文章编号　1004-924X(2026)16-2536-23

边缘引导选择与特征聚合的巡飞器航拍动态目标实例分割

官　冕，张印辉*，何自芬，陈光晨
(昆明理工大学 机电工程学院，云南 昆明 650500)

摘要：针对动态地物目标与背景特征类间相似以及弱边缘感知不足导致分割精度低的问题，构建边缘先验显式选择、空间-通道动态互补聚合，双域下采样高频补偿的特征建模链路，提出了边缘引导选择与特征聚合的巡飞器航拍动态目标实例分割方法。首先，设计多尺度边缘引导选择感知模块，将多尺度特征与局部均值之间的残差响应作为边缘先验并生成边缘响应图，再通过双域选择机制强化模型对动态目标弱边缘信息的感知能力。其次，构建自适应特征聚合模块利用通道混洗机制实现通道级交互以激活跨组语义信息，并聚合动态卷积与深度动态卷积自适应提取的空间语义信息与通道互补细节特征，以提高特征响应度从而解决目标类间特征相似的问题。最后，为克服普通卷积提取多尺度空间信息的局限性，设计双域下采样自适应模块融合双路径提取到的层级特征信息，实现多尺度信息的自适应聚合与特征强化。实验结果表明，本文提出的边缘引导选择与特征聚合的巡飞器航拍动态目标实例分割模型 mAP50-95 和 mAP50 分割精度分别达到 51.8% 和 84.3%，相较于基准模型提高 5.1% 和 4.6%，将模型部署在 Jetson Nano B01 移动开发平台，能够实现动态地物目标的分割。

关　键　词：实例分割；动态目标；边缘引导选择；自适应特征聚合；模型部署
中图分类号：TP394.1；TH691.9　　**文献标识码**：A
doi：10.37188/OPE.20263416.2536　　**CSTR**：32169.14.OPE.20263416.2536

Edge selection guidance and feature aggregation for dynamic object instance segmentation in aerial photography of patrol aircraft

GONG Mian, ZHANG Yinhui*, HE Zifen, CHENG Guangchen

(*Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology, Kunming* 650500, *China*)
* *Corresponding author, E-mail: zhangyinhui@kust.edu.cn*

Abstract: Aimed at the problem of low segmentation accuracy caused by the similarity between dynamic ground targets and background feature classes, as well as insufficient weak edge perception, a feature modeling link was constructed, which included edge prior explicit selection, spatial channel dynamic complementary aggregation, and dual domain downsampling high-frequency compensation. An edge guided selection and feature aggregation method for dynamic object instance segmentation in aerial photography of patrol aircraft was proposed. Firstly, a multi-scale edge guided selection perception module was designed, which took the residual response between multi-scale features and local means as the edge prior and gener-

收稿日期：2026-04-13；**修订日期**：2026-05-20.
基金项目：国家自然科学基金(No. 62561032，No. 62171206)；昆明理工大学分析测试基金(No. 2024P20231103003)

ated an edge response map. Then, a dual domain selection mechanism was applied to enhance the model's ability to perceive weak edge information of dynamic targets. Secondly, an adaptive feature aggregation module was constructed to utilize channel level interaction through a channel shuffling mechanism to activate cross-group semantic information, and to aggregate spatial semantic information extracted adaptively by dynamic convolution and deep dynamic convolution with complementary channel detail features, in order to improve feature responsiveness and solve the problem of feature similarity between target classes. Finally, to overcome the limitations of ordinary convolution in extracting multi-scale spatial information, a dual domain downsampling adaptive module was designed to fuse the hierarchical feature information extracted from dual paths, achieving adaptive aggregation and feature enhancement capabilities of multi-scale information. The experimental results show that the edge guided selection and feature aggregation proposed in this paper for the dynamic object instance segmentation models mAP50-95 and mAP50 in aerial photography achieve segmentation accuracies of 51.8% and 84.3%, respectively, which are 5.1% and 4.6% higher than the benchmark model. Deploying the models on the Jetson Nano B01 mobile development platform can achieve dynamic object segmentation.

Key words: instance segmentation; dynamic object; edge guided selection; adaptive feature aggregation; model deployment

1　引　言

巡飞器作为融合新一代电子信息技术与航空工业技术的智能空中平台，凭借其环境适应性强、机动灵活等优势，能够在复杂环境下获取大范围地表信息。在森林防护[1-2]、智慧交通[3-5]、灾害救援[6-7]、国防安全[8-9]及电力巡检[10-11]等领域发挥着关键作用。通过对航拍图像中着火点、受灾区域、潜在威胁或电力隐患等进行精细化分割与识别，为灾害应急响应、精准打击与基础设施维护提供关键的视觉感知信息。

实例分割作为视觉感知的核心任务之一，能够在像素级别实现目标的精确检测与轮廓提取，并区分同一类别中的不同实例。然而，巡飞器航拍的俯视成像视角不同于自然场景图像，目标尺度随飞行高度变化而呈现波动，空间分布呈现不均匀性，且受平台运动、光照变化及环境噪声的综合影响，图像中往往存在目标边缘模糊、细节丢失等问题导致目标与背景间的特征混淆度提高，从而影响实例分割模型的边界定位与实例区分能力。因此增强实例分割对动态地物目标的分割精度，对增强巡飞器的环境感知具有重要意义。

由于航拍图像在俯视成像、目标尺度变化及背景复杂度等方面存在差异，使得通用实例分割方法难以获得理想的结果。因此，针对航拍图像特征开展针对性的实例分割方法研究逐渐成为低空视觉领域的重要研究方向。Zamir等[12]提出的iSAID是面向航空影像实例分割的大规模基准，并指出该任务具有单幅图像实例数量多、尺度变化显著等区别于自然场景的典型难点。其实验结果也表明，直接将Mask R-CNN[13]和PANet[14]迁移到航拍场景时，其分割精度及效果不如自然图像上表现良好。类似地，Garg等在面向无人机场景的ISDNet[15]研究中也指出目标方向随机分布以及背景杂波共同构成了航拍实例分割的主要挑战。

针对航拍场景多尺度目标变换的问题，Su等人提出HQ-ISNet[16]，在Cascade Mask R-CNN框架[17]基础上引入高分辨率特征金字塔网络，通过增强多阶段掩码分支之间的信息交互从而提高图像实例分割精度。Chen等人[18]在BlendMask框架中引入尺寸平衡与类别平衡策略，以解决遥感图像中小目标比例高和类别不均衡问题。Lee等人[19]提出CenterMask，通过引入注意力机制优化特征融合阶段，使模型在保持高效率的同时具有较高分割精度。Huang[20]等人将幽灵卷积[21]与跨阶段部分网络[22]结合以增强网络的多层次感受野范围，使模型能够更有效地捕获多尺度目标特征。Zhou等人[23]提出动态核卷积生成方法，通

过卷积核融合机制与度量学习根据输入特征动态调整卷积核权重，实现网络参数自适应更新，从而增强模型对不同目标的学习能力。此外，Huang等人[24]提出通过分层偏移补偿模块在时序上对目标进行形变偏移补偿，并引入时序记忆更新模块建立帧间动态上下文关联，从而提升航拍视角下实例目标的分割精度。

近年来，一些研究将Transformer与卷积网络结合，以提升网络对全局上下文信息的建模能力。Han等人提出AerialFormer模型[25]通过CNN与Transformer的协同结构，实现了局部细节特征与全局语义信息的联合建模，从而在航拍数据集上获得更好的分割效果。Kirillov等人提出的Segment Anything Model(SAM)[26]通过提示机制实现通用图像分割，在多个视觉任务中表现出强大的泛化能力。针对航拍领域的特殊数据分布，Zhang等人提出RS-SAM[27]通过在SAM编码器中加入多尺度特征提取模块和渐进式细化结构，使模型能够解决图像中目标尺度变化大的问题。

上述方法分别从高分辨率特征融合、注意力增强以及通用提示分割等角度推动了航拍实例分割研究的发展，但部分方法依赖较大的骨干网络、多阶段掩码分支或复杂的训练与推理范式，容易受到计算资源、存储空间和实时性要求的限制。本文研究是在较轻量化模型上提升动态地物目标的分割精度。YOLO系列实例分割模型具有单阶段端到端预测、结构简洁、推理速度快和便于边缘端部署等特点。因此，本文选择YOLOv11n-seg作为基础模型，重点研究如何增强弱边缘感知、相似背景区分和多尺度特征表达能力。并针对动态目标存在背景特征相似与弱边缘感知不足导致分割精度低的问题，提出边缘引导选择与特征聚合的巡飞器航拍动态目标实例分割方法。

主要创新点如下：

(1)设计多尺度边缘引导选择感知模块，该模块的核心并非单纯引入多尺度深度卷积或边缘增强操作，而是将多尺度特征与其局部均值之间的残差响应作为显式边缘先验，生成具有尺度差异的边缘权重图，再经由双域选择机制将空间注意力加权的多尺度深度特征与通道自适应特征融合，最后通过分组多尺度深度卷积整合不同感受野的局部信息，从而强化模型对多粒度边缘结构及细节纹理的感知与选择能力。

(2)构建自适应特征聚合模块并建立动态互补特征聚合机制。不同于常见轻量注意力模块主要依赖全局池化生成静态通道权重，也不同于普通动态卷积主要对单一路径卷积核进行输入自适应调整，自适应特征聚合模块首先通过通道混洗打破分组特征之间的信息隔离，再利用动态卷积和深度动态卷积分别学习空间语义判别核与通道互补细节核，将“跨组重排-空间动态建模-通道互补建模”串联为统一的特征聚合链路，使动态权重不再只服务于卷积核自适应，而是用于缓解航拍目标与背景类间特征相似导致的判别信息不足问题。

(3)设计双域下采样自适应模块，不同于仅通过步长卷积、池化或空间重排完成尺度压缩的下采样方法，双域下采样自适应模块通过构建局部卷积与空洞卷积的并行双路架构同步捕获局部精细特征与全局上下文信息，在降低空间分辨率的同时保留近邻结构和远距离语义关系，并利用最大池化引导的高频特征增强机制实现目标特征的适应性增强。该模块提升模型对多尺度空间结构的层次化表征稳定性。

2　本文算法

本章以YOLOv11n-seg检测模型为基础架构，提出边缘引导选择与特征聚合的巡飞器航拍动态目标实例分割模型，其整体结构如图1所示。

首先替换主干部分的C3k2模块为多尺度边缘引导选择感知模块(Multi-scale Edge adaptive Selection perception Module，MESM)，通过自适应平均池化与局部卷积操作并行提取全局特征与边缘细节特征，并利用边缘选择增强策略计算各尺度特征与局部均值的残差响应进而生成具有尺度特异性的边缘权重图，再经由双域选择机制将空间注意力加权的多尺度深度特征与通道自适应特征融合，最后通过分组多尺度深度卷积整合不同感受野的局部信息，从而强化

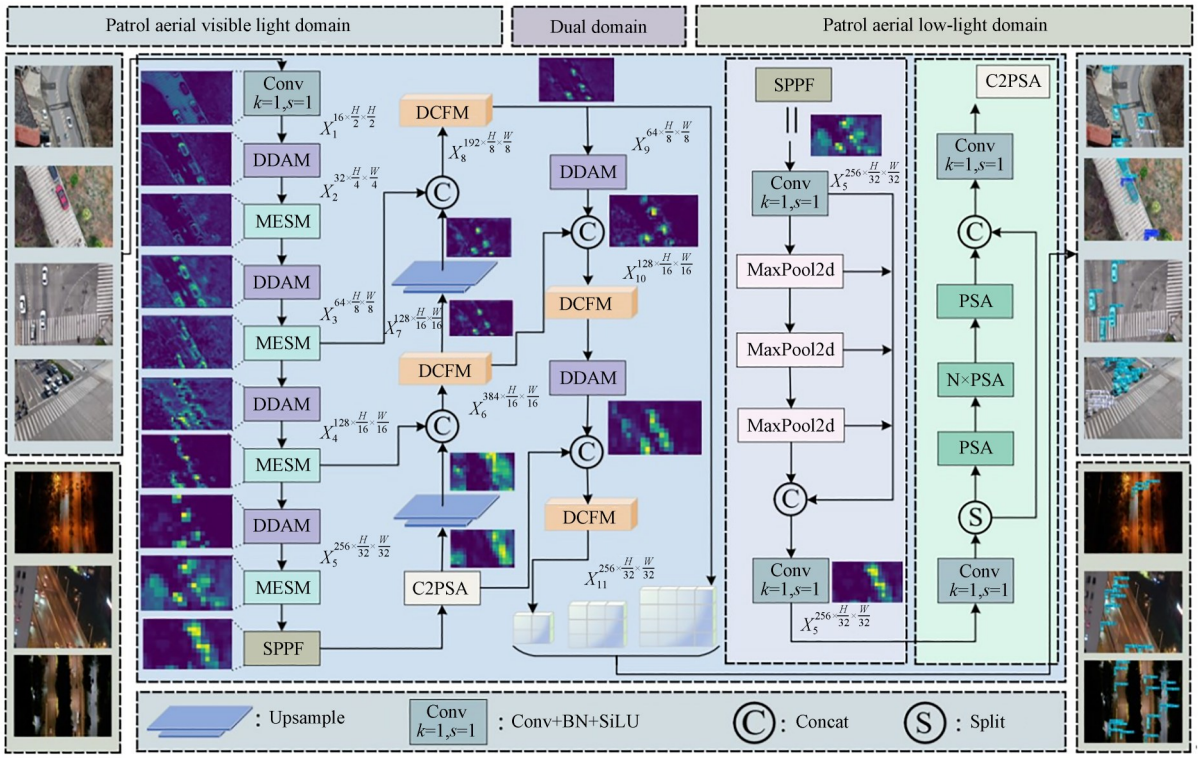

图 1　AEISM 模型结构

Fig. 1　AEISM structure

模型对多粒度弱边缘结构及细节纹理的感知与选择能力。其次在颈部特征融合阶段引入自适应特征聚合模块（Adaptive Feature Aggregation Module，AFAM），通过深度可分离卷积与通道混洗的协同机制，在通道维度上实现特征组的显式交互与重组，并结合动态权重生成机制自适应地融合多尺度空间语义信息，从而在特征提取阶段增强模型对动态目标关键特征的聚焦与判别能力。最后为进一步捕获多尺度信息，设计双域下采样自适应模块（Dual domain Downsampling Adaptive Module，DDAM），采用局部卷积与空洞卷积并行的双路结构，同步捕获局部细节与全局上下文信息，并结合特征切分与高频增强机制，以弥补普通卷积下采样丢失的语义信息。

从整体设计逻辑上看，AEISM 的三个模块分别对应巡飞器航拍动态目标实例分割中的三个误差来源。MESM 作用于主干特征提取阶段，在深层语义抽象前通过局部残差响应显式提取弱边缘先验，减少目标边界在早期卷积中的平滑丢失；AFAM 作用于颈部特征融合阶段，通过通道重排和双动态卷积聚合，将空间语义判别和通道细节互补解耦建模，缓解目标与背景外观相似导致的误激活；DDAM 作用于尺度压缩阶段，在下采样过程中同时保留局部结构、上下文语义和高频边缘信息。

2.1　多尺度边缘引导选择感知模块

针对巡飞器航拍视角灵活的成像特性导致的动态地物目标弱边缘、低对比度问题，以及由此引发的模型主干部分边缘信号丢失与语义对比度缺失现象，本文设计了多尺度边缘引导选择感知模块，首先通过自适应池化与局部卷积并行的双支路结构分别提取多尺度全局上下文特征与局部边缘细节特征，并利用边缘增强策略计算各尺度特征与其局部平均值的残差响应以生成尺度特异性边缘权重图，以增强模型对弱边缘的敏感度从而突出地物目标的边界信息，再经由双域选择机制融合空间注意力机制提取的多尺度深度特征与通道自适应对比度增强的各通道响应强度，最后通过分组多尺度深度卷积整合不同感受野的局部信息，从而强化模型对弱边缘的感知与选择能力。

与常规边缘增强模块相比，MESM的差异在于其边缘信息并非由单一卷积核或固定高通滤波器获得，而是通过多尺度特征与局部均值之间的残差响应自适应生成。当目标边缘与背景纹理差异较弱时，局部均值差分能够突出响应突变区域；当背景中存在强纹理干扰时，后续空间-通道双域选择机制又能对边缘候选进行二次筛选，减少无效纹理被误增强的风险。因此，MESM的作用不仅是增强边缘，而是在主干阶段完成边缘候选生成、边缘显著性选择和多尺度语义整合的联合建模。

针对巡飞器航拍视角灵活的成像特性导致的动态地物目标弱边缘、低对比度问题，以及由此引发的模型主干部分边缘信号丢失与语义对比度缺失现象，本文设计的多尺度边缘信息选择模块的整体结构如图2所示。输入特征 $F_{\text{input}}^{B\times C\times H\times W}$，其中 B 代表批次大小，C 代表通道数，$H\times W$ 代表输入特征图的尺寸，输入特征 $F_{\text{input}}^{B\times C\times H\times W}$ 首先经过3×3标准卷积进行特征变换从而提取像素领域的局部相关性。将输入特征通过自适应平均池化下采样并结合卷积实现通道降维得到四组特征尺寸即 $F_1^{B\times(C/4)\times 3\times 3}$，$F_2^{B\times(C/4)\times 6\times 6}$，$F_3^{B\times(C/4)\times 9\times 9}$ 和 $F_4^{B\times(C/4)\times 12\times 12}$，将得到的多粒度特征尺寸金字塔结构分别通过深度可分离卷积提取各尺度的上下文信息，其中特征 $F_1^{B\times(C/4)\times 3\times 3}$ 的3×3特征图保留了全局信息以提取整体结构信息，特征 $F_2^{B\times(C/4)\times 6\times 6}$ 的6×6特征图拥有更细致的全局信息以提取中等尺度的特征，特征 $F_3^{B\times(C/4)\times 9\times 9}$ 的9×9特征图具有较大感受野的局部信息以提取实例的细节结构特征，特征 $F_4^{B\times(C/4)\times 12\times 12}$ 的12×12特征图具有最接近原始的细节信息并增强精细的纹理边缘，将连续感受野覆盖的特征图分别通过双线性插值上采样到原始尺寸 $H\times W$。以第一组特征 $F_1^{B\times(C/4)\times H\times W}$ 为例，对特征信息进行3×3的自适应平均池化得到低频特征，即像素在每个位置的局部平均值即 $[\boldsymbol{A}_1,\boldsymbol{A}_2,\cdots,\boldsymbol{A}_n]$，其中 n 表示 $(C/4)\times H\times W$ 个特征值，输入的特征 $[\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_n]$ 减去平滑后的局部平均值 $[\boldsymbol{A}_1,\boldsymbol{A}_2,\cdots,\boldsymbol{A}_n]$ 得到高频信息，因为平均池化后的特征丢失了细节信息，所以输入特征减去平均值对应边缘等高频变化区域。

$$Edge=\sum_{i=1}^{n}(\boldsymbol{X}_i-\boldsymbol{A}_i),\tag{1}$$

其中：Edge表示高频信息，$\boldsymbol{X}_i$ 表示输入特征值，$\boldsymbol{A}_i$ 表示平均值，n 表示 $(C/4)\times H\times W$ 个特征值。将边缘特征经过一个卷积层并使用Sigmoid激活函数，将值映射到[0,1]之间自适应生成边缘权重表示边缘信息的重要程度。将边缘权重与输

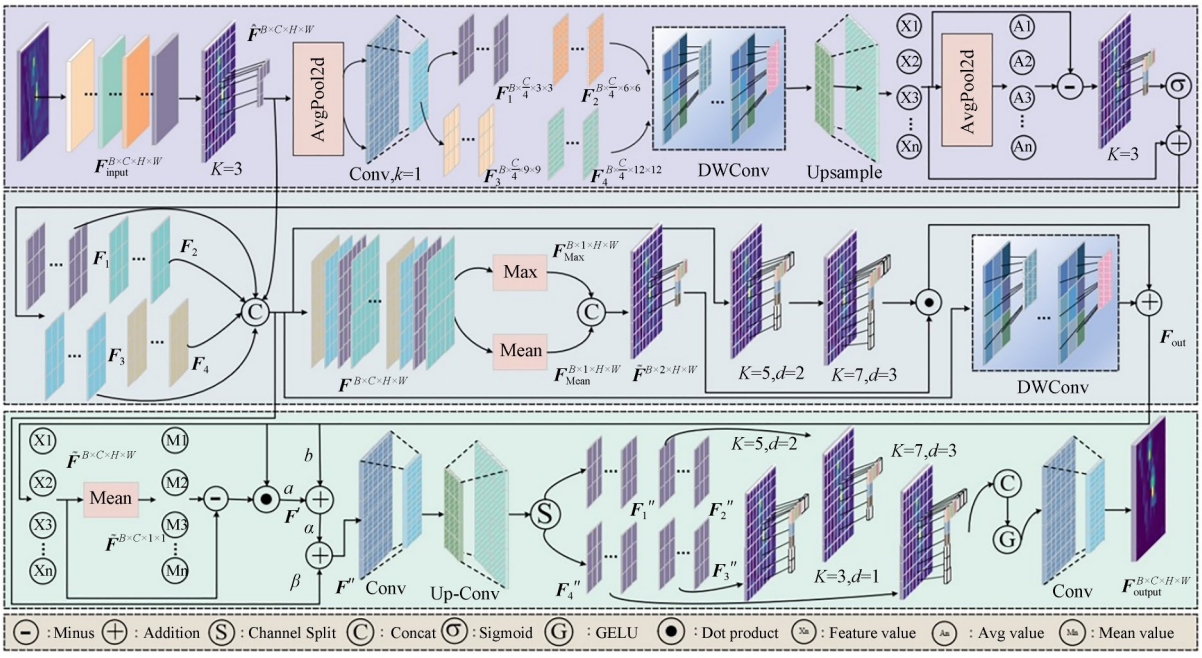

图2　多尺度边缘引导选择感知模块

Fig. 2　Multi-scale edge adaptive selection perception module

入信息相加,从而对输入特征图的像素根据边缘强度进行加权增强。

$$F_1=\sum_{i=1}^{n}(X_i+Sigmoid(Edge)),\tag{2}$$

其中:F_1表示输入特征$F_1^{B\times(C/4)\times H\times W}$经过边缘增强的输出,$Sigmoid(Edge)$表示高频信息通过Sigmoid激活函数的权重图。将经过上述操作的四组不同尺度的边缘增强特征F_1,F_2,F_3和F_4以及特征$\hat{F}^{B\times C\times H\times W}$沿通道维度拼接,形成包含全局轮廓、中等实例,局部细节、细微纹理以及空间信息的五层次复合特征张量即$F^{B\times 2C\times H\times W}$。复合特征张量$F^{B\times 2C\times H\times W}$沿通道维度取最大值即$F_{\mathrm{Max}}^{B\times 1\times H\times W}$,以突出该位置最重要的响应强化边缘信息,并沿通道维度取均值即$F_{\mathrm{Mean}}^{B\times 1\times H\times W}$,以提取该位置的整体特征信息。

$$\tilde{F}=Concat(F_{\mathrm{Max}}^{B\times 1\times H\times W},F_{\mathrm{Mean}}^{B\times 1\times H\times W}),\tag{3}$$

其中:$\tilde{F}^{B\times 2\times H\times W}$表示最大值$F_{\mathrm{Max}}^{B\times 1\times H\times W}$与均值$F_{\mathrm{Mean}}^{B\times 1\times H\times W}$沿通道维度拼接得到的既保留最显著特征又提高整体信息的特征图。通过3×3的卷积将特征$\tilde{F}^{B\times 2\times H\times W}$的双通道线性变换为单通道的空间注意力权重图。

$$\tilde{F}_{\mathrm{out}}=\left[(DC_{k=5,d=2}\times DC_{k=7,d=3})\times\tilde{F}+DC_{k=3}\right],\tag{4}$$

其中:DC表示$DWConv$深度可分离卷积,k表示卷积核,d表示扩张率。特征经过卷积核5×5和扩张率为2与卷积核7×7和扩张率为3的双尺度深度可分离卷积序列提取上下文信息,并与特征$\tilde{F}$逐元素相乘从而调整特征每个位置的重要性,再将特征通过3×3的深度可分离卷积提取到的局部细节特征相加从而得到$\tilde{F}_{\mathrm{out}}$。特征$\tilde{F}_{\mathrm{out}}$沿空间维度计算平均值即$\tilde{F}^{B\times C\times 1\times 1}$,得到每个通道的平均值$[M_1,M_2,\cdots,M_n]$,其中$n$表示通道数。将特征$\tilde{F}$的特征值$[X_1,X_2,\cdots,X_n]$减去平均值$[M_1,M_2,\cdots M_n]$,减去均值后,特征图中的每个像素值表示该位置相对于全局平均的偏离程度。引入可学习的参数a和b对特征进行非线性增强,实现自适应的特征选择与增强。将偏离程度与特征$\tilde{F}$逐元素相乘并通过参数a自适应调节这个乘积项,对特征$\tilde{F}$通过可学习参数b确保不会被丢弃。

$$F'=a\times\tilde{F}\times\left[\sum_{i=1}^{n}(X_i-M_i)\right]+b\times\tilde{F},\tag{5}$$

其中:F'表示由可学习参数自适应调节的局部对比度与输入特征相结合的综合特征。当参数$a>b$时,当$X_i>M_i$该位置权重会变大从而增强该位置的特征,当$X_i<M_i$该位置权重会变小从而抑制该位置的特征;当参数$a<b$时,即使局部对比度接近于0时,原始特征会被保留。复合特征$F^{B\times C\times H\times W}$与特征$F'$通过一组可学习参数$\alpha$和$\beta$再次自适应分配权重占比。

$$F''=\alpha\times F'+\beta\times F,\tag{6}$$

其中:F''表示由可学习参数自适应融合的综合特征与复合特征张量F'的特征增强。当参数$\alpha=0,\beta=1$时,综合特征没有提取到实例的信息,因此保留复合特征;当参数$\alpha=1,\beta=0$时,综合特征包含实例的重要特征信息,因此提高综合特征的比值。特征信息$F''\in R^{B\times 2C\times H\times W}$通过3×3的卷积将通道维度压缩到$F''\in R^{B\times C\times H\times W}$,再经过1×1的卷积将通道维度扩展为$F''\in R^{B\times 2C\times H\times W}$,使模块提取更丰富的特征表示,同时减小计算量并保留重要信息。随后将扩展后的特征$F''\in R^{B\times 2C\times H\times W}$沿通道维度均匀分成四组特征即每组通道的数量相等得到$F_1''\in R^{B\times[1\sim(C/2)]\times H\times W}$,$F_2''\in R^{B\times[(C/2)\sim C]\times H\times W}$,$F_3''\in R^{B\times[C\sim(3C/2)]\times H\times W}$和$F_4''\in R^{B\times[(3C/2)\sim 2C]\times H\times W}$。

$$F_i''=Split\left(F''\in R^{B\times\frac{C}{2}\times H\times W}\right),i=1,2,3,4,\tag{7}$$

其中:对F_1''保持不变以保留原始信息,对F_2''采用卷积核3×3和扩张率为1的深度可分离卷积提取精细局部信息,对F_3''采用卷积核5×5和扩张率为2的深度可分离卷积提取实例中等结构特征,对F_4''采用卷积核7×7和扩张率为3的深度可分离卷积提取大范围的上下文信息,随后沿通道维度拼接得到层次化的多尺度特征信息,并经过GELU激活函数引入非线性变换。

$$F_{\mathrm{output}}=F_1''+\lambda_{3,1}(F_2'')+\lambda_{5,2}(F_3'')+\lambda_{7,3}(F_4''),\tag{8}$$

其中:F_{output}表示输出特征,$\lambda_{3,1}$表示3×3的卷积核大小和扩张率为1的深度可分离卷积,$\lambda_{5,2}$表示5×5的卷积核大小和扩张率为2的深度可分离卷积,$\lambda_{7,3}$表示7×7的卷积核大小和扩张率为3的深度可分离卷积。最后通过1×1卷积将通道数压缩回初始维度,从而生成既包含像素级细节又涵盖整体结构的复合特征表示。

2.2　自适应特征聚合模块

针对动态环境下目标与背景特征相似导致的模型颈部特征融合阶段语义响应不足及背景误激活问题，现有轻量注意力模块通常通过全局平均池化、最大池化或全连接层生成通道/空间权重，本质上属于依赖全局统计量的静态重标定机制。尽管动态卷积方法可根据输入特征生成卷积核权重，但其通常仅作用于单一路径的空间特征提取，对通道组间交互及细节互补关系的建模能力有限。传统C3k2模块及CBAM等常见注意力机制在特征交互上多依赖固定权重或通过全局池化与全连接实现隐式通道加权，静态或基于全局统计的调整方式难以实现跨组结构重组与动态自适应融合。

因此，本文构建自适应特征聚合模块(Dynamic Complementary Feature aggregation Module，AFAM)来替换主干网络中的C3k2模块。所提出的AFAM模块首先构建深度可分离卷积与通道混洗的协同机制，通过逐通道细粒度特征提取与跨组交互增强特征的语义区分能力，并集成动态权重生成与多核互补聚合策略，通过全局平均池化捕获通道级全局统计信息，并经由轻量级的线性层与Sigmoid激活函数生成与输入内容相关的自适应权重系数。这些权重分别作用于多组并行的动态卷积与深度动态卷积核，使其能够根据输入特征动态调整参数，从而自适应地融合不同感受野下的空间语义信息与深度通道互补信息。

自适应特征聚合模块的整体结构如图3所示。输入特征 $\boldsymbol{X}_{\text{input}}\in\boldsymbol{R}^{C\times H\times W}$，其中 C 表示输入特征通道数，$H\times W$ 表示输入特征的尺寸，首先通过1×1的逐点卷积实现通道维度的全局线性重组得到 $\boldsymbol{X}_1^{C\times H\times W}$，将全局信息混合后的特征沿通道维度切分得到 $\boldsymbol{X}_1^{C/4}$ 与 $\boldsymbol{X}_1^{3C/4}$。通过分组卷积(GroupConv)处理 $\boldsymbol{X}_1^{C/4}$ 实现逐通道独立提取局部信息，为通道混洗提供局部通道特征。

$$\begin{cases}\boldsymbol{X}_1\in\boldsymbol{R}^{B\times C'\times H\times W}\rightarrow\boldsymbol{R}^{\left(\frac{B\times C'}{2}\right)\times 2\times(H\times W)}\\ \boldsymbol{X}_1\in\boldsymbol{R}^{\left(\frac{B\times C'}{2}\right)\times 2\times(H\times W)}\rightarrow\boldsymbol{R}^{2\times\left(\frac{B\times C'}{2}\right)\times(H\times W)}\\ \boldsymbol{X}_1\in\boldsymbol{R}^{2\times\left(\frac{B\times C'}{2}\right)\times(H\times W)}\rightarrow\boldsymbol{R}^{2\times\left(B\times\frac{C'}{2}\times H\times W\right)}\\ \boldsymbol{X}_1\in\boldsymbol{R}^{2\times\left(B\times\frac{C'}{2}\times H\times W\right)}\rightarrow\boldsymbol{R}^{B\times C'\times H\times W}\end{cases},\quad(9)$$

其中：表示通道混洗的步骤，其中 B 代表批次大小、C' 表示通道数 $C/4$，$H\times W$ 表示特征图尺寸。首先，计算中间维度即通道数分成两组，每组有 $C'/2$ 个通道得到 $\boldsymbol{X}_1\in\boldsymbol{R}^{[(B\times C'/2]\times 2\times(H\times W)}$，随后交换维度，使第一个维度为分组为2的维度得到 $\boldsymbol{X}_1\in\boldsymbol{R}^{2\times[(B\times C'/2]\times(H\times W)}$，恢复批次大小及维度，但保持分组得到两组形状为 $\boldsymbol{X}_1\in\boldsymbol{R}^{2\times[B\times(C/2)\times H\times W]}$ 的特征信息，最后沿通道维度拼接得到

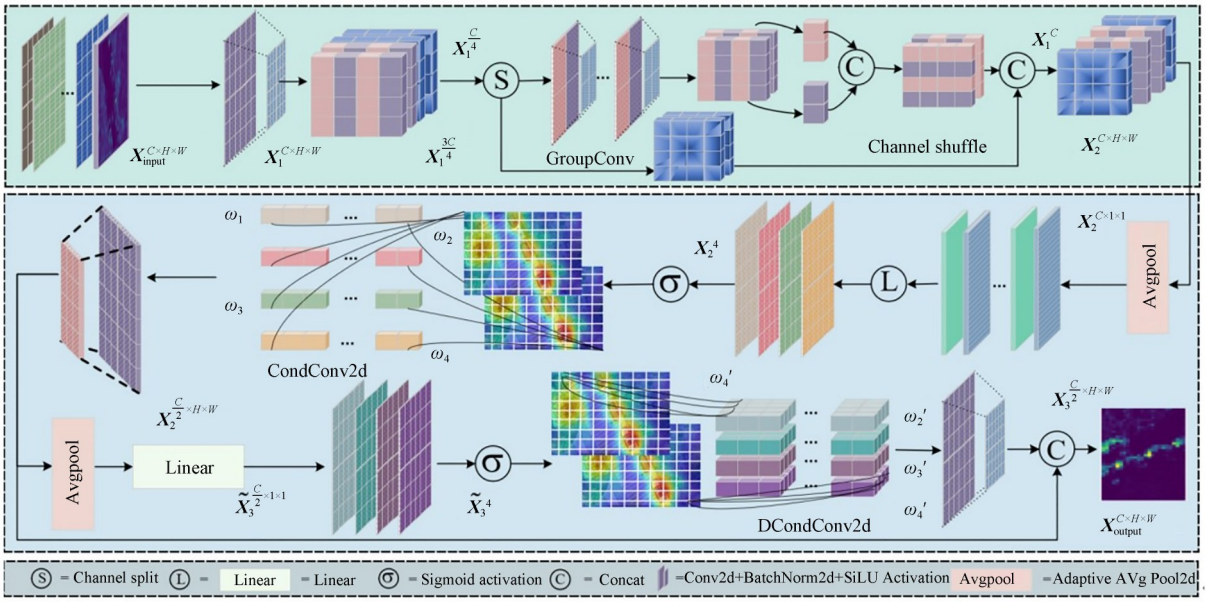

图3　自适应特征聚合模块

Fig. 3　Dynamic complementary feature aggregation module

$\boldsymbol{X}_1 \in \boldsymbol{R}^{B\times C'\times H\times W}$。利用通道混洗实现跨组语义信息交互，以及重新排列通道提升目标特征的聚集度。最后沿通道维度拼接通道混洗后的特征 $\boldsymbol{X}_1^{C/4}$ 与 $\boldsymbol{X}_1^{3C/4}$ 得到 $\boldsymbol{X}_2^{C\times H\times W}$，为自适应特征聚合提供更全局的特征信息以及更具判别性的跨组特征。

输入特征信息经过通道混洗得到的特征信息 $\boldsymbol{X}_2^{C\times H\times W}$，采用全局平均池化得到该通道在图像空间上的响应均值，并将图像空间展平去掉空间信息得到 $\boldsymbol{X}_2^{C\times 1\times 1}$，每个通道的全局空间统计信息，代表了该通道在整个特征图中的平均激活强度。随后，通过线性变换得到4个线性响应值，并通过Sigmoid激活函数使得到的线性响应值转换为4个独立的归一化权重即 ω_1，ω_2，ω_3 和 ω_4，每个权重代表四组动态卷积核相对应的权重。

$$CondConv2d = \alpha\omega_1 + \beta\omega_2 + \delta\omega_3 + \gamma\omega_4, \quad (10)$$

其中：ω_1，ω_2，ω_3 和 ω_4 代表卷积核权重，α，β，δ，γ 代表每个权重的系数。根据输入特征 $\boldsymbol{X}_2^{C\times H\times W}$ 自适应调节权重系数，并对四个权重值进行线性组合，生成一个动态的卷积核。该卷积核会随着输入特征的不同而动态变化以自适应提取输入特征的空间语义信息，使输出特征图带有样本自适应选择的知识即 $\boldsymbol{X}_2^{(C/2)\times H\times W}$。最后对输出特征图应用归一化层和 $SiLU$ 激活函数以提高模型的表征能力。

对具有空间感知的特征 $\boldsymbol{X}_2^{(C/2)\times H\times W}$，依次经过全局平均池化与空间维度的展平得到 $\boldsymbol{X}_2^{(C/2)\times 1\times 1}$，再通过线性变换将特征信息 $\boldsymbol{X}_2^{(C/2)\times 1\times 1}\rightarrow \boldsymbol{X}_2^4$，并通过Sigmoid激活函数得到归一化权重即 ω_1'，ω_2'，ω_3' 和 ω_4'，每一个权重代表对应深度动态卷积的权重。

$$DCondConv2d = \sum_{i=1}^{4}\eta_i\omega_i', \quad (11)$$

其中：η_i 表示四组深度卷积核的贡献度，ω_i' 表示四组深度动态卷积权重。深度动态卷积根据空间感知的特征 $\boldsymbol{X}_2^{(C/2)\times H\times W}$ 动态调整系数 η_i 的变化，并对四个深度权重值进行加权，生成一个深度动态卷积核。该深度动态卷积核会随着特征 $\boldsymbol{X}_2^{(C/2)\times H\times W}$ 的不同而动态变化以自适应提取特征 $\boldsymbol{X}_2^{(C/2)\times H\times W}$ 的深度通道互补信息即 $\boldsymbol{X}_3^{(C/2)\times H\times W}$，该特征图既包含动态卷积提取到的空间语义信息，又包含深度动态卷积捕捉到的互补通道信息。最后沿通道维度聚合跨空间语义特征与通道互补特征得到具有全局与局部信息特征 $\boldsymbol{X}_{\text{output}}^{C\times H\times W}$。

为解决动态地物目标与背景特征相似的问题，本文通过通道混洗机制完成像素级跨组重排、使目标特征在通道维度上物理交互，随后分别使用动态卷积与深度动态卷积提取到的跨空间语义信息与通道互补细节信息沿通道维度拼接，形成互补的特征张量，为模型在颈部结构特征融合阶段实现同时兼顾全局判别力与局部感知力的动态聚合。

2.3　双域下采样自适应模块

针对基准模型普通卷积在多尺度信息提取中存在感受野单一、尺度压缩过程中边缘和细节易丢失的问题，本文设计双域下采样自适应模块（Dual domain Downsampling Adaptive Module，DDAM），用以替代原有标准卷积下采样结构。与仅通过步长卷积、池化或空间重排完成尺度压缩的下采样方法不同，DDAM将下采样过程重新定义为局部信息、上下文信息和高频信息的协同选择过程。具体而言，模块首先构建局部卷积与空洞卷积并行的双域分支，同步捕获近邻细节和远距离上下文语义；随后通过特征切分与最大池化引导的高频增强分支，对尺度压缩后容易衰减的边缘和纹理响应进行补偿。因此，DDAM的创新点并非多分支结构本身，而是在下采样阶段显式建立“局部细节保持-上下文语义扩展-高频边缘补偿”的信息保持机制。

双域下采样自适应模块的整体结构如图4所示。以模型主干特征提取阶段的第二个卷积为例，即输入特征为 $\boldsymbol{X}_1^{16\times(H/2)\times(W/2)}$。经过步长为2，卷积核大小为3×3的卷积将输入特征 $\boldsymbol{X}_1^{16\times(H/2)\times(W/2)}$ 空间分辨率减半实现下采样的功能，同时为增加特征表示能力将特征维度提升二倍得到 $\hat{\boldsymbol{X}}_1^{32\times(H/4)\times(4)}$，在降低计算量的同时保留原始语义信息的完整，将得到的特征信息 $\hat{\boldsymbol{X}}_1$ 送入双域并行分支。首先，采用卷积核为3×3的深度可分离卷积DWConv聚焦于3×3的局部感受野并提取每个通道的精细局部信息得到 $\tilde{\boldsymbol{X}}_1^{32\times(H/4)\times(4)}$；同时并行分支采用深度可分离空洞卷积DWDConv，参数设置为空洞率为2的3×3卷积核捕获

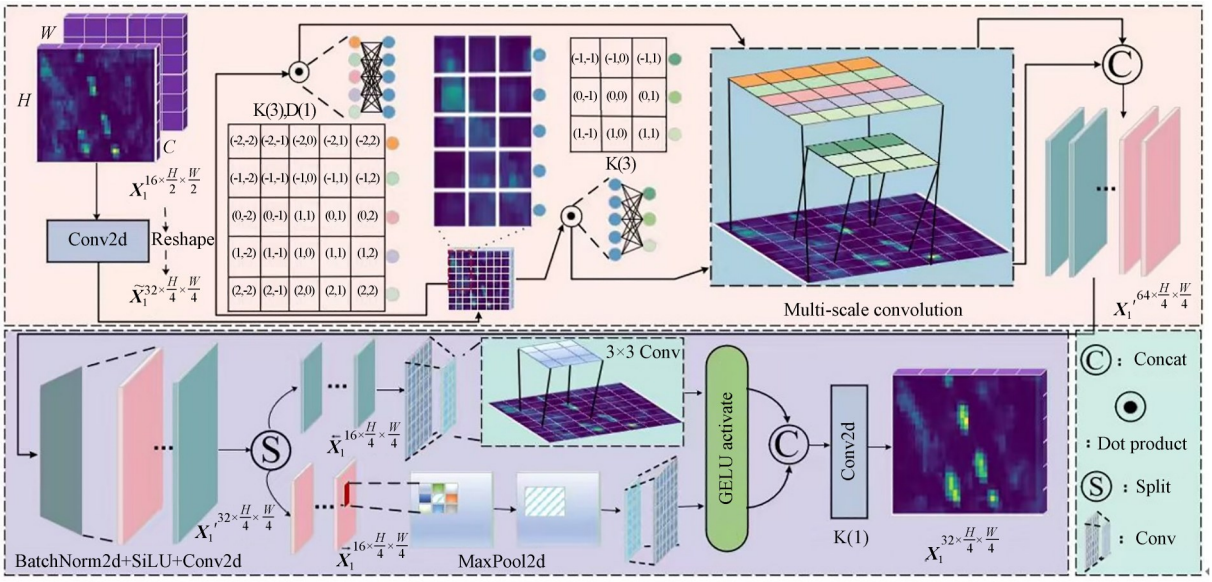

图 4　双域下采样自适应模块

Fig. 4　Dual-domain Downsampling Adaptive Module

像素间语义信息并提取上下文信息得到 $\hat{\boldsymbol{X}}_1^{32\times(H/4)\times(4)}$，如公式(12)所示：

$$\begin{cases}\tilde{\boldsymbol{X}}_1=DWConv_{3\times3}(\hat{\boldsymbol{X}}_1)\\ \hat{\boldsymbol{X}}_1=DWDConv_{3\times3,D=2}(\hat{\boldsymbol{X}}_1).\\ \boldsymbol{X}_1'=\tilde{\boldsymbol{X}}_1+\hat{\boldsymbol{X}}_1\end{cases}\tag{12}$$

深度可分离卷积提取到的局部细节信息 $\tilde{\boldsymbol{X}}_1$ 的每个位置都包含其相邻领域语义关联信息，而深度可分离空洞卷积提取到的上下文信息由于感受野增大，每个位置的特征包含 5×5 区域的上下文信息，最后沿通道维度拼接得到多尺度特征表示 $\boldsymbol{X}_1'\in\boldsymbol{R}^{64\times(H/4)\times(W/4)}$。基准模型使用的卷积一次只能在一个固定感受野内做线性映射，无法保留多尺度信息，而双域并行架构将局部精细特征与全局上下文特征沿通道维度拼接互补信息，实现多尺度空间信息的自适应聚合。将得到的特征信息 X_1' 依次经过 BatchNorm2d 对特征图实现批归一化，提高模块的稳定性；通过激活函数，引入非线性变化增强模型的表达能力，最后通过 1×1 逐点卷积将得到的特征维度映射为 $\boldsymbol{X}_1'\in\boldsymbol{R}^{32\times(H/4)\times(W/4)}$，使每个像素点具备局部细节与全局语义的双重表征。

为细化特征信息 $\boldsymbol{X}_1'$ 并促进特征交互，将特征信息沿通道维度切分为两个特征即 $\overleftarrow{\boldsymbol{X}}_1\in\boldsymbol{R}^{16\times(H/4)\times(W/4)}$ 与 $\vec{\boldsymbol{X}}_1\in\boldsymbol{R}^{16\times(H/4)\times(W/4)}$：

$$\overleftarrow{\boldsymbol{X}}_1,\vec{\boldsymbol{X}}_1=Split\left(\overleftrightarrow{\boldsymbol{X}}'_1\in\boldsymbol{R}^{32\times\frac{H}{4}\times\frac{W}{4}}\right),\tag{13}$$

其中：$\overleftrightarrow{\boldsymbol{X}}'_1$ 表示沿通道维度的中心切分得到 $\overleftarrow{\boldsymbol{X}}_1,\vec{\boldsymbol{X}}_1$。首先，将特征信息 $\overleftarrow{\boldsymbol{X}}_1$ 经过 3×3 的卷积进一步提炼局部特征，并利用 GELU 非线性激活函数增强模型的非线性表达能力，保留目标的局部特征信息；同时，将特征信息 $\vec{\boldsymbol{X}}_1$ 通过 3×3 的最大池化操作以保留局部区域的最大值，将得到的最大值经全连接层与 GELU 激活函数以增强目标的边缘等高频信息，沿通道维度拼接，实现局部信息与高频显著信息的多尺度特征融合，弥补双域聚合特征 $\boldsymbol{X}_1'$ 的细节信息，最后通过 1×1 的逐点卷积完成跨通道信息交互得到 $\boldsymbol{X}_1^{32\times(H/4)\times(W/4)}$，并生成与输入张量同形的残差项，并通过残差连接更新梯度。

3　实验结果及定量分析

3.1　数据集及实验环境配置

针对巡飞器对动态目标理解的需求，CO-Drone 数据集[28]采集自多个城市的多样化场景，包含 12 类地物目标(例如车辆、行人等)。该数据集的图像覆盖了白天、夜间、雪景等多种真实光照环境，从而有效弥补了现有巡飞器数据集在图像分辨率、目标类别丰富度、视角多样性及飞行

高度范围等方面的不足。VTUVA 数据集[29]通过 Zenmuse H20T 机载相机获取，飞行高度区间为 15～20 m，覆盖了可见光和弱光等多种复杂场景，同时兼顾不同光照和天气条件。与常规数据集相比，VTUVA 在视角动态变化及遮挡形态等方面展现出更高的复杂度，为巡飞器航拍视角下的实例分割任务提供了高质量的基准资源。

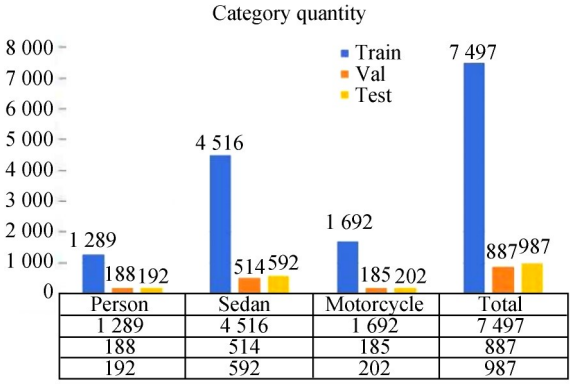

图 5　数据集构成

Fig. 5　Dataset composition

本文从 VTUVA 和 CODrone 两个公开数据集中筛选出弱光环境下的图像，共计 2 750 张，以构建面向航拍场景的目标感知数据集。针对这些图像中的行人（person）、车辆（sedan）及骑车的人（motorcycle）三类动态地物目标，采用人工标注的方式进行了精细的实例级掩码标注，为后续算法评估提供了像素级的真值标签。图像数量分别为 2 200 张、275 张和 275 张。在此基础上统计得到各子集的实例数量分布，如表 1 所示。训练集包含 7 497 个实例，其中行人的数量为 1 289、轿车的数量为 4 516、摩托车的数量为 1 692；验证集包含 887 个实例，其中行人的数量为 188、轿车的数量为 514、摩托车的数量为 185；测试集包含 987 个实例，其中行人的数量为 192、轿车的数量为 592、摩托车的数量为 202。整个数据集共计涵盖 9 371 个标注实例，充分覆盖了弱光条件下不同尺度、姿态和遮挡程度的航拍目标。

实验基于 Windows 10 操作系统，CPU 为英特尔 i7-13700KF，GPU 为 24G 版本 NVIDIA GeForce RTX 3090。深度学习框架为 PyTorch1.10.0，使用 CUDA11.3 加速模型训练，Python 版本为 3.9，以 Yolov11n-Seg 为基础模型，训练时图片输入尺寸为 640×640。在训练阶段超参数具体设置如表 1 所示。

表 1　超参数配置

Tab. 1　Hyperparameter configuration

超参数名称	超参数值
批次大小（Batch Size）	32
初始学习率（Learning Rate）	0.01
迭代次数（Epochs）	200
优化器	随机梯度下降（SGD）
Mosaic 增强	False

3.2　模型评价指标

为了客观评估本文所提出算法在实例分割任务上的性能，本文采用评估指标对模型进行定量分析，主要采用平均精度（mean Average Precision，mAP）作为核心性能指标，同时结合模型复杂度与推理效率相关指标，对算法性能进行多维度评估。主要采用 mAP50 即混淆矩阵 IOU 的阈值取 0.50 的平均精度（AP），mAP50-95 即 IOU 的阈值为 0.50 到 0.95 的平均精度，浮点运算次数（GFLOPs）、权重文件大小（MB）和推理速度（Inference）作为评价指标。

平均精度（Average Precision，AP）通常通过预测结果与真实标注之间的交并比（Intersection over Union，IoU）来衡量模型对目标实例的定位与分割准确程度。基于不同 IoU 阈值，可以进一步计算对应的精确率（Precision）与召回率（Recall），进而绘制精确率-召回率（Precision-Recall，PR）曲线，其积分面积即为平均精度，如图 6 所示。

mAP50 表示在 IoU 阈值设定为 0.50 时所计算得到的平均精度值，即当预测与真实掩码之间的交互比大于 0.50 时视为正确匹配，并据此计算 PR 曲线下面积。mAP50-95 的 IoU 阈值从 0.50 到 0.95，以 0.05 为步长均匀取值，共包含 10 个不同阈值。在每个 IoU 阈值下分别计算对应的平均精度，并最终取这些 AP 值的算术平均作为最终评价结果，如图 7 所示。

$$AP=\int_0^1 P(R)\mathrm{d}R, \tag{14}$$

$$mAP=\frac{\sum_{i=1}^{n}AP_i}{n},\tag{15}$$

其中：P为精准率，R为召回率，AP_i代表第i个类别的平均精度，n代表类别数量。

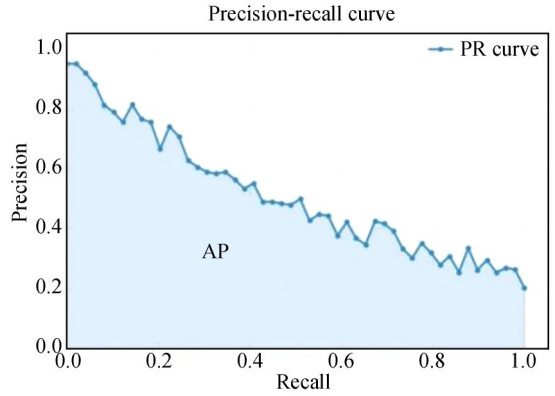

图 6　AP 曲线

Fig. 6　AP curve

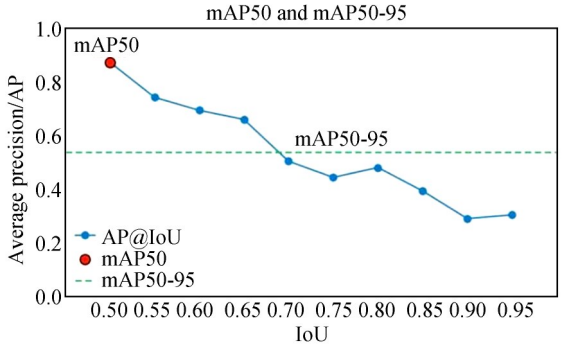

图 7　mAP50 与 mAP50-95 曲线

Fig. 7　mAP50 and mAP50-95 curves

3.3　模块消融实验

为探究多尺度边缘引导选择感知模块（MESM）、自适应特征聚合模块（AFAM）以及双域下采样自适应模块（DDAM）对基准网络性能的影响，本文设置五组消融实验分析改进模型对分割精度的有效性，以验证提出模块的优势。

根据表 2 的实验结果可知，YOLOv11n-Seg 基准分割模型的 mAP50-95 为 46.7%，mAP50 为 79.7%。在此基础上，本文分别引入多尺度边缘引导选择感知模块 MESM、双域下采样自适应模块 DDAM 和自适应特征聚合模块 AFAM，并结合图 1 中的网络结构及特征响应可视化结果，对各模块的有效性进行进一步分析。

首先，在基准模型主干网络特征提取阶段引入多尺度边缘引导选择感知模块 MESM，该模块通过多尺度边缘特征的精准感知与自适应选择机制，使模型 mAP50 提升 2.5% 至 82.2%，mAP50-95 提升 3.2% 至 49.9%，计算量 GFLOPs 增加 0.2 但模型大小由 5.73M 优化至 5.55M。结合图 1 可视化结果，MESM 被嵌入到主干网络的多级特征提取过程中，其作用并不是单纯增加网络深度，而是在不同分辨率阶段持续强化目标边缘和弱小目标候选区域。对应的特征图中，浅层阶段仍存在较多背景纹理响应，而经过 MESM 特征提取后，目标轮廓和局部结构区域的激活更加明显，表明该模块能够提升弱边缘目标和小目标的初始感知能力。因此其 mAP50 提升更明显，MESM 侧重边缘响应激活和目标检出，高 IoU 阈值下的掩码贴合度还依赖后续特征融合与掩码细化。

其次，为突破普通卷积在多尺度特征提取上的局限，采用双域下采样自适应模块 DDAM 替换基准下采样操作，通过局部与全局双路并行结构捕获多尺度上下文，mAP50 与 mAP50-95 分别提升 2.2% 与 2.2% 至 81.9% 与 48.9%，此时计算量增加 4.9 GFLOPs 至 15.1，模型尺寸增至

表 2　模块消融实验

Tab. 2　Models Ablation experiments

Models	MESM	DDAM	AFAM	mAP50/%	mAP50-95/%	GFLOPs	Size/MB	Para/M	Inference/ms
YOLOv11n				79.7	46.7	10.2	5.73	2.84	2.0
＋MESM	√			82.2	49.9	10.7	5.55	2.88	1.7
＋DDAM		√		81.9	48.9	15.1	9.09	4.56	2.8
＋AFAM			√	81.8	51.1	9.9	5.53	2.70	2.0
＋M＋D	√	√		83.0	51.1	15.6	9.46	4.61	2.4
AEISM	√	√	√	84.3	51.8	15.3	9.26	4.47	2.8

9.09 M，结合图 1 中 DDAM 所在的多级下采样路径可知，经过处理后的特征图在分辨率降低的同时，仍能保持对目标主体和边缘结构的有效响应，没有出现明显的目标区域响应消散现象，说明 DDAM 能够在尺度压缩过程中缓解普通下采样造成的小目标边界、高频纹理和局部结构信息丢失问题，然而，从复杂度指标来看，DDAM 引入后 GFLOPs 由 10.2 增加至 15.1，模型大小增至 9.09 M，复杂度增幅最为明显，原因在于采用局部卷积与空洞卷积并行的双域下采样结构，同时引入高频增强路径，以在尺度压缩过程中保留局部细节、上下文语义和边缘纹理信息。

随后，在基准模型颈部特征融合阶段引入自适应特征聚合模块 AFAM，模块通过通道混洗与动态卷积协同聚合语义与空间互补信息，使模型 mAP50 提升 2.1% 至 81.8%，mAP50-95 提升 4.4% 至 51.1%，计算量 GFLOPs 减小 0.3 但模型大小由 5.73 M 优化至 5.55 M，可能因动态卷积核随输入自适应变化，增强了航拍图像间特征提取的一致性。分析图 1 中双域交互区域和颈部融合路径可视化，不同尺度特征通过上采样和融合模块进行信息交互，使浅层空间细节与深层语义特征得到互补。经过 AFAM 融合后，特征响应由局部离散激活逐渐转向目标区域的连续响应，表明该模块有助于增强目标内部区域一致性和边界表达能力。对于航拍图像中车辆、行人等动态目标易与道路纹理、阴影区域和建筑边缘混淆的问题，AFAM 能够抑制背景误激活，强化目标区域的完整表达，从而提高预测掩码与真实标注之间的重合度。因此，相比主要提升目标检出能力的模块，AFAM 对 mAP50-95 的提升更加明显。

最后，当 MESM、DDAM 和 AFAM 同时引入后，AEISM 的 mAP50 和 mAP50-95 分别达到 84.3% 和 51.8%，相比基线分别提升 4.6% 和 5.1%。三者具有互补关系，即 MESM 主要增强弱边缘目标的感知与召回，DDAM 主要补偿下采样过程中的结构和高频信息损失，AFAM 主要提高多尺度特征融合质量和高 IoU 阈值下的掩码一致性。从复杂度看，AEISM 的 GFLOPs、模型大小和参数量分别为 15.3，9.26 MB 和 4.47 M，相比基线增加 5.1 GFLOPs，3.53 MB 和 1.63 M，单次推理时间由 2.0 ms 增至 2.8 ms。因此，AEISM 不是以最低复杂度为目标，而是在可接受计算开销内换取弱光航拍动态目标分割精度的提升。

3.4　不同分辨率下的精度与效率实验

为验证模型不同分辨率下的分割效率，统计了模型在 320，480，640，800 和 960 五种输入分辨率下的 mAP50、平均延时、mAP50-95、FPS、显存占用和平均功耗等指标。平均延时表示模型完成单张图像推理所需的平均时间，单位为 ms，数值越小表示模型响应速度越快；FPS 表示模型单位时间内能够处理的图像帧数，即推理吞吐率，数值越大表示实时性越好；显存表示模型推理过程中 GPU 的显存占用，单位为 MB，可反映模型部署时对硬件存储资源的需求；平均功耗表示推理过程中 GPU 的平均功率消耗，单位为 W，可用于评估模型在实际部署时的能耗水平。

由表 3 实验结果定量分析可知，AEISM 相比 YOLOv11n-seg 在多数输入分辨率下均取得了更高的实例分割精度，但同时带来一定推理效率损失。在 320，480，640 和 800 输入分辨率下，AEISM 的 mAP50 和 mAP50-95 均高于基线模型。其中，在本文主要采用的 640×640 输入尺度下，AEISM 的 mAP50 和 mAP50-95 分别达到 84.3% 和 51.8%，相比 YOLOv11n-seg 的 79.7% 和 46.7% 分别提升 4.6 和 5.1%，说明所提出的边缘引导选择、特征聚合和双域下采样机制能够有效增强弱光航拍动态目标的分割性能。

从效率指标来看，AEISM 的平均延时高于 YOLOv11n-seg，FPS 相应降低。在 640×640 输入尺度下，YOLOv11n-seg 的平均延时为 7.05 ms，FPS 为 141.81；AEISM 的平均延时为 14.30 ms，FPS 为 69.94。虽然 AEISM 推理速度下降，但仍能达到接近 70 FPS 的实时推理能力。与此同时，AEISM 在 640×640 下的显存占用为 3894 MB，仅比基线模型的 3 828 MB 增加 66 MB，说明模型结构改进并未导致显存占用大幅增加；其平均功耗为 153.12 W，高于基线模型的 113.93 W，表明新增模块确实带来了额外计算负载。

综上，AEISM 并非在最低 GFLOPs 或最小模型大小条件下获得最快速度，而是在可接受的

表 3　不同分辨率下的精度与效率对比

Tab. 3　Comparison of accuracy and efficiency at different resolutions

输入分辨率	模型	mAP50/%	mAP50-95/%	平均延时/ms	FPS	显存/MB	平均功耗/W
320	YOLOv11n-seg	22.1	11.6	5.36	186.55	3 804	172.30
	AEISM	31.2	16.4	12.41	80.57	3 853	171.78
480	YOLOv11n-seg	57.6	31.1	6.37	156.87	3 948	118.36
	AEISM	62.9	35.2	13.64	73.33	4 132	146.62
640	YOLOv11n-seg	79.7	46.7	7.05	141.81	3 828	113.93
	AEISM	84.3	51.8	14.30	69.94	3 894	153.12
800	YOLOv11n-seg	74.6	41.1	8.31	120.25	3 859	139.85
	AEISM	77.2	43.2	15.43	64.80	3 939	171.90
960	YOLOv11n-seg	60.4	32.2	8.81	113.45	3 961	154.88
	AEISM	56.7	30.9	16.24	61.57	3 973	185.40

计算开销范围内提高弱光航拍动态目标实例分割精度。

3.5　主干网络不同注意力的消融实验

为分析多尺度边缘引导选择感知模块 MESM 在主干部分特征提取中的有效性，将 EIEM，MLCA，MSBlock，LFE，DBB，FMB 以及 MESM 嵌入模型主干网络的同一位置以验证有效性。

由表 4 实验结果定量分析可知，在主干网络同一位置嵌入不同特征增强模块后，模型在分割精度和复杂度方面表现出明显差异。多尺度边缘引导选择感知模块 MESM 的 mAP50 达到 82.2%，在所有对比模块中最高，相较 EIEM，MLCA，MSBlock，LFE，DBB 和 FMB 分别提升 0.5%，1.3%，7.1%，0.3%，2.3% 和 1.6%，说明 MESM 在目标检出、弱边缘响应和粗粒度边界定位方面具有更强的促进作用。由于 mAP50 更侧重于预测掩码与真实掩码是否达到基本重合，因此 MESM 能够更有效地增强弱光航拍场景下动态目标的边缘显著性和目标召回能力。

表 4　主干网络不同注意力消融实验

Tab. 4　Different Attention Ablation in Backbone Networks

Attention	mAP50/%	mAP50-95/%	GFLOPs	Size/MB	Para/M	Inference/ms
EIEM	81.7	50.1	10.0	5.50	2.69	2.1
MLCA	80.9	48.8	10.2	5.74	2.84	2.1
MSBlock	75.1	43.9	10.2	5.84	2.76	2.3
LFE	81.9	49.9	10.0	5.55	2.74	1.5
DBB	79.9	47.3	10.2	6.55	2.83	1.8
FMB	80.6	50.2	10.4	5.76	2.84	1.6
MESM	82.2	49.9	10.7	5.55	2.88	1.7

从 mAP50-95 综合分割精度看，MESM 为 49.9%，与 LFE 持平，相较 MLCA、MSBlock 和 DBB 分别提高 1.1%，6.0% 和 2.6%，但相较 EIEM 和 FMB 分别低 0.2% 和 0.3%。由结果可知 MESM 在高 IoU 阈值下的精细掩码贴合度并非绝对最优，EIEM 和 FMB 对部分目标的精细轮廓拟合具有一定优势；但二者在 mAP50 上均低于 MESM，表明其对弱边缘目标的整体检出能力和目标召回能力不如 MESM。因此，MESM 的优势主要体现在主干特征提取阶段对弱边缘、小目标和低对比度目标的早期感知。

在复杂度 GFLOPs 方面，MESM 为 10.7 GFLOPs，相较于对比模块为最高，但模型尺寸 Size 为 5.55 MB，与 LFE 并列，优于 MLCA 的 5.74 MB，MSBlock 的 5.84 MB，DBB 的 6.55 MB 和 FMB 的 5.76 MB。实验结果表明，本文在

主干网络嵌入MESM能够强化模型特征提取的能力。

3.6　颈部不同注意力对比实验

为验证所提自适应特征聚合模块的优势AFAM，本文将RVB，WDBB，Star，AdditiveBlock，CaFormer，IDWB以及AFAM嵌入模型颈部特征融合阶段的同一位置以验证其有效性。

由表5实验结果定量分析可知，在模型颈部特征融合阶段嵌入不同特征增强模块后，模型在分割精度和复杂度方面表现出明显差异。自适应特征聚合模块AFAM的mAP50达到81.8%，相较基准模型YOLOv11n-seg提高2.1%，并且高于RVB，WDBB，Star，AdditiveBlock，CaFormer和IDWB，分别提升1.5%，2.5%，4.8%，4.7%，2.2%和2.3%，说明AFAM在颈部特征融合阶段能够更有效地增强动态目标的语义响应和目标区域判别能力。由于颈部网络主要承担多尺度特征融合任务，因此AFAM的精度提升表明，其能够在融合不同层级特征时强化目标区域的有效信息表达，缓解航拍图像中动态目标与复杂背景之间特征相似所造成的误检和漏检问题。从mAP50-95综合分割精度看，AFAM达到51.1%，相较基准模型YOLOv11n-seg提高4.4%，同时相较RVB，WDBB，Star，AdditiveBlock，CaFormer和IDWB分别提高2.0%，3.9%，4.6%，5.3%，4.7%和4.6%，在所有对比模块中取得最优结果。

表5　颈部不同注意力消融实验

Tab. 5　Neck different attention ablation experiment

Attention	mAP50/%	mAP50-95/%	GFLOPs	Size/MB	Para/M	Inference/ms
YOLOv11n-seg	79.7	46.7	10.2	5.73	2.84	2.0
RVB	80.3	49.1	9.9	5.47	2.68	1.3
WDBB	79.3	47.2	10.2	7.00	2.84	1.2
Star	77.0	46.5	10.2	5.67	2.79	2.3
AdditiveBlock	77.1	45.8	10.4	5.55	2.89	1.3
CaFormer	79.6	46.4	10.3	5.64	2.79	2.3
IDWB	79.5	46.5	10.1	5.58	2.75	1.6
AFAM	81.8	51.1	9.9	5.53	2.59	2.0

在模型效率方面，AFAM的参数量Para为2.59 M，不仅低于基准模型YOLOv11n的2.84 M，也低于其余对比模块。在模型复杂度GFLOPs方面，AFAM相比于基准模型下降0.3GFLOPs，优于嵌入其余注意力模型，由于引入动态卷积自适应提取目标的特征信息，虽然与基准模型的推理速度相同，但慢于嵌入RVB，WDBB，AdditiveBlock和IDWB模块的模型推理速度。实验结果表明，本文提出AFAM模块兼顾模型复杂度与检测精度，可以解决动态实例目标与背景类间特征相似的问题。

3.7　不同主干网络对比实验

为探究所提多尺度边缘引导选择感知模块与双域下采样自适应模块对模型主干部分特征高效提取的能力及分割精度的影响，将ResNet18，ResNet50，Mobilenet，Fasternet，Unireplknet，Vanillanet，Repvit，RevCol及Swin Transformer分别替换基准模型YOLOv11n-seg的主干部分，以验证MESM与DDAM模块对网络提取特征的能力。

由表6消融实验结果定量分析可知，在YOLOv11n-seg中嵌入MESM和DDAM后，模型的mAP50与mAP50-95分别达到84.3%和51.8%，相较于基准模型YOLOv11n-seg分别提高4.6%和5.1%，与其他典型主干网络相比，mAP50指标上均取得最优结果，相较ResNet18，ResNet50，MobileNet，UniRepLKNet，VanillaNet，RepViT，RevCol和Swin Transformer分别提高2.0%，0.7%，5.8%，2.9%，4.9%，3.2%，11.4%，5.3%和3.9%；在mAP50-95指标上分别提高3.0%，1.0%，9.5%，3.5%，4.2%，4.2%，9.0%，7.2%和3.6%。实验结果表明，相

表 6　不同主干网络对比实验

Tab. 6　Comparative experiment of different backbone networks

Backbone	mAP50/%	mAP50-95/%	GFLOPs	Size/MB	Para/M	Inference/ms
YOLOv11n-seg	79.7	46.7	10.2	5.73	2.84	2.0
Resnet18	82.3	48.8	39.0	26.1	13.58	2.4
Resnet50	83.6	50.8	79.5	55.3	28.79	4.6
MobileNet	78.5	42.3	9.6	4.69	2.29	1.6
FasterNet	81.4	48.3	13.1	8.26	4.15	1.9
Unireplknet	79.4	47.6	18.0	12.4	6.06	3.4
Vanillanet	81.1	47.6	99.1	57.0	23.94	6.2
Repvit	72.9	42.8	20.9	13.4	6.68	2.2
RevCol	79.0	44.6	8.8	4.94	2.35	1.4
SwinTransformer	80.4	48.2	81.5	57.7	29.97	8.6
YOLOv11-seg+M,D	84.3	51.8	15.3	9.26	4.61	2.8

较于直接替换为通用主干网络，在 YOLOv11n-seg 原有轻量化结构基础上引入 MESM 和 DDAM，更有利于针对巡飞器航拍动态目标的弱边缘、低对比度和尺度变化问题进行特征增强。

从模型复杂度看，YOLOv11n-seg+M,D 的 GFLOPs 为 15.3，相较基准模型的 10.2 增加 5.1，模型大小由 5.73 MB 增至 9.26 MB，参数量由 2.84 M 增至 4.61 M，说明 MESM 和 DDAM 在提升分割精度的同时引入了一定计算和参数开销。但与 ResNet18，ResNet50，UniRepLKNet，Vanillanet，RepViT 和 Swin Transformer 等主干网络相比，本文模型的 GFLOPs 和参数量均明显更低。与 MobileNet，Fasternet 和 RevCol 等更轻量化主干相比，本文模型的参数量和 GFLOPs 相对更高。其中 MobileNet 和 RevCol 的参数量分别为 2.29 M 和 2.35 M，低于本文模型的 4.61 M；Fasternet 的参数量为 4.15 M，也略低于本文模型。但这些轻量化主干在 mAP50 和 mAP50-95 上均低于 YOLOv11n-seg+M,D，说明过度压缩主干网络会削弱模型对弱边缘目标、多尺度结构和复杂背景语义的表达能力。相比之下，本文方法在适度增加参数量的条件下获得了更高的分割精度，体现出更优的精度-复杂度折中。

3.8　不同尺度选择实验

为验证多尺度边缘引导选择感知模块中四种尺度变换的有效性，分别选取等比缩放、奇数尺寸、斐波那契、偶数尺寸和极端跨度的尺寸替换多尺度边缘引导选择感知模块 MESM 中多粒度特征尺寸金字塔结构，以验证其四种尺度构成的金字塔结构的优势。

由表 7 实验结果定量分析可知，等比缩放策略(2,4,8,16)虽覆盖从精细到全局的倍数增长尺度，但对中间尺度特征提取不足，导致中等尺度目标的边缘信息感知不充分；奇数尺寸策略(1,3,5,7)聚焦于小尺度，虽能避免对齐偏差，但缺乏大尺度上下文引导，对远距离语义关联响应较弱；斐波那契策略(2,3,5,8)遵循自然增长规

表 7　不同尺度选择实验

Tab. 7　Selecting experiments at different scales

Select	Value	mAP50/%	mAP50-95/%	GFLOPs	Inference/ms
等比缩放	2,4,8,16	81.5	48.2	10.7	2.2
奇数尺寸	1,3,5,7	78.8	47.3	10.7	2.3
斐波那契	2,3,5,8	82.6	49.2	10.7	1.7
偶数尺寸	4,8,12,16	80.8	50.3	10.7	1.9
极端跨度	1,4,10,16	79.3	48.3	10.7	2.3
MESM	3,6,9,12	82.2	49.9	10.7	1.7

律,兼顾细节与中等尺度特征,在mAP50分割精度达到最优为82.6%,但对更大尺度上下文信息的覆盖有限;偶数尺寸策略(4,8,12,16)关注规则倍数尺度,在mAP50-95分割精度表现最佳为50.3%,但对细粒度边缘细节的提取能力不足;极端跨度策略(1,4,10,16)虽同时覆盖极细与长距离特征,但中间尺度缺失导致对常规尺度目标的边缘连续性保持较差。单一倾向的尺度配置易导致模型对不同粒度目标的边缘感知不均,而多尺度边缘信息选择模块MESM采用线性递增尺度(3,6,9,12),通过均匀覆盖从细节到中等偏大的尺度范围,并利用双域选择机制进行自适应融合,在mAP50分割精度达到82.2%,mAP50-95分割精度49.9%等精度指标上取得均衡表现,且保持最优推理效率为1.7 ms,实现了对实例目标多粒度边缘结构感知精度与计算效率的有效平衡。

3.9　不同下采样对比实验

为验证所提双域下采样自适应模块的有效性,本节将SCDown,SPDConv,PSConv,LDConv,RFAConv以及DDAM嵌入模型的同一位置以验证其有效性。

表8　不同下采样之间的实验

Tab.8　Experiments between different downsampling methods

Conv	mAP50/%	mAP50-95/%	GFLOPs	Size/MB	Para/M	Inference/ms
YOLOv11n-seg	79.7	46.7	10.2	5.73	2.84	2.0
SCDown	76.3	44.9	9.3	4.63	2.25	1.7
V7DownSampling	82.5	46.7	9.5	5.09	2.48	1.6
SPDConv	80.1	48.8	15.2	9.56	4.84	1.3
PSConv	78.6	44.6	10.1	5.52	2.71	1.7
LDConv	76.1	44.3	9.2	4.95	2.42	2.6
RFAConv	79.7	47.6	10.5	5.88	2.89	1.6
DDAM	81.9	48.9	15.1	9.09	4.56	2.8

由表8实验结果分析可知,在模型中嵌入双域下采样自适应模块时,mAP50较基准网络YOLOv11n提高2.2%达到81.9%,优于其他对比方法;在mAP50-95检测精度上,较基准网络分别提高2.2%达到48.9%,SCDown下降1.8%,V7DownSampling保持不变、PSConv下降2.1%,LDConv下降2.4%,而SPDConv提高1.8%,RFAConv提高0.9%。本文提出的双域下采样自适应模块通过构建局部卷积与空洞卷积并行的双域下采样架构,同步提取局部细节特征与全局上下文信息,而SCDown,SPDConv等方法多侧重于单一尺度的空间压缩或特征重组,缺乏多尺度特征的互补性融合。其次,在特征增强机制上,双域下采样自适应模块引入通道切分与最大池化引导的高频增强路径,显式强化边缘及纹理等高频信息,可以缓解下采样过程中的细节丢失问题,而PSConv与LDConv等方法未针对高频特征进行专门优化。因此,本文提出的双域下采样自适应模块通过双域并行架构与高频增强机制的协同作用,提升了模型对多尺度空间结构的表征能力。

3.10　不同YOLO系列实例分割模型对比实验

为了验证所提网络AEISM的有效性,将模型与YOLOv8-seg,YOLOv11-seg[30],YOLOv12-seg[31]和YOLOv13-seg[32]系列网络进行分析,引入多种评价指标,如预测阈值为0.50的平均精度(mAP50)、阈值为0.50到0.95的平均精度(mAP50-95)、浮点数(GFLOPs)、推理速度(Inference)以及权重大小(Size),对网络进行综合评估。

由表9实验结果可得如下结论:

(1)平均分割精度对比,AEISM的mAP50分割精度为84.3%,较YOLOv8n-seg,YOLOv8s-seg,YOLOv11n-seg,YOLOv11s-seg,YOLOv11m-seg,YOLOv12n-seg,YOLOv12s-seg,YOLOv13n-seg,YOLOv13s-seg,YOLOv26n-seg和YOLOv26s-seg相比分别高出5.9%,1.6%,4.6%,2.1%,1.1%,6.9%,

表 9　不同网络模型对比实验

Tab. 9　Comparative experiments of different network models

Models	mAP50/%	mAP50-95/%	GFLOPs	Size/MB	Para/M	Inference/ms
YOLOv8n-seg	78.4	44.5	12.0	6.45	3.26	1.2
YOLOv8s-seg	82.7	47.7	42.4	22.7	11.78	3.0
YOLOv11n-seg	79.7	46.7	10.2	5.73	2.84	2.0
YOLOv11s-seg	82.2	49.4	35.3	19.5	10.07	3.0
YOLOv11m-seg	83.2	52.2	123.0	43.0	22.34	6.0
YOLOv12n-seg	77.4	45.4	9.7	5.70	2.76	1.7
YOLOv12s-seg	83.1	49.9	33.3	19.0	9.73	3.3
YOLOv13n-seg	77.3	44.9	10.0	5.67	2.70	2.0
YOLOv13s-seg	81.2	47.5	34.0	19.0	9.66	3.9
YOLOv26n-seg	76.5	46.4	9.6	6.25	2.69	1.7
YOLOv26s-seg	80.0	49.7	36.6	22.2	10.37	2.4
AEISM	84.3	51.8	15.3	9.26	4.47	2.8

1.2%，7.0%，3.1%，7.8% 和 4.3%。其 mAP50-95 的平均分割精度为 51.8%，较上述模型相比分别提高 7.3%，4.1%，5.1%，2.4%，−0.4%，6.4%，1.9%，6.9%，4.3%，5.4% 和 2.1%，说明所提方法能够有效提升动态目标召回能力和实例定位能力。

(2)模型复杂度和参数量对比，AEISM 的浮点运算次数为 15.3 GFLOPs，模型大小为 9.26 MB，参数量为 4.47 M，与 YOLOv8s-seg，YOLOv11s-seg，YOLOv11m-seg，YOLOv12s-seg 和 YOLOv13s-seg 相比具有明显的轻量级优势，与轻量级 YOLOv8n-seg，YOLOv11n-seg，YOLOv12n-seg 和 YOLOv13n-seg 接近。

(3)模型分割速度对比，AEISM 的推理速度为 2.8 ms，优于 YOLOv8s-seg，YOLOv11s-seg，YOLOv11m-seg，YOLOv12s-seg 和 YOLOv13s-seg，推理速度介于 YOLOv8n-seg，YOLOv11n-seg，YOLOv12n-seg 和 YOLOv13n-seg 之间，保持了较高的实时性。

3.11　分割结果分析

本文选取 YOLOv8n-seg，YOLOv11n-seg 以及 YOLOv13n-seg 网络模型与本文提出的 AEISM 网络模型进行对比试验，并选取航拍可见光场景下实例目标与背景特征相似的问题(a)、目标边缘信息响应微弱的问题(f)与(g)、航拍弱光环境下存在低对比度及弱边缘的问题(b)与(c)，实例目标存在纹理特征相似及尺度变化的问题(d)与(e)。红色框代表漏检，橘色框代表误检，通过分析图 8 分割网络结果的可视化效果可知(彩图见期刊电子版)。

(1)面对航拍可见光场景下实例目标与背景特征相似时(a)，模型会出现识别混淆。YOLOv8n 存在行人目标与车辆目标的漏检以及背景的误检问题；YOLOv11n 对行人目标产生漏检现象；YOLOv13n 可以有效分割车辆目标，但存在会将背景误检为车辆的问题。相比之下，本文 AEISM 模型引入自适应特征聚合模块，能够更精准地提取与融合目标的关键鉴别性特征，有效增强对动态目标的判别能力，从而减少背景干扰下的漏检与误检。

(2)面对目标边缘信息响应微弱时(f)与(g)，YOLOv8n 可将图像中的车辆目标分割，而 YOLOv11n 和 YOLOv13n 则会存在将背景误检为车辆目标的问题，本文模型 AEISM 通过多尺度边缘引导选择感知模块增强对弱边缘信号的捕捉与重构，使模型能够更清晰地分辨目标轮廓，因此在实例目标边缘响应微弱时取得更准确的分割效果。

(3)面对弱光环境下存在低对比度及弱边缘问题时(b)与(c)，YOLOv8n 与 YOLOv11n 在弱光环境中存在车辆目标的漏检问题；而 YOLOv11n 与 YOLOv13n 存在背景信息的误检现象，本文模型 AEISM 通过多尺度边缘引导选择感知模块与自适应特征聚合模块的协同作用，

增强了模型对低可见度目标的关注度，提升了目标区域的响应强度，从而更好地将其与背景分离。

(4)面对航拍图像存在纹理特征相似及尺度变化的问题(d)与(e)，在(d)图像中YOLOv8n可以有效检测车辆目标，但YOLOv11n，YOLOv13n基准模型均会将背景误检为车辆目标；在(e)图像YOLOv8n与YOLOv13n基准模型均会存在车辆目标的漏检问题，而YOLOv11n则会将背景误检为车辆目标。本文模型AEISM通过引入自适应特征聚合模块与双域下采样自适应模块结合自适应特征聚合模块与双域下采样自适应模块，增强模型对图像边缘区域目标的特征挖掘能力，并实现了对不同尺度目标特征的更有效捕获，从而在航拍场景下实现更准确的分割效果。

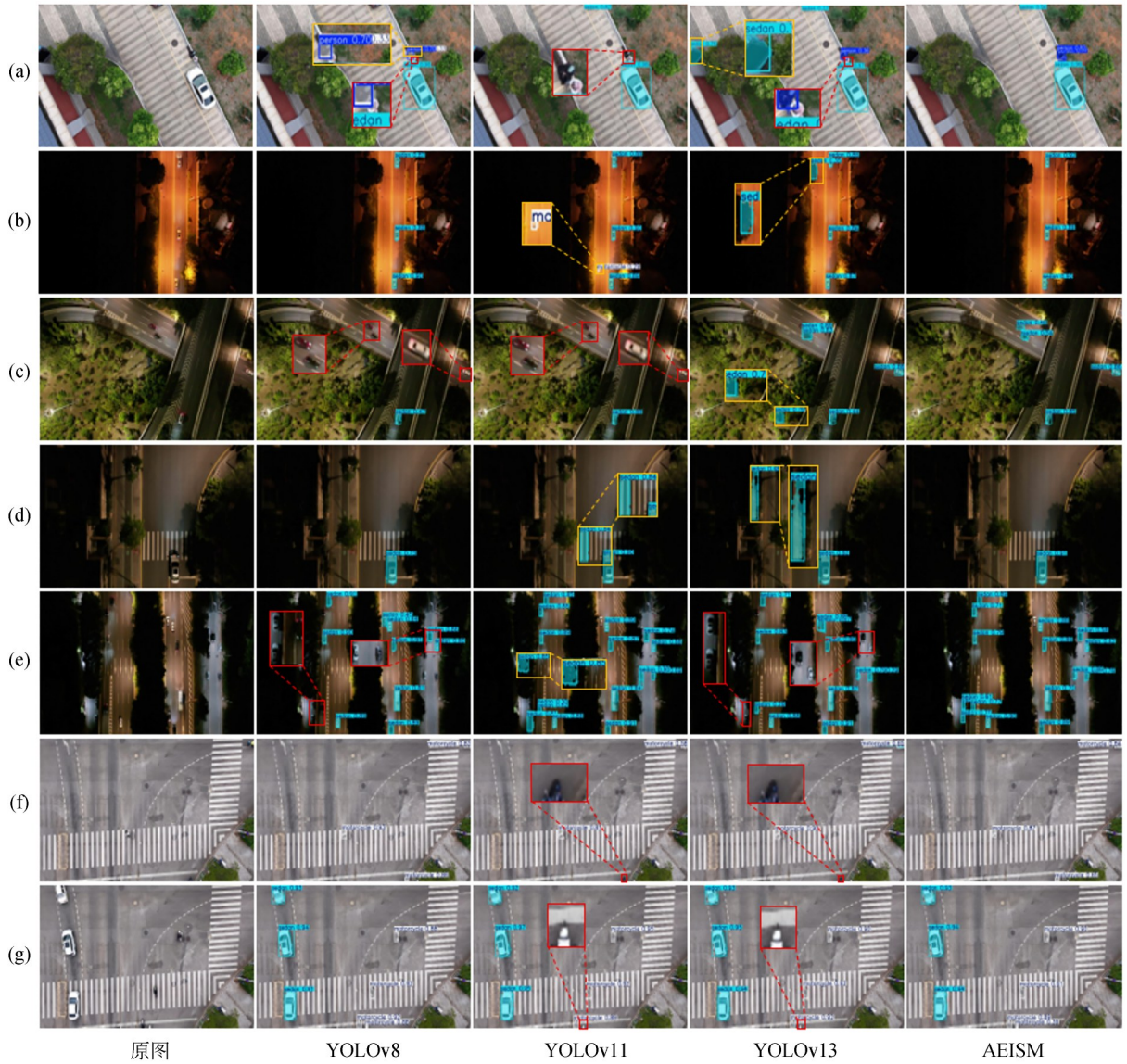

图 8　网络分割结果可视化

Fig. 8　Visualization of network segment results

3.12　复杂环境下可视化分析

在巡飞器航拍动态地物目标分割任务中，由于实例目标常处于复杂多变的自然环境中，光照条件及大气能见度等因素的剧烈变化会显著增加分割难度，因此为增强模型对复杂环境的适应能力与泛化性能，本文采用针对性数据增强策

略，通过模拟 0.2～0.5 范围浓度雾、灰色雾霾及限制型光照变化等典型环境干扰，生成高真实性的图像。红色框代表漏检，即模型未能检测出的真实地物实例；橘色框代表误检，即模型错误识别出的非目标区域。通过分析图 9 分割结果的可视化效果可知（彩图见期刊电子版）。

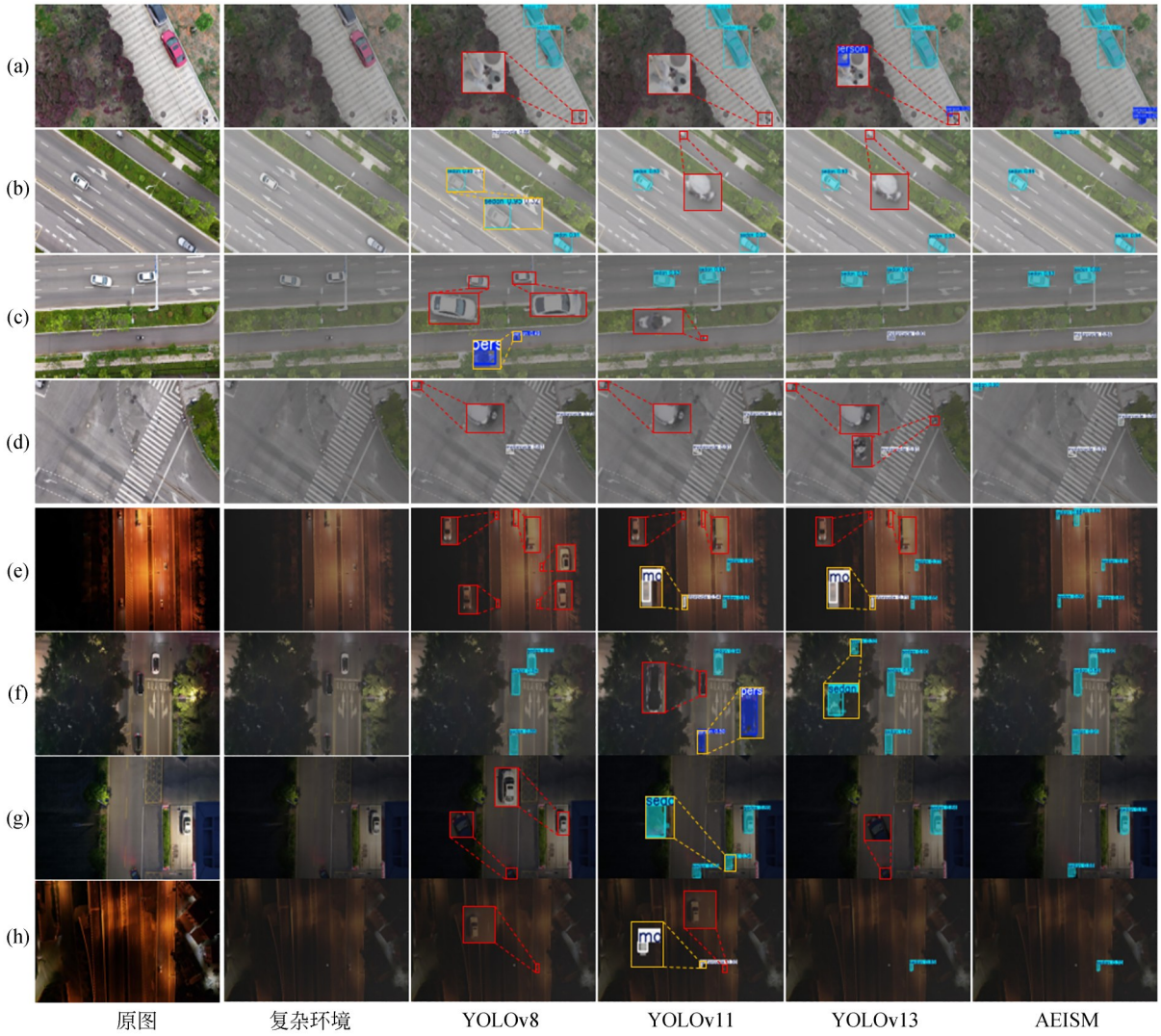

图 9　复杂环境网络分割结果可视化

Fig. 9　Visualization of complex environmental network segmentation results

3.13　模型部署

为验证本文所提模型 AEISM 能有效部署在计算资源受限的实例分割任务中，将本文边缘引导选择与特征聚合的巡飞器航拍动态目标实例分割部署到边缘设备 Jetson Nano B01，此设备技术参数如表 10 所示。

如图 10 所示为模型部署图，将获取到的巡飞器航拍图像数据作为基准模型的输入，在 PC 端进行批处理大小为 200 次的模型训练得到 AEISM 模型的预训练权重，得到的预训练权重部署到 Jetson Nano B01 边缘设备，并在该设备运行预训练检测模型对动态目标进行推理，输出动态目标类别及目标置信度，实验结果如表 11 所示。

由表 11 实验结果的定量化分析可知，本文提出的边缘引导选择与特征聚合的巡飞器航拍动态目标实例分割模型部署到边缘设备时，巡飞器航拍图像的动态目标类别 Person，sedan，motorcycle 较基准模型 YOLOv11n-seg 及 YOLOv26n-seg 均有所提升，其中 Person 类别的检测精度

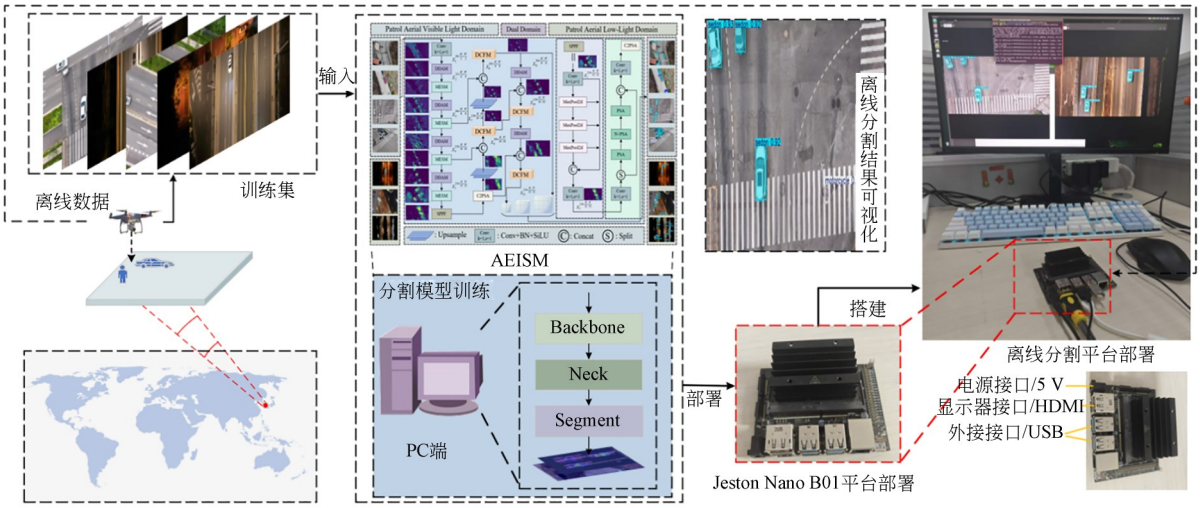

图 10　模型部署

Fig. 10　Model deployment

表 10　Jetson Nano B01 参数表

Tab. 10　Jetson Nano B01 technical data

组件	参数
CPU	四核 ARM Cortex-A57@143 GHz
GPU	128 核 NVIDIA MaxellTM GPU
显存	4 GB 64 位 LPDDR4 内存
存储	32 G Micro SD 卡
浮点数	473 GFLOPS
规格尺寸/mm	100×80×29

表 11　不同模型部署分割结果

Tab. 11　Deployment and segmentation results of different models

Classes	Model	mAP50/%	mAP50-95/%
Person	YOLOv11n-seg	69.5	27.0
	YOLO26n-seg	64.4	27.9
	AEISM	76.1	34.3
Sedan	YOLOv11n-seg	87.7	67.9
	YOLOv26n-seg	84.0	64.5
	AEISM	87.8	69.6
Motorcycle	YOLOv11n-seg	81.9	45.3
	YOLOv26n-seg	81.2	46.8
	AEISM	89.1	51.6

mAP50 和 mAP50-95 分别达到 76.1% 和 34.3%；对 sedan 类别的检测精度 mAP50 和 mAP50-95 分别达到 87.8% 和 69.6%；对 motorcycle 类别的检测精度 mAP50 和 mAP50-95 分别达到 89.1% 和 51.6%。相比之下，本文提出的模型 AEISM 相比于基准模型 YOLOv11n-seg 及 YOLOv26n-seg 在动态目标分割精度上均有所提升。

4　结　论

本文针对巡飞器航拍动态地物目标在可见光与弱光双源环境下存在的类间特征相似、弱边缘感知不足的问题，提出边缘引导选择与特征聚合的动态目标实例分割模型。以 YOLOv11n-seg 为基础框架，围绕边缘先验显式选择-空间/通道动态互补聚合-双域下采样高频补偿的设计链路，对主干特征提取、颈部特征融合和下采样过程进行针对性优化，使模型由通用特征增强转向面向弱边缘保持、相似背景区分和尺度压缩补偿的协同建模。

通过构建可见光与弱光环境下的航拍动态目标实例分割数据集进行实验验证，本文提出的 AEISM 模型在动态地物目标分割任务中取得了较好的性能，其 mAP50 和 mAP50-95 分别达到 84.3% 和 51.8%，相较于基准模型 YOLOv11n-seg 分别提升 4.6% 和 5.1%。实验结果表明，本文方法能够有效增强模型对动态目标的识别与分割能力，提高复杂航拍环境下目标边界感知、语义判别和实例区域定位的准确性。

本文的主要贡献如下：

(1)设计多尺度边缘引导选择感知模块MESM。针对弱光、俯视和运动成像条件下目标边缘响应较弱、主干网络早期卷积容易平滑边界信息的问题，在语义抽象前显式引入边缘先验，通过多尺度特征与局部均值之间的残差响应生成边缘权重图，并结合空间—通道双域选择机制强化弱边缘结构感知，从而改善动态目标边缘模糊导致的分割精度下降问题。

(2)构建自适应特征聚合模块AFAM。针对车辆、行人与道路、阴影、建筑等背景区域存在较强外观相似性，导致模型在特征融合阶段容易产生语义混淆的问题，通过通道混洗实现跨通道语义交互，并结合动态卷积与深度动态卷积自适应聚合空间语义信息和通道细节特征，从而提升模型对目标与相似背景的判别能力。

(3)设计双域下采样自适应模块DDAM。针对普通下采样过程容易压缩小目标边缘、高频纹理和局部结构信息的问题，构建局部卷积与空洞卷积并行的双域下采样结构，同步提取局部细节信息与全局上下文特征，并通过高频增强机制补偿尺度变换过程中的边缘与纹理损失，从而增强模型对多尺度空间结构的表达能力。

作者贡献声明：

宫冕：负责论文的整体撰写，创新点的构思与设计，以及实验的初步实施；

张印辉：作为导师，提供了论文的理论指导，对研究方法论进行了深入的指导和建议；

何自芬：作为导师，对论文进行了理论指导，并对研究的创新性和实用性进行了评估；

陈光晨：作为宫冕的师兄，对论文的创新改进提供了宝贵的指导和建议。

参考文献：

[1] 卢汉，崔博伦，万华洋，等. 半监督式野生动物夜间目标端到端检测[J]. 光学 精密工程，2025，33(5)：789-801.
LU H，CUI B L，WAN H Y，*et al*. End-to-end recognition of nighttime wildlife based on semi-supervised learning[J]. *Opt. Precision Eng.*，2025，33(5)：789-801.(in Chinese)

[2] WANG H，ZHANG X H，MENG X Y，*et al*. Electronic sheepdog：a novel method in with UAV-assisted wearable grazing monitoring[J]. *IEEE Internet of Things Journal*，2023，10(18)：16036-16047.

[3] 陈仁祥，邓力珩，杨黎霞，等. 改进YOLO11n的雾天路面缺陷轻量化检测[J]. 光学 精密工程，2026，34(4)：640-651.
CHEN R X，DENG L H，YANG L X，*et al*. Lightweight detection of foggy pavement defects based on improved YOLO11n[J]. *Opt. Precision Eng.*，2026，34(4)：640-651.(in Chinese)

[4] 刘媛媛，朱凯，顾志辉，等. 自适应特征的轻量化路面裂缝检测方法[J]. 光学 精密工程，2026，34(2)：336-351.
LIU Y Y，ZHU K，GU Z H，*et al*. Lightweight road crack detection method with adaptive features[J]. *Opt. Precision Eng.*，2026，34(2)：336-351.(in Chinese)

[5] KONG X J，NI C H，DUAN G H，*et al*. Energy consumption optimization of UAV-assisted traffic monitoring scheme with tiny reinforcement learning[J]. *IEEE Internet of Things Journal*，2024，11(12)：21135-21145.

[6] 彭勃，白吉康，陈伟文，等. 基于深度学习的无人机搜救方法研究进展[J]. 航空学报，2025，46(23)：1-18.
PENG B，BAI J K，CHEN W W，*et al*. Research progress for UAV search and rescue methods based on deep learning[J]. *Acta Aeronautica et Astronautica Sinica*，2025，46(23)：1-18.(in Chinese)

[7] KHALIL H，RAHMAN S U，ULLAH I，*et al*. A UAV-swarm-communication model using a machine-learning approach for search-and-rescue applications[J]. *Drones*，2022，6(12)：372.

[8] 姚雨，宋春林，邵江琦. 无人机航拍军事车辆实时检测及定位算法[J]. 兵工学报，2024，45(S1)：354-360.
YAO Y，SONG C L，SHAO J Q. Real-time detection and localization algorithm for military vehicles in drone aerial photography[J]. *Acta Armamentarii*，2024，45(S1)：354-360.(in Chinese)

[9] 王乾胜，展勇忠，邹宇. 基于改进Yolov5n的无人机对地面军事目标识别算法[J]. 计算机测量与控制，2024，32(6)：189-197，226.

WANG Q S, ZHAN Y Z, ZOU Y. Recognition algorithm for UAV ground military targets based on improved Yolov5n[J]. *Computer Measurement & Control*, 2024, 32(6): 189-197, 226. (in Chinese)

[10] 朱广, 顾晨, 徐立云, 等. 改进YOLOv8的风机叶片多尺度缺陷检测[J]. 光学 精密工程, 2025, 33(9): 1496-1514.
ZHU G, GU C, XU L Y, *et al*. Improvement of YOLOv8 for multi-scale defect detection in wind turbine blades[J]. *Opt. Precision Eng.*, 2025, 33(9): 1496-1514. (in Chinese)

[11] 刘传洋, 吴一全, 刘景景. 无人机航拍图像中绝缘子缺陷检测的深度学习方法研究进展[J]. 电工技术学报, 2025, 40(9): 2897-2916.
LIU C Y, WU Y Q, LIU J J. Research progress of deep learning methods for insulator defect detection in UAV based aerial images[J]. *Transactions of China Electrotechnical Society*, 2025, 40(9): 2897-2916. (in Chinese)

[12] ZAMIR S W, ARORA A, GUPTA A, *et al*. iSAID: A Large-scale Dataset for Instance Segmentation in Aerial Images[EB/OL]. 2019: *arXiv*: 1905.12886. https://arxiv.org/abs/1905.12886

[13] HE K M, GKIOXARI G, DOLLAR P, *et al*. Mask R-CNN[C]. 2017 *IEEE International Conference on Computer Vision* (*ICCV*). *October* 22-29, 2017. *Venice*. IEEE, 2017: 2980-2988.

[14] WANG K X, LIEW J H, ZOU Y T, *et al*. PANet: few-shot image semantic segmentation with prototype alignment[C]. 2019 *IEEE/CVF International Conference on Computer Vision* (*ICCV*). *October* 27-*November* 2, 2019. *Seoul*, *Korea*. IEEE, 2019: 9196-9205.

[15] GARG P, CHAKRAVARTHY A S, MANDAL M, *et al*. ISDNet: AI-enabled instance segmentation of aerial scenes for smart cities[J]. *ACM Transactions on Internet Technology*, 2021, 21(3): 1-18.

[16] SU H, WEI S J, LIU S, *et al*. HQ-ISNet: high-quality instance segmentation for remote sensing imagery[J]. *Remote Sensing*, 2020, 12(6): 989.

[17] CAI Z W, VASCONCELOS N. Cascade R-CNN: delving into high quality object detection[C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. *June* 18-23, 2018, *Salt Lake City*, *UT*, *USA*. IEEE, 2018: 6154-6162.

[18] CHEN H, SUN K Y, TIAN Z, *et al*. BlendMask: top-down meets bottom-up for instance segmentation[C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 13-19, 2020, *Seattle*, *WA*, *USA*. IEEE, 2020: 8570-8578.

[19] LEE Y, PARK J. CenterMask: Real-time Anchor-free Instance Segmentation[C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 13-19, 2020, *Seattle*, *WA*, *USA*. IEEE, 2020: 13903-13912.

[20] HUANG P, YIN Y, HU K F, *et al*. MonoSeg: an infrared UAV perspective vehicle instance segmentation model with strong adaptability and integrity[J]. *Sensors*, 2025, 25(1): 225.

[21] HAN K, WANG Y H, TIAN Q, *et al*. GhostNet: more features from cheap operations[C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 13-19, 2020, *Seattle*, *WA*, *USA*. IEEE, 2020: 1577-1586.

[22] WANG C Y, MARK LIAO H Y, WU Y H, *et al*. CSPNet: a new backbone that can enhance learning capability of CNN[C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (*CVPRW*). *June* 14-19, 2020, *Seattle*, *WA*, *USA*. IEEE, 2020: 1571-1580.

[23] ZHOU J K, WANG Y, ZHOU W L. Efficient instance segmentation framework for UAV-based pavement distress detection[J]. *Automation in Construction*, 2025, 175: 106195.

[24] HUANG Y, ZHANG Y H, HE Z F, *et al*. Video instance segmentation through hierarchical offset compensation and temporal memory update for UAV aerial images[J]. *Sensors*, 2025, 25(14): 4274.

[25] HANYU T, YAMAZAKI K, TRAN M, *et al*. AerialFormer: multi-resolution transformer for aerial image segmentation[J]. *Remote Sensing*, 2024, 16(16): 2930.

[26] KIRILLOV A, MINTUN E, RAVI N, *et al*. Segment anything[C]. 2023 *IEEE/CVF International Conference on Computer Vision* (*ICCV*). *October* 1-6, 2023, *Paris*, *France*. IEEE, 2024: 3992-4003.

[27] ZHANG E K, LIU J J, CAO A D, *et al*. RS-SAM: integrating multi-scale information for enhanced remote sensing image segmentation[C].

Computer Vision-ACCV 2024. Singapore: *Springer*, 2025: 280-296.

[28] ZHANG P Y, ZHAO J, WANG D, *et al*. Visible-thermal UAV tracking: a large-scale benchmark and new baseline [C]. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 18-24, 2022, *New Orleans*, *LA*, *USA*. IEEE, 2022: 8876-8885.

[29] ZHANG P, ZHAO J, WANG D, *et al*. Visible-thermal UAV tracking: A large-scale benchmark and new baseline [C]. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). IEEE, 2022: 8876-8885.

[30] KHANAM R, HUSSAIN M. YOLOv11: An Overview of the Key Architectural Enhancements [EB/OL]. 2024: *arXiv*: 2410.17725. https://arxiv.org/abs/2410.17725.

[31] TIAN Y J, YE Q X, DOERMANN D. YOLOv12: Attention-centric Real-time Object Detectors [EB/OL]. 2025: *arXiv*: 2502.12524. https://arxiv.org/abs/2502.12524

[32] LEI M Q, LI S Q, WU Y H, *et al*. YOLOv13: Real-time Object Detection with Hypergraph-enhanced Adaptive Visual Perception [EB/OL]. 2025: *arXiv*: 2506.17733. https://arxiv.org/abs/2506.17733

作者简介：

宫　冕(2001—),男,山东泰安人,硕士研究生,2023年于潍坊学院获得学士学位,主要从事图像处理、机器视觉及机器智能等方面的研究。E-mail: 13853877423@163.com

通讯作者：

张印辉(1977—),男,河北故城人,博士,教授。博士生导师,2000年、2005年于西安理工大学分别获得学士和硕士学位,2010年于昆明理工大学获得博士学位,主要从事图像处理、机器视觉及机器智能等方面的研究。E-mail: zhangyinhui@kust.edu.cn