性。

为进一步展示各模块的有效性，采用特征图进行对比，图6所示为依次加入各模块后特征图的对比，颜色越亮表示模型对该区域的关注度越高。

MSFA-YOLO模型在测试集上的检测对比效果如图7所示。实验使用热力图的形式展示，选择多

种检测场景作为检测样本，包括复杂背景、遮挡、暗光、密集、俯视拍摄角度等检测场景，从图 7 中使用蓝色框标出的部分可以看出，对于小目标的检测效果更佳。与改进前的模型相比，所提出的模型更好地完成了检测任务，在各种场景中能够准确识别并确定目标位置，并且相较原始模型有明显改善。

2.4　与其他主流方法的对比实验

将 MSFA-YOLO 与最新的单阶段目标检测算法进行对比，对比模型根据参数量相近进行选择，这类模型易于部署，并已在生产生活的各个领域得到广泛应用，包括 LWUVADet（lightweight UAV object detection network）（Min 等，2024）、Hyper-YOLO（Feng 等，2025）、Mamba-YOLO（Wang 等，2025）、LUD

表 3　VisDrone2019-DET 测试集 MSFA-YOLO 消融实验结果

Table 3　MSFA-YOLO ablation results on VisDrone2019-DET test set

MEF	FSA	LFFE	Precision/%	Recall/%	mAP_{50}/%	mAP_{50-95}/%	FPS /（帧/s）	参数量/MB	模型大小/MB	FLOPs/G
×	×	×	39.1	29.5	27.2	15.1	142	2.58	5.2	6.3
√	×	×	39.6	30.9	28.1	15.8	135	2.60	5.3	7.7
×	√	×	39.4	30.5	27.8	15.6	131	2.94	6.0	7.6
×	×	√	39.7	30.4	27.7	15.4	**172**	**2.24**	**4.6**	**6.2**
√	×	√	41.2	32.3	30.1	17.1	147	2.72	5.6	9.2
×	√	√	42.3	32.1	29.9	16.8	136	3.06	6.2	9.1
√	√	×	43.2	32.8	30.8	17.5	129	3.42	6.9	10.5
√	√	√	**44.1**	**33.3**	**31.4**	**17.7**	133	3.08	6.1	10.4

注：加粗字体表示各列最优结果。“√”表示使用该模块，“×”表示未使用该模块。

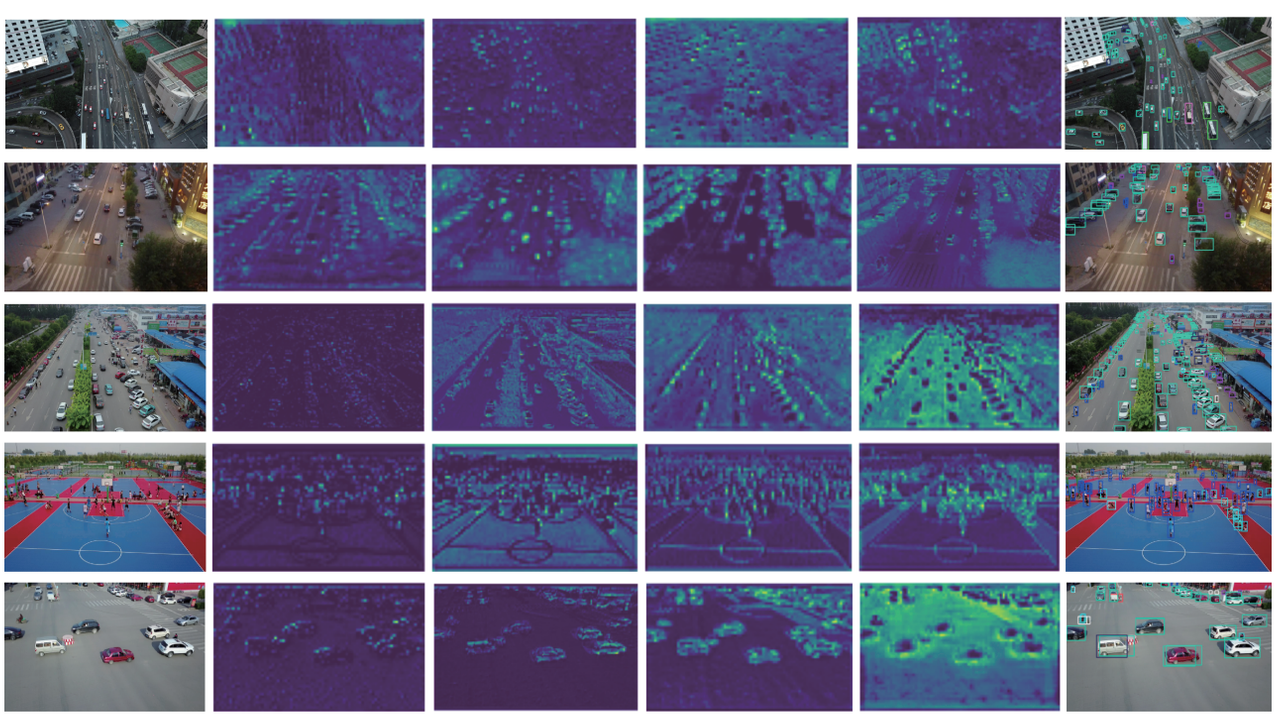


图 6　添加模块后的特征图

Fig. 6　Feature maps with added module



图7　改进前后的热力图对比

Fig. 7　Comparative heatmaps of original and improved methods

(lightweight UAV small object detection)-YOLO(Fan等,2025)以及YOLO系列最新模型等,性能评估指标使用所有类别的mAP_{50}、每个类别的平均精度以及模型参数量Parameters,对比实验结果如表4所示。表4列出了MSFA-YOLO以及其他9种对比算法的实验结果。在所有类别(All)的mAP_{50}以及除了公共汽车(bus)的单类别精度中,MSFA-YOLO的指标都取得了最佳效果,整体的mAP_{50}达到了38.1%,即使在数据量较少的自行车(bicycle)和遮阳三轮车(tricycle)中,也是分别取得了2.4%～4.9%和0.5%～4%的提升,在目标数量最多的汽车(car)类别中取得了1.6%～3.9%的提升,在参数量上分别比LWUAVDet少6.62 M和比Mamba-YOLO少2.58 M,但是其精度分别高于两者0.7%和1.7%。总之,所提出的MSFA-YOLO在解决无人机航拍图像的小目标检测任务中表现出高精度、强泛化能力和高鲁棒性。

2.5　其他数据集上的实验

为了进一步验证MSFA-YOLO的鲁棒性和实用性,将其应用于UAVDT数据集进行泛化性实验,表5展示了经典算法模型以及本研究设计的模型在该数据集上的检测效果比较,评估指标选择Precision、Recall、mAP_{50}、$mAP_{50\text{-}95}$和Parameters。由表5可知,MSFA-YOLO在UAVDT数据集上,比其他经典算法模型在评估指标上表现都更佳,与基线相比各评估指标分别有8.5%、0.6%、2.1%和1.5%的提升,由此证明了所提出的MSFA-YOLO在不同类型的数据集下仍具有较高的目标检测精度,表明该模型具有较强的泛化能力。图8展示了所提模型在UAVDT数据集上的检测效果。

表 4　VisDrone2019-DET 验证集上对比实验结果

Table 4　Comparison of experimental results on VisDrone2019-DET validation set

方法	参数量/M	mAP_{50}/%										
		All	pedestrian	people	bicycle	car	van	truck	tricycle	Awning tricycle	bus	motor
YOLOv8n	3.01	32.8	34.4	27.3	8.3	76.1	38.6	27.7	21.3	10.7	47.3	36.5
YOLOv9t	1.97	32.2	33.6	27.6	7.5	75.4	40.0	27.7	19.9	12.2	43.8	34.6
YOLOv10n	2.27	32.9	34.6	27.7	8.3	75.5	37.8	28.4	20.5	11.3	48.3	36.4
YOLO11n(基线)	2.58	33.8	35.6	27.9	9.2	76.2	39.4	29.8	21.3	11.9	49.2	37.2
YOLO12n	2.56	32.7	34,3	27.5	8.1	75.8	38.1	28.7	20.1	11.0	48.5	36.7
LUD-YOLO	2.81	35.2	36.9	29.3	9.97	77.4	41.8	31.4	22.2	13.6	49.8	39.4
LWUAVDet	9.70	37.4	–	–	–	–	–	–	–	–	–	–
Hyper-YOLO	2.68	33.7	36.3	28.3	7.6	76.5	38.2	30.5	22.1	11.7	50.3	37.4
Mamba-YOLO	5.66	36.4	38.1	29.8	10.0	77.7	42.1	33.1	23.1	14.2	**54.8**	40.6
MSFA-YOLO(本文)	3.08	**38.1**	**41.3**	**31.4**	**12.4**	**79.3**	**45.0**	**36.6**	**25.0**	**14.7**	53.0	**42.2**

注：加粗字体表示各列最优结果。“–”表示没有该部分的数据可用。

表 5　UAVDT 数据集上对比实验结果

Table 5　Comparison of experimental results on UAVDT dataset

模型	Precision/%	Recall/%	mAP_{50}/%	mAP_{50-95}/%	参数量/M
Faster RCNN	29.8	24.5	23.4	11.0	41.39
SSD	27.3	20.6	21.4	9.3	50.40
YOLOv8n	36.4	29.7	29.5	16.8	3.01
YOLOv9t	35.7	29.4	28.2	16.2	**1.97**
YOLOv10n	35.9	30.4	30.1	17.2	2.27
YOLO11n(基线)	36.6	30.6	30.2	17.8	2.58
YOLO12n	36.8	30.9	30.6	17.5	2.56
Hyper-YOLO	39.1	30.8	30.9	18.1	2.68
Mamba-YOLO	42.5	31.1	31.7	18.6	5.66
MSFA-YOLO(本文)	**45.1**	**31.2**	**32.3**	**19.3**	3.08

注：加粗字体表示各列最优结果。

3　结　论

本文提出了一种面向无人机航拍小目标检测的多尺度频域自适应网络 MSFA-YOLO。该方法包含3个核心模块：MEF 模块通过基于 Sobel 的梯度提取保留细粒度边缘特征，有效减轻传统下采样操作固有的信息损失；FSA 模块结合空频域联合表示和多形态感受野增强小目标检测能力；LFFE 模块采用多分支可重参数化卷积在保持精度的同时降低计算复杂度。

在 VisDrone2019-DET 和 UAVDT 数据集上的实验表明，MSFA-YOLO 性能优于现有先进方法。消融实验验证了各模块的有效性和互补性，证明了该方法在无人机小目标检测任务中的优越性能和泛化能力。未来工作将重点提升算法在低光照和恶劣天气等极端条件下的鲁棒性。

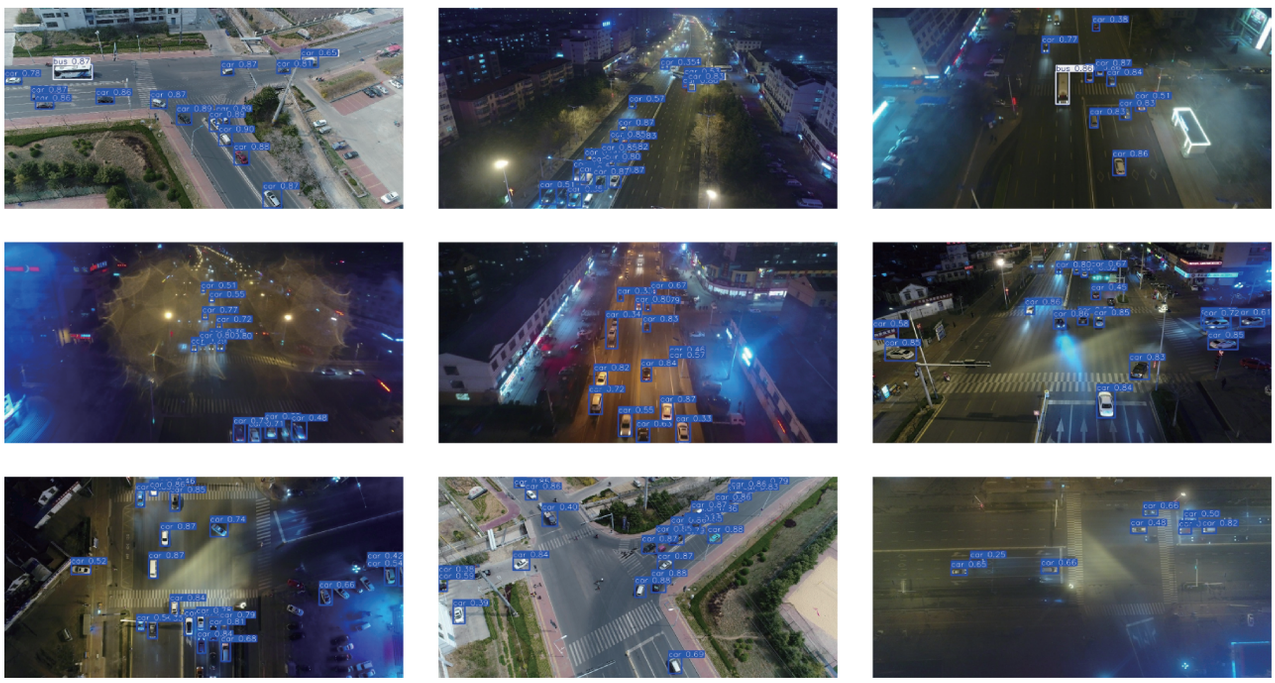


图8　MSFA-YOLO在UAVDT数据集上的检测效果

Fig. 8　Detection performance of MSFA-YOLO on UAVDT dataset

参考文献(References)

Chen J F, Hu Z H, Wu W, Zhao Y C and Huang B. 2025. LKPF-YOLO: a small target ship detection method for marine wide-area remote sensing images. IEEE Transactions on Aerospace and Electronic Systems, 61(2): 2769-2783 [DOI: 10.1109/TAES.2024.3476459]

Cui Y N, Ren W Q and Knoll A. 2024. Omni-kernel network for image restoration//Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI: 1426-1434 [DOI: 10.1609/aaai.v38i2.27907]

Du D W, Qi Y K, Yu H Y, Yang Y F, Duan K W, Li G R, et al. 2018. The unmanned aerial vehicle benchmark: object detection and tracking//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 375-391 [DOI: 10.1007/978-3-030-01249-6_23]

Du D W, Zhu P F, Wen L Y, Bian X, Lin H B, Hu Q H, et al. 2019. VisDrone-DET2019: the vision meets drone object detection in image challenge results//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision Workshop. Seoul, Korea (South): IEEE: 213-226 [DOI: 10.1109/ICCVW.2019.00030]

Fan Q S, Li Y T, Deveci M, Zhong K Y and Kadry S. 2025. LUD-YOLO: a novel lightweight object detection network for unmanned aerial vehicle. Information Sciences, 686: #121366 [DOI: 10.1016/j.ins.2024.121366]

Feng Y F, Huang J G, Du S Y, Ying S H, Yong J H, Li Y P, et al. 2025. Hyper-YOLO: when visual object detection meets hypergraph computation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 47(4): 2388-2401 [DOI: 10.1109/TPAMI.2024.3524377]

Hao C Y, Jin Y, He G Q, Zhang H, Zhu X Q and Song W R. 2025. Adaptive slicing-assisted enhanced method for small object detection. Journal of Image and Graphics, 30(7): 2389-2407 (郝川艳, 金怡, 何桂琴, 张昊, 祝翔祺, 宋婉茹. 2025. 自适应切片辅助增强的小物体目标检测. 中国图象图形学报, 30(7): 2389-2407) [DOI: 10.11834/jig.240450]

Hui Y M, Wang J and Li B. 2024. STF-YOLO: a small target detection algorithm for UAV remote sensing images based on improved Swin-Transformer and class weighted classification decoupling head. Measurement, 224: #113936 [DOI: 10.1016/j.measurement.2023.113936]

Jin T and Hu P Y. 2025. HK-DETR: improved knife-holding dangerous behavior detection algorithm based on RT-DETR. Journal of Image and Graphics, 30(4): 1027-1040 (金涛, 胡配雨. 2025. 改进实时目标检测Transformer的持刀危险行为检测算法. 中国图象图形学报, 30(4): 1027-1040) [DOI: 10.11834/jig.240295]

Kou R K, Wang C P, Fu Q, Yu Y and Zhang D D. 2022. Infrared small target detection based on the improved density peak global search and human visual local contrast mechanism. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 15: 6144-6157 [DOI: 10.1109/JSTARS.2022.3193884]

Min X L, Zhou W, Hu R, Wu Y Y, Pang Y R and Yi J. 2024. LWUAVDet: a lightweight UAV object detection network on edge devices. IEEE Internet of Things Journal, 11(13): 24013-24023

［DOI：10.1109/JIOT.2024.3388045］

Peng G L，Yang Z J，Wang S B and Zhou Y. 2023. AMFLW-YOLO：a lightweight network for remote sensing image detection based on attention mechanism and multiscale feature fusion. IEEE Transactions on Geoscience and Remote Sensing，61：#4600916［DOI：10.1109/TGRS.2023.3327285］

Shi C J，Zheng X，Zhao Z B，Zhang K，Su Z B and Lu Q C. 2024. LSKF-YOLO：large selective kernel feature fusion network for power tower detection in high-resolution satellite remote sensing images. IEEE Transactions on Geoscience and Remote Sensing，62：#5620116［DOI：10.1109/TGRS.2024.3389056］

Wang Z Y，Li C，Xu H Y，Zhu X Z and Li H B. 2025. Mamba YOLO：a simple baseline for object detection with state space model//Proceedings of the 39th AAAI Conference on Artificial Intelligence. Philadelphia，USA：AAAI：#912［DOI：10.1609/aaai. v39i8. 32885］

Xiao Z J，Li S B，Qu H C and Li F K. 2025. Remote sensing image detection guided by context information and multi-scale feature sequences. Journal of Image and Graphics，30(7)：2570-2583（肖振久，李士博，曲海成，李富坤. 2025. 上下文信息和多尺度特征序列引导的遥感图像检测. 中国图象图形学报，2025，30(7)：2570-2583）［DOI：10.11834/jig.240361］

Zhang Y，Ye M，Zhu G Y，Liu Y，Guo P Y and Yan J H. 2024. FFCA-YOLO for small object detection in remote sensing images. IEEE Transactions on Geoscience and Remote Sensing，62：#5611215［DOI：10.1109/TGRS.2024.3363057］

Zhu H，Ni H P，Liu S M，Xu G X and Deng L Z. 2020. TNLRS：target-aware non-local low-rank modeling with saliency filtering regularization for infrared small target detection. IEEE Transactions on Image Processing，29：9546-9558［DOI：10.1109/TIP. 2020. 3028457］

作者简介

郭宋浩，男，硕士研究生，主要研究方向为计算机视觉与无人机应用。E-mail：1902128057@qq.com

卢晓光，通信作者，男，副教授，主要研究方向为雷达信号处理、图像处理与识别和监视数据处理。

E-mail：xglu@cauc.edu.cn

王阳，男，高级工程师，主要研究方向为无人机应用与深度学习。E-mail：jake_wang@163.com