北京林业大学学报
JOURNAL OF BEIJING FORESTRY UNIVERSITY

掩膜分辨率自适应与边缘协同优化的野生动物实例分割

杨锋 赵超越 王升 韩柏迅 杨林哲 许福

Adaptive mask resolution with collaborative edge optimization for wildlife instance segmentation

Yang Feng, Zhao Chaoyue, Wang Sheng, Han Boxun, Yang Linzhe, Xu Fu

引用本文:
杨锋, 赵超越, 王升, 韩柏迅, 杨林哲, 许福. 掩膜分辨率自适应与边缘协同优化的野生动物实例分割[J]. 北京林业大学学报, 2026, 48(2):26–42. doi: 10.12171/j.1000–1522.20250211
Yang Feng, Zhao Chaoyue, Wang Sheng, Han Boxun, Yang Linzhe, Xu Fu. Adaptive mask resolution with collaborative edge optimization for wildlife instance segmentation[J]. Journal of Beijing Forestry University, 2026, 48(2):26–42. doi: 10.12171/j.1000–1522.20250211

在线阅读 View online: https://doi.org/10.12171/j.1000–1522.20250211

您可能感兴趣的其他文章

Articles you may be interested in

基于实例分割模型的原木检尺径方法

A method of log diameter measurement based on instance segmentation model

北京林业大学学报, 2023, 45(3): 153–159. https://doi.org/10.12171/j.1000–1522.20220345

基于高空间分辨率时间序列影像的对象级侵占林地图斑快速检测

Rapid detection of object–level invaded forest map patch based on high spatial resolution time series image

北京林业大学学报, 2022, 44(11): 60–69. https://doi.org/10.12171/j.1000–1522.20210261

基于YOLOv7的红外相机野生动物图像筛选

Wildlife image screening for infrared cameras based on YOLOv7

北京林业大学学报, 2024, 46(2): 143–154. https://doi.org/10.12171/j.1000–1522.20230112

MAFF–YOLO: 面向造林验收的明穴目标检测模型

MAFF–YOLO: a target detection model for planting holes in afforestation acceptance

北京林业大学学报, 2025, 47(4): 142–154. https://doi.org/10.12171/j.1000–1522.20240353

基于无人机影像和MDIEA–YOLO苗木识别模型的造林验收智能系统

An intelligent system for afforestation acceptance based on UAV images and MDIEA–YOLO seedling recognition model

北京林业大学学报, 2025, 47(5): 14–25. https://doi.org/10.12171/j.1000–1522.20240354

基于树皮纹理的轻量化YOLOv11树种识别方法

Lightweight YOLOv11 tree species recognition method based on bark texture

北京林业大学学报, 2025, 47(8): 134–148. https://doi.org/10.12171/j.1000–1522.20250151

DOI:10.12171/j.1000−1522.20250211

掩膜分辨率自适应与边缘协同优化的野生动物实例分割

杨 锋 赵超越 王 升 韩柏迅 杨林哲 许 福
（北京林业大学信息学院（人工智能学院），国家林业和草原局林业智能信息处理工程技术研究中心，
林木资源高效生产全国重点实验室，北京 100083）

摘要:【目的】野生动物实例分割是获取高质量动物行为追踪、栖息地利用分析和种群动态统计等生态参数的基础。现有模型普遍采用编码器−解码器架构，通过多级下采样提取特征后固定上采样至原始分辨率生成掩膜，这种标准流程不仅因下采样与固定感受野卷积破坏目标边界结构连续性，导致个体轮廓模糊破碎，更未能依据像素所处遮挡复杂度、边缘锐度等上下文实施难度感知的资源分配，严重制约了行为学、种群动态学等高精度生态分析的数据信度。为应对上述挑战，本研究构建一个包含 4 231 张图像的综合性野生动物数据集，并提出融合动态分辨率与边缘增强的实例分割框架，旨在通过难度自适应的资源调度机制实现野外复杂场景下动物个体的精准轮廓提取。**【方法】**该框架通过通道注意力深度自适应模块（SE-DAM）集成可变形卷积，精准捕捉细粒度空间细节以强化特征定位能力；并采用自适应分辨率模块（ARM），基于实例尺度与遮挡特性动态调整掩膜分辨率，显著缓解下采样所致的空间信息损失；同时引入概率驱动协同优化分割模块（PDCO），利用离散余弦变换（DCT）进行像素难度分类，并建模难易区域间的相互依赖关系，通过优先处理简单像素进而指导复杂区域细化，避免了冗余的前景/背景计算。3 个模块在特征提取、分辨率决策和掩膜细化环节形成闭环优化，共同提升对遮挡、伪装及小目标的分割鲁棒性。本方法在实验数据集上通过 mAP、mI^{2}oU、FPS 等多指标定量评估、主流基线模型性能对比及消融实验验证了其分割精度与效率优势。**【结果】**基于 ResNet-101-FPN 骨干网络，本方法 mAP 达 53.7%，mI^{2}oU 达 86.71%，较 Mask R-CNN 基线分别提升了 11.2 和 7.7 个百分点，浮点运算量仅提升 0.3 GFLOPs，帧率稳定在 12.4 帧/s，验证了算法在精度与效率间的有效平衡。当采用 Swin-B 骨干网络时，mAP 进一步达到 54.8%，明显优于 Mask Transfiner（50.8%）与 QueryInst（52.4%）等现有最优方法，展现出良好的骨干网络泛化能力。此外，相较于采用 112 × 112 像素固定分辨率掩膜的方法，动态分辨率策略的浮点运算量为 1.31 GFLOPs，显著低于对比方法，且在相似个体堆叠、枝叶遮挡及红外图像等复杂场景中均保持优异的分割性能，证实了方法对野外监测环境的强适应性。**【结论】**本文提出的融合动态掩模分辨率与边缘细节增强的野生动物实例分割方法，通过难度感知的自适应资源分配机制有效解决了传统方法中固定分辨率导致的边界模糊与计算冗余问题，在遮挡、伪装和小目标等挑战性场景下的识别精度与轮廓质量显著提升，为高效、高精度的野生动物自动监测与生态研究提供了可直接部署的技术方案，对推动智慧生态保护具有明确的应用价值。

关键词： 计算机视觉；实例分割；自适应分辨率；遮挡感知分割；小目标分割；难度感知学习；摄像机陷阱图像分析；野生动物监测

中图分类号： TP751.1；Q958.1 **文献标志码：** A **文章编号：** 1000−1522(2026)02−0026−17

引文格式： 杨锋，赵超越，王升，等. 掩膜分辨率自适应与边缘协同优化的野生动物实例分割［J］. 北京林业大学学报，2026，48(2)：26−42. Yang Feng, Zhao Chaoyue, Wang Sheng, et al. Adaptive mask resolution with collaborative edge optimization for wildlife instance segmentation［J］. Journal of Beijing Forestry University, 2026, 48(2): 26−42.

Adaptive mask resolution with collaborative edge optimization for wildlife instance segmentation

收稿日期： 2025−06−13 **修回日期：** 2025−12−10

基金项目： 国家重点研发计划项目（2022YFF1302700），中央高校基本科研业务费优秀青年团队项目（QNTD202504）。

第一作者： 杨锋。主要研究方向：人工智能。Email：fengyang@bjfu.edu.cn 地址：100083 北京市海淀区清华东路 35 号北京林业大学信息学院（人工智能学院）。

责任作者： 许福，教授。主要研究方向：智慧林业、智慧园林、人工智能、海量遥感信息处理。Email：xufu@bjfu.edu.cn 地址：同上。

本刊网址： https://j.bjfu.edu.cn；https://journal.bjfu.edu.cn

Yang Feng　Zhao Chaoyue　Wang Sheng　Han Boxun　Yang Linzhe　Xu Fu
(School of Information Science and Technology (School of Artificial Intelligence), Beijing Forestry University, Engineering Research Center for Forestry-oriented Intelligent Information Processing, State Key Laboratory of Efficient Production of Forest Resources, Beijing 100083, China)

Abstract: [**Objective**] Wildlife instance segmentation underpins critical ecological analyses, including behavior tracking, habitat utilization assessment, and population dynamics monitoring. Existing encoder-decoder architectures extract features via multi-level downsampling and then apply fixed upsampling to generate masks. This pipeline degrades boundary continuity, yielding blurry and fragmented contours, and lacks difficulty-aware resource allocation based on pixel-wise occlusion complexity and edge sharpness. These limitations severely compromise data fidelity for high-precision ecological research. [**Method**] To address these challenges, we construct a comprehensive wildlife dataset comprising 4 231 images and propose a novel instance segmentation framework that fuses dynamic resolution with edge enhancement. Our approach enables accurate animal contour extraction through difficulty-adaptive resource scheduling. The framework integrates deformable convolutions within a Channel Attention Deep Adaptive Module (SE-DAM) to capture fine-grained spatial details. An Adaptive Resolution Module (ARM) dynamically adjusts mask resolution based on instance scale and occlusion characteristics, mitigating spatial information loss. Additionally, a Probability-Driven Collaborative Optimization (PDCO) module employs Discrete Cosine Transform (DCT) for pixel difficulty classification and models interdependencies between easy and hard regions. By prioritizing simple pixels to guide the refinement of complex regions, PDCO avoids redundant foreground/background computations. These three modules form a closed-loop optimization across feature extraction, resolution decision, and mask refinement, collectively enhancing robustness to occlusion, camouflage, and small targets. [**Result**] Experimental results demonstrate that with a ResNet-101-FPN backbone, our method achieves an mAP of 53.7% and an mI^2oU of 86.71%, surpassing Mask R-CNN by 11.2 and 7.7 percentage points, respectively. The computational cost increases by only 0.3 GFLOPs with an inference speed of 12.4 FPS, showing an effective balance between accuracy and efficiency. Using a Swin-B backbone, the mAP reaches 54.8%, outperforming state-of-the-art methods such as Mask Transfiner (50.8%) and QueryInst (52.4%). Our dynamic resolution strategy maintains superior performance in challenging scenarios, including overlapping individuals, branch occlusion, and infrared imagery, while achieving significantly lower computational costs than fixed-resolution approaches. [**Conclusion**] We present a wildlife instance segmentation method that integrates dynamic mask resolution with edge detail enhancement. Through a difficulty-aware adaptive resource allocation mechanism, our approach effectively mitigates boundary blur and computational redundancy inherent in conventional fixed-resolution methods. This yields significant improvements in recognition accuracy and contour quality under challenging conditions including occlusion, camouflage, and small targets. The framework provides a deployable solution for efficient, high-precision automatic wildlife monitoring and ecological research, offering clear practical value for intelligent ecological conservation.

Key words: computer vision; instance segmentation; adaptive resolution; occlusion-aware segmentation; small object segmentation; difficulty-aware learning; camera trap image analysis; wildlife monitoring

受栖息地丧失、气候变化及人类活动影响，全球野生动物种群数量正显著下降，这使得生物多样性保护与生态平衡维护成为全球紧迫议题[1]。野生动物作为物种多样性中的重要一环，是自然和人类共同的宝贵财富，因此野生动物监测起着重要作用[2]。有效的野生动物监测能提供种群规模、分布规律及行为特征等关键数据，对于生态动态理解[3]，潜在威胁识别以及保护成效评估[4]具有决定性作用。

计算机视觉技术的快速发展为野生动物保护工作带来了革命性变化[5]。通过提供更高效、精准且可扩展的监测手段，该技术显著提升了相关研究的效率与能力[6]。然而，尽管技术不断进步，在自然环境中复杂场景下对野生动物进行准确个体区分[7]与追踪[8]仍然面临严峻挑战。作为计算机视觉的重要分

支，实例分割技术能够精确分割图像中每个独立目标，并为不同实例赋予唯一标签[9]，为解决上述问题提供了新的技术路径。这一方法为野生动物研究开辟了新途径，有助于实现精确的个体识别[10]、细致的行为分析以及可靠的种群数量估算[11]。

为此，研究者已提出多种基于卷积神经网络的动物实例分割方法。其中，Mask R-CNN 模型[12]被成功应用于牲畜自动检测与计数任务，通过生成二值掩膜实现像素级分割，但其 RoI Align 操作在目标严重遮挡时易引入空间错位误差，导致处理尺寸差异显著的野生动物个体时效率较低；CondInst 网络[13]创新性地采用动态实例感知机制，规避了传统 ROI 裁剪与特征对齐操作，在提升分割精度的同时，显著加快推理速度，但由于掩膜头设计仍依赖固定分辨率特征图，导致难以自适应地捕捉复杂背景中野生动物的精细边缘细节；SOLOv2 模型[14]则引入矩阵非极大值抑制技术，有效增强了分割精确性与运算效率，但大范围下采样导致的特征图分辨率损失仍制约了其对小目标及模糊边界的分割精度；最新提出的 Mask Transfiner 模型[15]通过基于四叉树的 Transformer 网络架构，可针对性优化易错区域分割结果，进一步提升整体性能，但其分层处理机制计算开销较高，在资源受限的野外监测设备上部署面临挑战。

计算机视觉迭代虽将野生动物监测推向像素级精度时代，但上述模型仍沿用“全图均一分辨率、全像素等量计算”范式。面对动物与背景难辨、光照多变、个体遮挡和种内差异，现有实例分割方法的两大局限被成倍放大：其一，由于下采样与卷积操作导致的空间信息丢失，模型难以捕捉细粒度特征（如目标边界与复杂纹理），进而产生边界模糊与定位精度下降的问题，尤其在枝叶遮挡或动物躯干局部暴露的场景中，Mask R-CNN、CondInst 等方法常将背景植被的纹理错误地融入动物轮廓，如大熊猫（*Ailuropoda melanoleuca*）黑色毛发边缘的绿色晕染，或无法清晰界定动物与相似背景的交界，如远东豹（*Panthera pardus orientalis*）在枯草中的轮廓模糊；其二，这些方法通常对图像所有像素进行均一化处理，忽视了不同区域分割难度的差异性，对扭角羚（*Budorcas tibetana*）背部大片均匀毛皮区域等前景主体和背景等简单区域投入了过多计算资源，却对真正决定分割精度的边界像素（如东北虎 *Panthera tigris altaica* 与背景交界的过渡区）及严重遮挡部位等复杂区域关注不足或处理能力有限。虽然背景或目标表面等简单区域较易分割[16]，但边界区域及遮挡部位等复杂场景仍存在显著挑战[17]。现有方法因未能区分这些差异性，导致计算资源分配失当，最终在关键挑战区域表现欠佳。例如，QueryInst 方法[18]采用动态卷积生成掩膜，但未显式分离前景/背景特征。当背景像素占比显著高于目标边缘时（如枝叶遮挡动物躯干），卷积核采样会偏向背景主导区域，导致上采样后掩膜分配错误，且密集预测机制计算开销较大；TensorMask 方法[19]基于四维张量的密集滑窗预测框架缺乏通道注意力引导，在低对比度场景（如深色皮毛与阴影）中，因纹理特征采样丢失，将相似亮度区域归类为同质块，同时全局参数计算导致冗余开销。

针对上述边界模糊与空间信息损失，以及资源分配失当导致复杂区域性能不佳等关键挑战，本研究拟开展以下系统性工作以提升野生动物实例分割精度与效率。（1）构建一个涵盖大熊猫、东北虎、远东豹、赛加羚羊（*Saiga tatarica*）及金丝猴（*Rhinopithecus roxellanae*）5 种代表性野生动物的高质量实例分割数据集，目标包含 4 231 幅标注精确的图像，旨在为复杂野外场景下的模型训练与评估提供标准化数据支撑。（2）引入通道注意力深度自适应模块（SE-depthwise adaptive module，SE-DAM），通过融合通道注意力机制与可变形卷积，旨在动态优化卷积核参数，重点提升模型对目标边界等复杂空间结构的特征表征能力与定位精度。（3）开发自适应分辨率模块（adaptive resolution module，ARM），依据实例特征分布动态确定最优掩膜分辨率，力求在保留关键空间细节与提升分割准确性的同时，显著降低冗余计算开销。（4）提出一种基于块离散余弦变换的条件随机场优化器（patch-based discrete cosine transform conditional random field optimizer，PDCO），根据特征复杂度将像素分类为易分割与难分割类别，通过条件随机场建模像素类别间上下文关联性，以高置信度像素的可靠信息为引导，驱动对低置信度像素的精细化分割过程，实现分割掩膜优化。

1 掩膜野生动物实例分割方法

1.1 基于 Mask R-CNN 框架实现特征增强与分割优化

本研究针对野生动物图像中目标易伪装、边缘模糊、存在遮挡的特殊挑战，通过改进 Mask R-CNN 基础框架来实现实例分割精度与计算效率的平衡。如图 1 所示，改进后的模型架构旨在增强野生动物图像中个体实例的表征能力：在 FPN 和 RPN 输出端集成改进的 SE-DAM 模块[20]，在借鉴通道注意力机制提取多通道形态和纹理信息的基础上，融合可变形卷积（deformable convolution）机制动态调整特征采样位置，有效补偿野生动物运动与遮挡导致的非线性偏移，并结合深度可分离卷积（depthwise separable

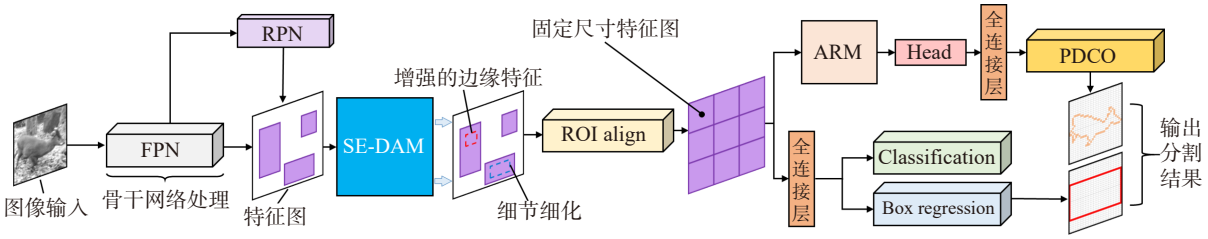


图 1　自适应掩膜深度学习野生动物实例分割方法总体架构图

Fig. 1　Overall architecture diagram of the adaptive-mask deep learning method for wildlife instance segmentation

convolution)特性，增强模型对非刚性形变的适应能力，显著提升特征提取精度并降低误差；通过 RoI Align 双线性插值精确处理 ROI 特征后，模型执行物种分类与边界框定位；ARM 模块引入通道注意力机制有效区分不同野生动物实例，根据实例尺寸与复杂度动态调整掩膜分辨率；最终，PDCO 模块对掩膜边缘进行分类，结合条件随机场(Conditional Random Fields，CRF)进一步优化掩膜边缘，以适应野生动物图像的高边缘变异特性，从而提升模型在复杂自然背景下的分割质量。

1.2　引入 SE-DAM 解决特征空间错位与 RoI 对齐误差

针对野生动物因躲避、休憩或快速移动导致的局部暴露问题，以及 RoI Align 阶段植被遮挡与个体重叠造成双线性插值难以处理非线性偏移的难题，本文提出如图 2 所示的 SE-DAM 模块解决方案。

该模块采用通道注意力机制，自适应强化因动物姿态变化或局部遮挡而受阻的关键特征。摒弃传统模型对所有特征通道均等加权的策略，优先强化对动物边缘、纹理等关键信息敏感的通道。具体而言，先对输入特征图F执行全局平均池化生成通道描述符z_c，使网络动态识别并增强对动物运动及遮挡等变化敏感的通道；继而通过可学习权重，按以下公式对各通道响应强度进行重标定。

$$z_c = \mathrm{GAP}(\boldsymbol{F}_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \boldsymbol{F}_{ijc} \tag{1}$$

$$s = \sigma(\boldsymbol{W}_2 \delta(\boldsymbol{W}_1 z)) \tag{2}$$

式中：δ表示 ReLU 激活函数，σ代表 Sigmoid 函数，z_c表示经全局平均池化得到的通道描述符，$\boldsymbol{F}_c$为输入特征图，H和W分别为特征图的高度与宽度，$\boldsymbol{F}_{ijc}$表示通道c中(i,j)处的特征值，s为通道注意力机制中的缩放参数，$\boldsymbol{W}_1$和$\boldsymbol{W}_2$为全连接层的权重矩阵，z为经过通道重标定后的调整特征。

针对快速运动与局部遮挡导致的特征错位问题，SE-DAM 模块引入可变形卷积机制。采用 3 × 3 卷积层计算偏移量Δp，通过 4 个最近邻整数坐标点(i,j)、$(i+1,j)$、$(i,j+1)$和$(i+1,j+1)$(其中$i=\lfloor x \rfloor$，$j=\lfloor y \rfloor$)来适应非线性偏移或非整数位置(x,y)的情况，权重根据偏移点到这 4 个参考点的距离进行计算。

随后基于上述权重对 4 个坐标点的像素值进行加权平均，以补偿偏移量并估算目标像素值$V_{(x,y)}$。

$$V_{(x,y)} = a \cdot f(i,j) + b \cdot f(i+1,j) + c \cdot f(i,j+1) + d \cdot f(i+1,j+1) \tag{3}$$

式中：f表示对应位置的像素值。

随后通过可变形卷积层按以下公式调整特征图。

$$F'_{ij} = \sum_{n=1}^{N} \omega_n \cdot F(i + \Delta p^x_{i,j,n}, j + \Delta p^y_{i,j,n}) \tag{4}$$

式中：F'_{ij}表示可变形卷积操作后位置(i,j)处的新特征值，ω_n代表卷积核第n个元素的权重，$\Delta p^x_{i,j,n}$和$\Delta p^y_{i,j,n}$分别表示卷积核在x和y方向上的偏移量，N表示可变形卷积核的采样点总数，采用标准 3 × 3 卷积核，N的值为 9。

通过这一过程，SE-DAM 模块在 RoI Align 操作前自适应调整特征图的空间排布，使其更精确地反

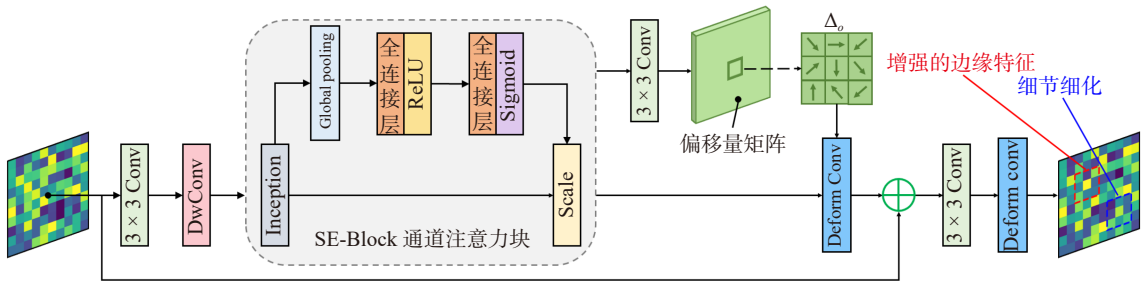


图 2　通道注意力深度自适应模块(SE-DAM)结构示意图

Fig. 2　Schematic diagram of the channel SE-depthwise adaptive module (SE-DAM)

映野生动物的实际空间分布，从而有效减少上采样和 RoI Align 过程中的错位误差与信息损失。

1.3 引入自适应分辨率模块解决掩膜分辨率分配与背景分离问题

针对野生动物与自然环境高度相似的色彩、纹理及轮廓特征（即天然伪装能力）导致的背景分离难题，传统固定分辨率掩膜因无法适应不同实例的特性，往往导致分割细节丢失或计算资源浪费。为此，本文参考 Li 等[9]提出的掩膜切换模块（Mask Switching Module，MSM），结合通道注意力设计出 ARM，其具体结构如图 3 所示。

ARM 的核心功能是从预设分辨率集合 $[r_1, r_2, \cdots, r_K]$ 中为每个动物实例动态选择最优掩膜分辨率，从而降低实例分割的计算资源消耗。具体而言，该模块以区域 RoI 特征作为输入，首先通过全局平均池化（Global Average Pooling，GAP）提取全局上下文信息并降维，生成每个通道的紧凑特征表示，其数学表达为

$$G_c = \frac{\sum_{i=1}^{H}\sum_{j=1}^{W} F_c(i,j)}{H \times W} \tag{5}$$

式中：G_c 表示第 c 个通道对应的全局特征，$F_c(i,j)$ 为原始 RoI 特征图中位置 (i,j) 处的特征值，H 和 W 分别代表特征图的高度与宽度。

经 GAP 处理后的特征图 G 输入 SE 模块，该模块通过深度可分离卷积层提取不同通道中个体动物的分辨率特征并进行优化，输出二维特征图 $\boldsymbol{V}''_A$ 后执行展平操作。展平操作将三维张量（尺寸为高度 H × 宽度 W × 通道数 C）转化为一维特征向量，具体实现时按通道顺序逐行拼接直至最后一个通道，最终生成的一维特征向量 $\boldsymbol{F}$ 可表示为

$$\boldsymbol{F} = \mathrm{cat}(\mathrm{vec}(\boldsymbol{V}''_{A:,:,1}), \mathrm{vec}(\boldsymbol{V}''_{A:,:,2}), \cdots, \mathrm{vec}(\boldsymbol{V}''_{A:,:,C})) \tag{6}$$

式中：$\mathrm{vec}(\boldsymbol{V}''_{A:,:,k})$ 表示将第 k 个通道的特征图展平为向量的过程，cat 则指将所有通道展平后的向量进行整合与连接。

展平操作后，采用 Gumbel-Softmax 重参数化技术将特征向量 $\boldsymbol{F}$ 转化为离散分辨率选择。该方法通过随机过程实现可微变换，允许从离散概率分布中采样以实现端到端训练。具体而言，特征向量 $\boldsymbol{F}$ 经全连接层输出类别概率分布 $\boldsymbol{P} = [p_1, p_2, \cdots, p_K]$，该分布对应野生动物实例分割掩膜的分辨率选择概率。针对每个类别概率 p_K，引入从 Gumbel(0,1) 分布采样的独立同分布噪声项 g_K 以确保变换过程的可微性，其采样生成方式如下。

$$g_K = -\log(-\log(u_K)) \tag{7}$$

式中：$u_K \sim \mathrm{Uniform}(0,1)$。随后应用 Gumbel-Softmax 函数获得近似独热编码输出 $\boldsymbol{Y} = [y_1, y_2, \cdots, y_K]$，其中 y_k 定义为

$$y_k = \frac{\exp\left(\dfrac{\log(p_k) + g_k}{\tau}\right)}{\sum_{k'} \exp\left(\dfrac{\log(p_{k'}) + g_{k'}}{\tau}\right)} \tag{8}$$

式中：k 表示当前计算的类别索引，而 k' 则表示除当前类别外的其他所有可能类别；τ 表示温度参数，用于控制输出分布的尖锐程度。当 τ 趋近于 0 时，输出分布趋近于真实的独热编码（one-hot encoding），模型会明确选择单一分辨率（如“small”、“medium”和“large”），适用于特征丰富度高、形态稳定的野生动物实例，如成体东北虎、远东豹，因其轮廓清晰且纹理显著，模型可基于高置信度选择最优分辨率；而较大的 τ 值会产生更平滑的输出分布，生成的概率向量中各分辨率选项的权重差异减小，平滑分布可保留多分辨率探索的可能性，更适应特征模糊、形态多变的场景，如幼崽、部分遮挡个体或低对比度红外图像，避免因过度离散化选择而丢失潜在有效信息。

经 Gumbel-Softmax 运算生成的概率向量 $\boldsymbol{Y} = [y_1, y_2 \cdots, y_K]$ 描述各分辨率选项的选取概率。在输出向量 $\boldsymbol{Y}$ 中，被选中的分量对应索引 k，直接映射到预设分辨率类别（“small”、“medium”和“large”），其选择结果由以下公式确定。

$$\hat{k} = \arg\max_k(y_k) \tag{9}$$

式中：$\hat{k}$ 表示概率向量 $\boldsymbol{Y}$ 中最大值的索引，对应所选

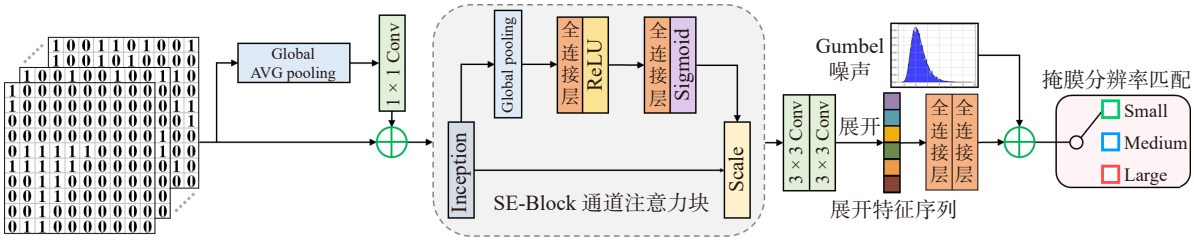


图 3 自适应分辨率模块（ARM）结构示意图

Fig. 3 Schematic diagram of Adaptive Resolution Module (ARM)

分辨率类别。

ARM 有效解决了野生动物天然伪装特性与复杂背景在识别与分割任务中带来的挑战。该模块通过动态调整每个实例的掩膜分辨率,不仅显著减少了不必要高分辨率处理导致的计算资源浪费,同时确保了分割过程中细节特征的精准捕捉。

1.4　引入条件随机场优化器实现掩膜边缘细化

在野生动物实例分割任务中,由掩膜头(mask head)根据 ARM 输出特征图生成的掩膜常存在边界模糊与背景噪声问题。该现象在动物轮廓边缘区域尤为显著,会严重降低分割精度并导致细节特征丢失,特别是当动物体色或纹理与背景高度相似时。

针对上述问题,本文提出的 PDCO 模块在 Wen 等[21]提出的 PatchDCT 框架基础上进行改进,重点处理包含边界信息的混合像素块,将计算资源集中分配至需精细处理的区域(如图 4 所示)。该模块将 $K\times K$ 尺寸的掩膜划分为 $m\times m$ 个像素块进行独立优化,同时引入 CRF 层,利用像素间空间关系进一步优化掩膜,从而增强前景与背景边界的对比度,并提升边缘清晰度。

1.4.1　分类器的设计

本分类器将图像块划分为 3 类:前景块、背景块及混合块。前景块仅包含目标像素;背景块不含任何目标像素;混合块则通常出现在目标边界区域,同时包含目标与背景像素。为精准处理混合块,本研究采用全卷积网络架构的分类器。对于前景块,使用离散余弦变换(Discrete Cosine Transform, DCT)进行特征表征,设定阈值 $\epsilon=0.05$ 以滤除细微噪声与图像变异,从而实现准确的前景块识别。在 DCT 变换中,直流分量(T_{DC})作为变换结果的首个元素,其计算公式为

$$T_{\mathrm{DC}}=\frac{1}{K^2}\sum_{x=0}^{K-1}\sum_{y=0}^{K-1}M(x,y) \tag{10}$$

式中:$M(x,y)$ 表示图像块内位置 (x,y) 处的像素值,$K\times K$ 为图像块尺寸。若块内所有像素 $M(x,y)$ 均接近理想前景,则 T_{DC} 趋近于 1。本研究引入阈值 ϵ 允许直流分量存在微小偏差,当满足 $|T_{\mathrm{DC}}-1|\leqslant\epsilon$ 时判定为前景块。对于背景块,通过卷积层计算表征背景常量的标量值 β,当满足 $|T_{\mathrm{DC}}-\beta|\leqslant\epsilon$ 时判定为背景块。其余高频 DCT 分量表征像素值变化的频率分布,在纯前景块中其值应趋近于 0,计算公式[21]为

$$D(u,v)=\frac{2}{K}\sum_{x=0}^{K-1}\sum_{y=0}^{K-1}M(x,y)\cos\frac{\pi(2x+1)u}{2K}\cos\frac{\pi(2y+1)v}{2K} \tag{11}$$

式中:u 和 v 分别表示离散余弦变换(DCT)中频域坐标的水平方向与垂直方向的频率索引;当 $u,v>0$ 时,根据余弦函数特性,当像素值 $M(x,y)$ 相等且对称分布时,高频分量 $D(u,v)$ 会相互抵消并趋近于 0。基于直流分量与高频分量 $D(u,v)$ 的阈值判断机制,未满足前景块与背景块判定条件的图像块将自动归类为混合块,该设计显著简化了目标边界区域的识别流程。

1.4.2　回归器的设计

该回归器专门处理分类器识别的混合块(包含实例边界处前景与背景像素的混合区域),这些区域对实现精确分割至关重要。为优化关键区域的掩膜边界,回归器针对性调整混合块的中高频 DCT 系数:首先计算每个混合块的 DCT 变换,根据尺寸和频率位置(特别是中高频段)校准系数以识别关键边界特征;通过抑制低幅值系数减少掩膜噪声,同时相位调

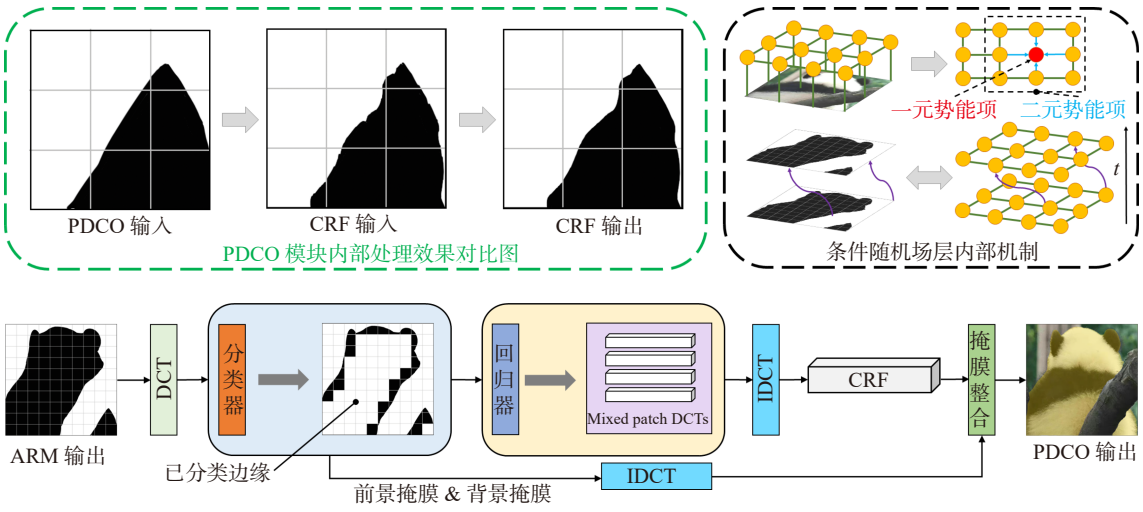


图 4　基于块的离散余弦变换的条件随机场优化器(PDCO)结构示意图

Fig. 4　Structural diagram of patch-based discrete cosine transform conditional random field optimizer (PDCO)

整确保 DCT 系数修改能增强边界清晰度，并与真实边界更好地对齐；最终通过离散余弦逆变换（inverse discrete cosine transform，IDCT）结合优化后的 DCT 系数重建掩膜。

1.4.3 条件随机场（CRF）掩膜优化

CRF 层[22]旨在进一步优化经回归器输出 IDCT 生成的预处理掩膜。为降低背景块大量像素对预测结果的计算影响，该层选择性聚焦于 IDCT 后前景块与混合块区域的像素，尤其强化混合块区域的计算。CRF 层建立全连接图模型，图中每个节点代表掩膜中的一个像素，节点间边权由像素颜色相似性与空间邻近度定义。该层计算混合块内每个像素时，会与已确定前景块的所有像素建立关联关系，从而实现边界区域的精准识别。CRF 能量函数由一元势能与二元势能构成：一元势能直接取自 IDCT 输出，表示各像素属于前景或背景的概率；二元势能强调像素相似性，通过指数函数削弱颜色与位置差异显著像素间的标签一致性，其优化函数[22]定义为

$$E(x)=\sum_i \psi_u(x_i)+\sum_{j<i}\psi_p(x_i,x_j) \tag{12}$$

$$\psi_p(x_i,x_j)=\mu(x_i,x_j)\cdot\omega_1\exp\left(-\frac{|p_i-p_j|^2}{2\sigma_\alpha^2}-\frac{|I_i-I_j|^2}{2\sigma_\beta^2}\right)+\omega_2\exp\left(-\frac{|p_i-\mathrm{p}_j|^2}{2\sigma_\gamma^2}\right) \tag{13}$$

式中：$\psi_u(x_i)$表示从预处理掩膜获取的一元势能；$\psi_p(x_i,x_j)$为基于像素i与j的位置(p_i,p_j)和颜色(I_i,I_j)计算的二元势能；参数ω_1调节动物纹理与背景的色彩差异敏感性，ω_2约束相邻像素在遮挡边界的连续性；σ_α、σ_β和σ_γ这 3 个参数协同控制不同野生动物的边缘响应特性。本方案采用平均场近似算法优化 CRF，通过迭代更新每个像素的标签概率直至收敛，实现整体能量函数的最小化。在每次迭代中，各像素的标签分布依据其邻域像素的当前估计值进行更新，其数学表示[22]为

$$q_i(x_i)\propto\exp\left(-\psi_u(x_i)-\sum_{j\neq i}\psi_p(x_i,x_j)\sum_{x_j}q_j(x_j)\right) \tag{14}$$

式中：$\psi_u(x_i)$表示像素i处于状态x_i的概率，$q_i(x_i)$表示像素i的标签（前景/背景）的近似边际概率分布，通过迭代考虑其邻域像素j的当前估计分布$q_j(x_j)$来更新，该更新过程循环执行直至所有像素的标签分布收敛，最终实现分割边缘的精准优化。经过 CRF 处理后，通过合并操作（merge）将 CRF 优化后的边缘与经 IDCT 处理的前景掩膜融合，最终生成完整且精度更高的分割掩膜。

2 实验验证

2.1 数据集构建

2.1.1 数据采集与标注

为支持野生动物监测研究，本研究构建了一个野生动物实例分割数据集。数据集采集自天台山、唐家河及东北虎豹国家公园，包含大熊猫、赛加羚羊、金丝猴、东北虎和远东豹 5 类野生动物的图像。这些图像来源于自然光照条件下（包括正午、黄昏及夜间环境下拍摄）的地面陷阱相机拍摄，同时包含可见光与红外两种成像模式。2015—2023 年间的监测影像中，共筛选出 4 231 张原始图像，其中大熊猫 317 张、赛加羚羊 1 010 张、金丝猴 1 032 张、东北虎 963 张及远东豹 909 张，所有图像均已完成实例分割标注。图像分辨率包含 3 种规格（2 048 × 1 536、4 000 × 3 000 和 1 280 × 720 像素）。数据集示例及标注效果如图 5 所示。

2.1.2 数据集划分与预处理

本文提出的实例分割方法在自建的 5 类物种数据集与文献 [23] 发布的 LoTE-Animal 数据集上进行了联合训练、验证与测试。特别需要说明的是，LoTE-Animal 数据集不仅包含本研究的 5 个目标物种（大熊猫、赛加羚羊、金丝猴、东北虎和远东豹），还涵盖其他 6 种濒危物种的长期连续行为数据，共为实例分割任务提供 28 000 张高分辨率图像。该数据集通过系统记录自然保护区内动物长期连续的行为特征，为包括目标物种在内的野生动物精确行为学与形态学分析提供重要数据支撑。

为增强模型鲁棒性，本研究对两个训练集实施了系统的预处理策略。该策略涵盖图像旋转、饱和度与曝光度调整、模糊处理、噪声添加及马赛克效应模拟（图 6）。其中，马赛克效应系通过随机叠加实心黑色块，模拟野外相机镜头被泥土或毛发遮挡的真实场景。此外，数据集中有超过 5 000 张图像经局部



图 5 野生动物数据集标注效果图

Fig. 5 Annotation results of the wildlife dataset

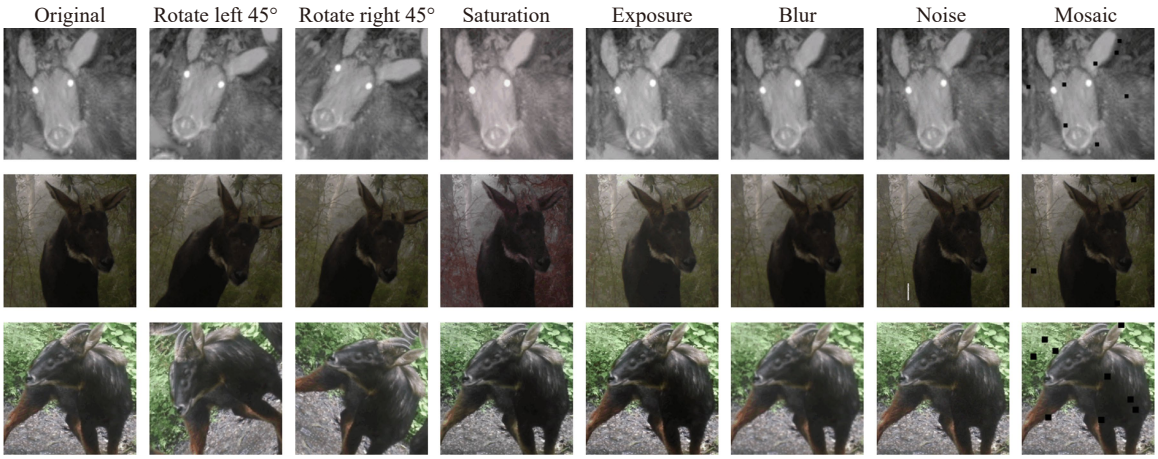


图 6　数据集数据增强效果示意图

Fig. 6　Illustration of data augmentation effects on the dataset

遮挡与掩膜标注调整(图 7)，专门用于模拟动物在自然环境中隐藏与伪装的行为特征。

2.2　主流方法对比基准

为全面评估所提方法在野生动物实例分割任务中的性能，选取 Mask Transfiner、QueryInst 等通用实例分割领域的主流领先方法作为基准，从分割精度、计算效率及资源消耗等多维度展开对比。所有实验均在统一硬件平台与软件环境下开展，确保结果可比。为验证模型在不同特征提取能力下的鲁棒性，分别在 ResNet-101-FPN、ResNeXt-101-FPN 及 Swin-B 三种表征能力各异的骨干网络上进行测试。

为进一步评估本方法相对于现有主流模型的整体性能，以 Mask R-CNN 为基线模型，超参数和损失函数设置均与 MMDetection 实现保持一致(学习率 0.001，批量大小 8)。评估指标涵盖平均精度均值(mean Average Precision，mAP)、平均实例交并比(mI^{2}oU)、浮点运算量(Floating Point Operations，FLOPs)及推理速度(Frames Per Second，FPS，帧/s)，并与近期主流实例分割模型进行对比。所有实验在 NVIDIA RTX 3090 GPU(PyTorch 框架)上完成。为综合考察收敛性与泛化能力，采用 150 轮与 300 轮双周期训练策略：其中 150 轮设定模拟濒危物种图像稀缺场景，与领域基准保持一致以保障对比公平性。骨干网络选用 ResNet-50-FPN、ResNet-101-FPN 及 ResNeXt-101-FPN。

鉴于野生动物监测常部署于算力受限的边缘设备或电池供电的野外相机，本研究重点评估模型的计算复杂度与运行效率，通过 FLOPs 量化理论计算量并与基准方法对比。在精度评估方面，引入平均实例交并比 mI^{2}oU 以精细衡量预测掩膜与真实标注的像素级吻合度，尤其关注边缘区域的分割准确性。此外，通过对比实验验证本方法自适应掩膜分辨率策略相较于多种固定分辨率方案在平衡精度与资源消耗方面的优势。

2.3　模块改进消融实验

为定量评估各核心模块的贡献度，采用控制变量法设计消融实验，系统探究 SE-DAM、ARM 及 PDCO 的有效性与协同效应。评估涵盖分割精度与计算效率双维度：精度指标选用实例分割通用的 mAP 及对掩膜边缘质量敏感的 mI^{2}oU；效率指标则通过 FLOPs 和 FPS 衡量。所有实验均在统一硬件平台(NVIDIA RTX 3090 GPU)与软件环境(PyTorch)下执行，以排除环境干扰。

3　结果与分析

3.1　与主流方法的定量对比及性能优势

在ResNet-101-FPN、ResNeXt-101-FPN 和Swin-B 骨干网络下，本方法分别取得 53.7%、54.2% 和 54.8%

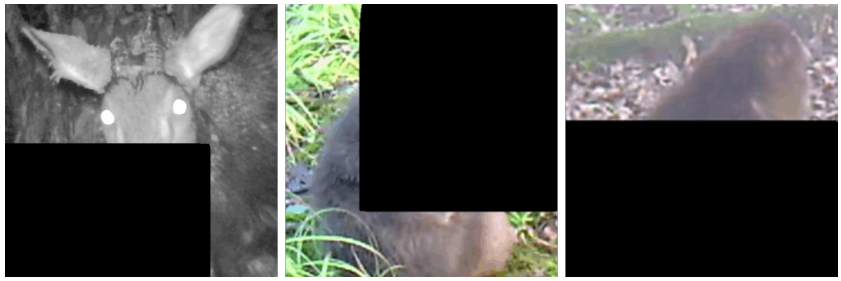


图 7　数据集局部遮挡数据调整效果图

Fig. 7　Illustration of data adjustment for partial occlusion simulation

的 mAP 值(表 1),均超越现有主流方法。需特别说明的是,虽然 Mask Transfiner 和 QueryInst 的分割精度与本方法相当,Mask Transfiner(采用 Swin-B 主干网络)和 QueryInst(采用 ResNet-101-FPN 主干网络)分别取得了 50.8% 和 52.4% 的 mAP 得分,表现出较高的分割精度,但其性能提升以显著的内存和资源消耗为代价,FPS 分别仅为 6.5 和 7.3 帧/s,明显低于本方法在相应主干网络下取得的 10.9 帧/s(Swin-B)和 12.4 帧/s(ResNet-101-FPN),这在本研究针对野生动物监测的实地应用中不具备可行性。此外,在小目标及遮挡动物分割任务中,本方法以 4.2 个百分点的优势在 ResNet-101-FPN 主干网络上超越其他方法的最高 mAP 值(如表 1 所示,HTC[24]在该主干网络上 mAP 值为 49.5%,本方法 mAP 得分为 53.7%)。这得益于通过通道注意力机制分析跨通道特征信息,有效校正空间信息损失并区分前景与复杂背景,从而实现对小型遮挡目标的高效分割。

尽管动态掩膜在整体精度上略低于 112 × 112 像素固定掩膜,但其在计算效率(浮点运算量为 1.31 GFLOPs,显著低于 BlendMask[25]、DCT-Mask[26]和 RefineMask[27]的值;FPS 达到 12.4 帧/s,较 BlendMask、DCT-Mask 和 RefineMask 的平均值(8.5 帧/s)提升了 3.9 帧/s)和资源适应性方面具有显著优势(表 2),对部署在保护区边缘计算节点的实际监测系统至关重要。特别地,对于面积小于 32 × 32 像素的小型目标,即使采用高分辨率策略,其分割精度提升也较为有限;而中大型目标在动态掩膜下仍能保持 80% 以上的 mI^{2}oU 精度(表 3 中本方法 mI^{2}oU$_m$ 和 mI^{2}oU$_L$ 得分分别为 82.37% 和 86.46%)。动态掩膜确保了对电池供电的红外相机等低算力设备的兼容性,同时通过后续 PDCO 模块补偿了分辨率损失。实验表明,虽然 BMask R-CNN[28]、MEInst[29]和 QueryInst 具有更高的 FPS 与更低的 FLOPs,显示其处理效率较高,但其分割结果欠佳,mAP 等指标较低;而 PointRend[30]等模型虽获得与本方法相当的分割效果,但计算资源消耗显著增加,处理效率大幅降低(浮点运算量高达 3.64 GFLOPs,FPS 最低仅 8.0 帧/s)。本研究还采用 mI^{2}oU 指标评估了部分先进方法,筛选 mI^{2}oU 值较高的方案进行对比(如表 3 所示),结果表明本方法整体性能优于其他方案。

3.2 相较 Mask R-CNN 基线的提升

高精度与高效率协同优化的框架设计显著提升了复杂场景下的分割能力。本研究提出的融合空间—通道自适应机制与分层优化策略的框架,有效解决了 Mask R-CNN 基线中由下采样/卷积导致的空间信息损失和异构区域同质化处理两大核心瓶颈。相

表 1 不同实例分割方法评估结果对比(含 mAP/FLOPs/FPS 指标)

Tab. 1 Comparison of evaluation results across different instance segmentation models, covering metrics mAP, FLOPs, and FPS

方法	骨干网络	mAP/%	mAP$_{50}$/%	mAP$_{75}$/%	mAP$_s$/%	mAP$_m$/%	mAP$_L$/%	mAP$_B$/%	FPS/(帧·s^{-1})
Mask R-CNN	ResNet-101-FPN	42.5	61.7	43.3	21.7	44.0	46.3	24.2	13.9
BMask R-CNN	ResNet-101-FPN	43.3	62.4	45.7	21.3	45.9	49.9	24.7	13.1
TensorMask	ResNet-101-FPN	42.9	59.5	44.6	19.6	43.9	47.6	24.5	11.9
MEInst	ResNet-101-FPN	38.4	58.8	41.9	22.8	41.6	43.2	21.9	10.3
Mask Scoring	ResNet-101-FPN	44.2	64.1	47.1	20.9	46.0	51.4	25.2	12.2
BlendMask	ResNet-101-FPN	42.1	61.6	44.0	18.4	44.2	48.7	24.0	10.7
CondInst	ResNet-101-FPN	47.8	62.2	51.8	24.7	50.9	53.3	27.3	9.8
PointRend	ResNeXt-101-FPN	51.2	68.3	56.5	28.5	50.5	49.2	27.2	8.2
HTC	ResNet-101-FPN	49.5	67.2	52.6	25.4	51.4	54.6	26.4	7.1
DCT-Mask	ResNet-101-FPN	46.4	63.9	51.1	22.3	50.6	53.5	26.4	12.9
SOLOv2	ResNet-101-FPN	40.1	58.9	43.0	19.4	43.4	49.8	22.8	11.6
RefineMask	ResNet-101-FPN	46.6	62.9	49.3	23.7	48.5	50.2	26.5	9.6
Mask Transfiner	Swin-B	50.8	68.3	54.8	28.1	50.9	51.9	27.4	6.5
QueryInst	ResNet-101-FPN	52.4	70.3	53.7	28.5	51.3	54.4	26.8	7.3
本方法	ResNet-101-FPN	53.7	70.9	59.7	33.5	51.6	57.7	27.5	12.4
	ResNeXt-101-FPN	54.2	71.5	61.3	34.7	52.8	59.2	27.6	12.1
	Swin-B	54.8	72.1	61.6	34.8	53.3	60.5	31.1	10.9

注:mAP为平均精度均值,mAP$_{50}$、mAP$_{75}$分别为交并比(IoU)阈值取 50%、75% 时的平均精度,mAP$_s$、mAP$_m$、mAP$_L$分别为小、中、大型目标的平均精度,mAP$_B$为背景干扰下的平均精度,FPS为每秒处理的图像帧数。

表 2　不同实例分割方法掩膜分辨率对比（mAP 指标）

Tab. 2　Comparison of evaluation results across different instance segmentation methods using default mask resolutions and dynamic resolution masks

方法	骨干网络	掩膜分辨率	mAP/%	mAP_{50}/%	mAP_B/%	浮点运算量/GFLOPs	FPS/(帧·s^{-1})
Mask R-CNN	ResNet-101-FPN	56 × 56像素	42.8	62.8	23.5	1.01	13.2
BlendMask	ResNet-101-FPN	112 × 112像素	42.6	60.3	23.9	2.03	8.2
PointRend	ResNet-101-FPN	224 × 224像素	52.4	68.7	27.4	3.64	8.0
DCT-Mask	ResNet-101-FPN	112 × 112像素	45.1	61.3	25.8	1.89	8.9
MEInst	ResNet-101-FPN	56 × 56像素	39.4	59.5	22.7	1.44	12.1
RefineMask	ResNet-101-FPN	112 × 112像素	45.9	61.6	25.0	1.97	8.5
QueryInst	ResNet-101-FPN	28 × 28像素	40.6	60.7	21.9	0.73	13.1
CondInst	ResNet-101-FPN	56 × 56像素	47.2	61.1	24.1	1.26	9.9
本方法	ResNet-101-FPN	动态调整	53.7	70.9	27.5	1.31	12.4

表 3　实例分割方法评估结果对比（mI^2oU 指标）

Tab. 3　Comparison of evaluation results across different instance segmentation models, covering metric mI^2oU　%

方法	骨干网络	mI^2oU	mI^2oU_s	mI^2oU_m	mI^2oU_L
Mask R-CNN	ResNet-101-FPN	79.01	47.49	73.18	81.31
Mask Scoring	ResNet-101-FPN	83.34	53.10	74.31	83.27
MEInst	ResNet-101-FPN	80.75	51.01	73.59	82.03
DCT-Mask	ResNet-101-FPN	84.01	53.95	74.48	84.13
HTC	ResNet-101-FPN	80.68	51.75	73.02	82.27
Mask Transfiner	ResNet-101-FPN	87.15	57.29	75.63	87.29
QueryInst	ResNet-101-FPN	85.03	52.53	75.52	86.07
TensorMask	ResNet-101-FPN	86.64	56.67	74.77	86.57
PointRend	ResNet-101-FPN	85.90	54.65	75.61	85.04
本方法	ResNet-101-FPN	86.71	75.61	82.37	86.46

注:mI^2oU为平均交并比均值;mI^2oU_s、mI^2oU_m、mI^2oU_L分别为小、中、大型野生动物目标的平均交并比。

较于主流实例分割流程，本方法通过 SE-DAM 动态校正空间错位，并结合 ARM 优化计算资源分配。该框架在多种骨干网络下均显著提升了分割性能(如 Swin-B 骨干下 mAP 达 54.8%)，同时在计算效率(如 ARM 模块下 FPS 稳定在 12.4 帧/s)上展现出显著优势(表 1)。通过协同优化，满足了生态监测应用对野外复杂环境下(如遮挡、伪装)实现高效、精准个体实例分割的迫切需求。本模型整体性能优于 Mask R-CNN:在相同骨干网络下，150 和 300 训练周期时 mAP 分别平均提升 8.4 个百分点和 9.1 个百分点;在 mAP_{50} 指标上分别平均提升 6.4 个百分点和 6.5 个百分点;在 mAP_{75} 指标上则分别取得 13.3 个百分点和 14.1 个百分点的平均提升(表 4)。mAP_{50} 作为基础检测精度的体现，主要受益于 SE-DAM 模块在特征对齐方面的改进，其 6.5 个百分点的提升反映了模型对目标主体识别能力的增强;而 mAP_{75} 对边界精度要求更高，其显著提升(14.1 个百分点)则得益于 PDCO 模块在边缘优化方面的专门优化。这说明本方法在保持高召回率的同时，显著提升了分割边界的精细化程度。本研究的 SE-DAM 模块通过通道注意力机制与可变形卷积校正特征图偏移和空间错位，其中通道注意力通过重标定特征通道权重，增强了动物轮廓等高频信息的表征能力，同时 PDCO 模块采用了频率域分析与空间域优化相结合的策略，对混合块分类优化实现物体边缘的高精度分割。这些组件的协同作用使本方法在更高 IoU 阈值下具有更强的边界识别能力，显著提升了分割结果的精细度与准确性。

3.3　核心模块贡献度

3.3.1　模块添加消融实验对比

本研究提出的 3 个模块在高精度实例分割、遮挡目标分割及边缘优化方面均表现出有效性能。当单独启用 SE-DAM 模块时，模型达到 46.1% mAP 和 81.04% mI^2oU，1.15 GFLOPs，12.8 帧/s;单独启用

表 4 与基准模型（Mask R-CNN）对比实验结果

Tab. 4 Comparison of evaluation results with the baseline (Mask R-CNN) %

方法	骨干网络	训练轮数	mAP	mAP_{50}	mAP_{75}	mAP_s	mAP_m	mAP_L	mAP_B
Mask R-CNN	ResNet-50-FPN	150	36.5	52.9	37.0	16.1	33.6	34.2	20.4
本方法	ResNet-50-FPN		45.7	59.4	49.8	24.4	36.5	40.5	23.3
Mask R-CNN	ResNet-50-FPN	300	39.9	59.8	41.8	18.9	35.9	37.3	22.3
本方法	ResNet-50-FPN		48.9	63.5	53.3	26.2	40.3	44.7	24.9
Mask R-CNN	ResNet-101-FPN	150	38.6	55.9	39.1	20.5	38.9	40.8	22.6
本方法	ResNet-50-FPN		47.3	62.4	52.5	26.2	41.9	46.5	25.1
Mask R-CNN	ResNet-101-FPN	300	42.5	61.7	43.3	21.7	44.0	46.3	24.2
本方法	ResNet-50-FPN		53.7	70.9	59.7	33.5	51.6	57.7	27.5
Mask R-CNN	ResNeXt-101-FPN	150	41.2	61.8	43.2	20.9	39.7	43.6	23.6
本方法	ResNet-50-FPN		48.5	67.9	57.0	28.0	43.1	49.8	26.1
Mask R-CNN	ResNeXt-101-FPN	300	42.4	65.1	45.5	22.4	42.5	44.7	23.9
本方法	ResNet-50-FPN		54.2	71.5	61.3	34.7	52.8	59.2	27.6

ARM 模块时，性能为 45.6% mAP 和 80.39% mI^2oU，1.16 GFLOPs，12.9 帧/s；单独启用 PDCO 模块不含 CRF 时，性能为 47.3% mAP 和 81.18% mI^2oU，1.20 GFLOPs，12.5 帧/s（表 5）。这是因为 SE-DAM 模块通过通道注意力机制强化对动物关键部位（如边缘、纹理显著区域）的特征响应，同时利用可变形卷积自适应调整采样位置，有效补偿了因动物运动或遮挡导致的特征空间错位；ARM 模块则通过 Gumbel-Softmax 决策机制，实现了从像素级特征到实例级分辨率分配的映射，使模型能够根据实例的语义复杂度和空间尺度动态分配计算资源；PDCO 模块通过 DCT 域的频率分析，将图像块依据频谱能量分布划分为前景、背景和混合区域，从而实现对边界区域的精准聚焦。当模块协同工作时，性能保持持续提升，表明各模块间耦合度低且无相互干扰。

SE-DAM 与 ARM 组合时，mAP 提升至 49.5%，mI^2oU 达 84.29%，1.27 GFLOPs，12.4 帧/s；SE-DAM 与 PDCO（不含 CRF）组合时，mAP 进一步增至 49.7%，mI^2oU 为 84.26%，1.30 GFLOPs，11.9 帧/s；而当 3 个模块（SE-DAM、ARM 和 PDCO 不含 CRF）共同启用时，mAP 达到 52.2%，mI^2oU 为 86.91%，1.37 GFLOPs，11.4 帧/s，显示出叠加效应。SE-DAM 作用于特征提取阶段，优化的是特征的空间对齐与通道权重；ARM 在特征编码后介入，解决的是分辨率分配策略问题；PDCO 则专注于解码后的掩膜优化层面。三者分别对应于分割流程中的前、中、后 3 个不同阶段，形成了从全局特征优化到局部边缘细化的完整链路。特别地，当 3 个模块均未启用时，模型性能与 Mask R-CNN 相当。具体而言，在 PDCO 模块中省略 CRF 会使分割性能略有下降（与未省略

表 5 消融实验评估结果

Tab. 5 Results of the ablation study evaluation

SE-DAM	ARM	PDCO（−CRF）	PDCO（＋CRF）	mAP/%	mI^2oU/%	浮点运算量/GFLOPs	FPS/（帧·s^{-1}）
√				46.1	81.04	1.15	12.8
	√			45.6	80.39	1.16	12.9
		√		47.3	81.18	1.20	12.5
			√	46.9	81.10	1.25	12.3
√	√			49.5	84.29	1.27	12.4
√		√		49.7	84.26	1.30	11.9
√			√	49.2	83.37	1.35	12.1
	√	√		48.8	83.16	1.31	12.2
	√		√	48.1	83.18	1.36	12.5
√	√	√		52.2	86.91	1.37	11.4
√	√		√	54.8	87.58	1.42	10.9

CRF 的 PDCO 模块相比，mAP 降低了 2.6 个百分点，mI^2oU 降低了 0.67 个百分点)，但能减少资源消耗(浮点运算量降低了 0.05 GFLOPs)，并提升处理速度(帧率增加了 0.5 帧/s)。这证实了 CRF 在提升精度的同时带来计算开销。在资源受限或特殊条件下，省略用于掩膜边界优化的 CRF 可提升处理效率。

针对野生动物运动、遮挡导致的空间错位问题，SE-DAM 模块融合了通道注意力与可变形卷积机制。如表 1 所示，该模块在 ResNet-50/101-FPN 骨干下实现了 45.7% ~ 51.4% 的 mAP，显著高于 Mask R-CNN 基线模型的 36.5% ~ 42.4%。通道注意力机制强化了对动物姿态变化和局部暴露区域的关键特征响应，而可变形卷积则自适应调整感受野，精准补偿因植被遮挡或个体重叠造成的非线性偏移。相较于 CondInst 等动态卷积方案仅优化掩膜头特征，SE-DAM 在特征提取阶段即进行空间自适应校准，从根源上减少了 RoI Align 过程中的信息损失，为后续高精度定位奠定了坚实基础。

3.3.2　ARM 模块以最小计算开销实现了关键细节的精准捕捉

为应对野生动物天然伪装与背景高相似性挑战，ARM 模块创新性地为每个实例动态分配掩膜分辨率(范围 28 × 28 像素至 112 × 112 像素)。如表 6 和表 7 所示，在 ResNet-101-FPN 骨干下，ARM 仅需

表 6　动态与固定分辨率掩膜性能对比（含 mAP/FLOPs/FPS 指标）

Tab. 6　Comparison of evaluation results between dynamic resolution masks and fixed resolution masks, covering metrics mAP, FLOPs, and FPS

掩膜尺寸	骨干网络	mAP/%	mAP_{50}/%	mAP_{75}/%	mAP_s/%	mAP_m/%	mAP_L/%	mAP_B/%	浮点运算量/GFLOPs	FPS/(帧·s^{-1})
28 × 28像素	ResNet-101-FPN	47.0	61.1	51.6	27.6	46.9	52.5	26.7	0.55	16.6
	ResNeXt-101-FPN	47.1	62.5	52.4	28.5	48.4	54.2	26.7	0.58	16.2
	Swin-B	48.1	62.5	53.4	28.9	49.1	54.9	27.2	0.64	14.5
56 × 56像素	ResNet-101-FPN	51.3	66.6	55.1	30.7	49.1	55.3	33.3	1.12	12.9
	ResNeXt-101-FPN	52.2	67.8	56.0	30.8	49.2	54.6	33.9	1.19	12.4
	Swin-B	53.0	68.9	55.9	31.4	50.2	55.7	34.4	1.27	10.3
112 × 112像素	ResNet-101-FPN	56.2	71.1	59.8	36.1	52.6	57.5	37.9	1.65	9.7
	ResNeXt-101-FPN	57.4	72.2	59.8	37.2	54.3	59.4	38.3	1.78	8.8
	Swin-B	57.7	72.4	61.9	37.4	54.5	59.6	38.1	1.86	7.1
动态掩膜	ResNet-101-FPN	53.7	70.9	59.7	33.5	51.6	57.7	27.5	1.31	12.4
	ResNeXt-101-FPN	54.2	71.5	61.3	34.7	52.8	59.2	27.6	1.35	12.1
	Swin-B	54.8	72.1	61.6	34.8	53.3	60.5	31.1	1.42	10.9

表 7　动态与固定分辨率掩膜的 mI^2oU 指标对比

Tab. 7　Comparison of evaluation results between dynamic resolution masks and fixed resolution masks, covering metric mI^2oU

%

掩膜尺寸	骨干网络	mI^2oU	mI^2oU_s	mI^2oU_m	mI^2oU_L
28 × 28像素	ResNet-101-FPN	81.07	73.34	76.37	80.39
	ResNeXt-101-FPN	81.50	73.36	76.51	80.86
	Swin-B	82.18	73.70	76.97	80.77
56 × 56像素	ResNet-101-FPN	85.65	74.47	81.25	86.07
	ResNeXt-101-FPN	86.02	75.83	82.04	86.04
	Swin-B	86.97	75.89	82.19	86.73
112 × 112像素	ResNet-101-FPN	89.83	76.52	85.82	91.09
	ResNeXt-101-FPN	89.97	76.90	86.09	91.59
	Swin-B	90.37	76.96	86.31	91.85
动态掩膜	ResNet-101-FPN	86.71	75.61	82.37	86.46
	ResNeXt-101-FPN	87.14	75.99	82.58	86.93
	Swin-B	87.58	76.04	82.79	87.21

1.31 GFLOPs(显著低于 ResNet-101-FPN 主干网络下 112 × 112 像素固定分辨率的 1.65 GFLOPs)，同时维持了 86.71% 的高 mI^{2}oU 精度。核心突破在于：通过通道注意力驱动的 Gumbel-Softmax 决策，实现了计算资源的按需分配——对大型、特征丰富的个体(如东北虎)采用高分辨率保留细节，而对小型或背景单一目标则降低分辨率以节省资源。该动态策略有效克服了固定分辨率掩膜(如 PointRend 的 224 × 224 方案计算开销巨大)在资源受限的边缘设备部署时的局限性，为保护区长期自动化监测提供了可行技术路径。图 8 和图 9 的对比清晰展示了其在保持边界精度的同时显著降低计算负担的优势。

在固定与动态掩膜分辨率性能对比方面，表 6 显示：当分辨率从 28 × 28 像素提升至 112 × 112 像素时，固定分辨率掩膜分配策略使分割精度显著提高(mAP 从 47.0% 增至 56.2%)，但伴随计算资源消耗增加与模型处理效率下降(浮点运算量从 0.55 GFLOPs 升至 1.65 GFLOPs，FPS 从 16.6 帧/s 降至 9.7 帧/s)。采用基于 Gumbel-Softmax 函数动态分配分辨率的 ARM，在分割精度与计算效率间实现更好平衡：相较于中等尺寸掩膜，ARM 模块在资源消耗小幅增加的情况下确保更高分割精度，例如在 ResNet-101-FPN 骨干网络下，相较于 56 × 56 像素下，mAP 提升至 53.7% 的同时仅需增加 0.19 GFLOPs。

对于边界特征较少且精度较低的小目标，即使提升掩膜分辨率，mI^{2}oU 分数改善也微乎其微；但随着目标尺寸及特征密度的增加，mI^{2}oU 分数呈现显著提升。中分辨率掩膜的性能增益最为突出，而动态分辨率掩膜的整体性能略优于固定中分辨率方案。相较于 112 × 112 像素固定分辨率策略，虽然整体精度略有下降，但计算资源消耗显著降低(表 7)。为验证 ARM 模块的分割效果，图 8 展示了不同分辨率下掩膜边缘的细节对比，图 9 则呈现了具体的动态掩膜分割结果。

3.3.3 PDCO 模块驱动分层优化显著提升了边缘与遮挡区域的分割质量

针对均质化处理忽视区域难度差异的问题，PDCO 模块采用了基于 DCT 的块分类与 CRF 引导的分层优化策略。在 ResNet-101-FPN 骨干下，该模块使边界精度指标(mAP_B)达到 27.5%，显著优于 Mask R-CNN 的 24.2%，如表 1 所示。其核心在于：首先利用 DCT 频率特性将像素块分类为前景、背景及混合块，聚焦资源优化包含边界的混合块；继而通过 CRF 建模像素间上下文关系，以高置信度区域(易分割像素)引导低置信度区域(难分割边缘与遮挡部位)的精细化分割。

3.4 完整实例分割结果系统验证

组合图像首列在枝叶遮挡场景下，BMask R-CNN、CondInst 和 DCT-Mask 等方法表现欠佳，常将背景植被误判为动物躯体部分。这类模型因前景背

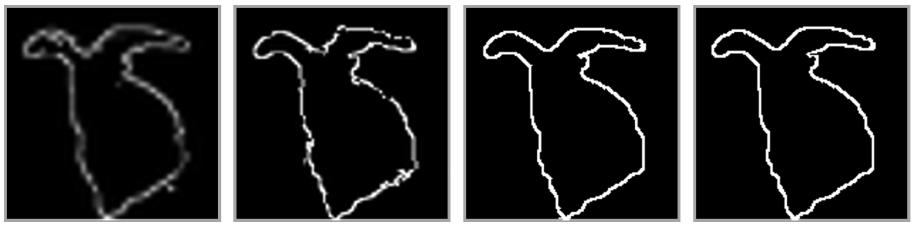


图 8 多分辨率掩膜边界对比图

Fig. 8 Comparison of mask boundaries at different resolutions

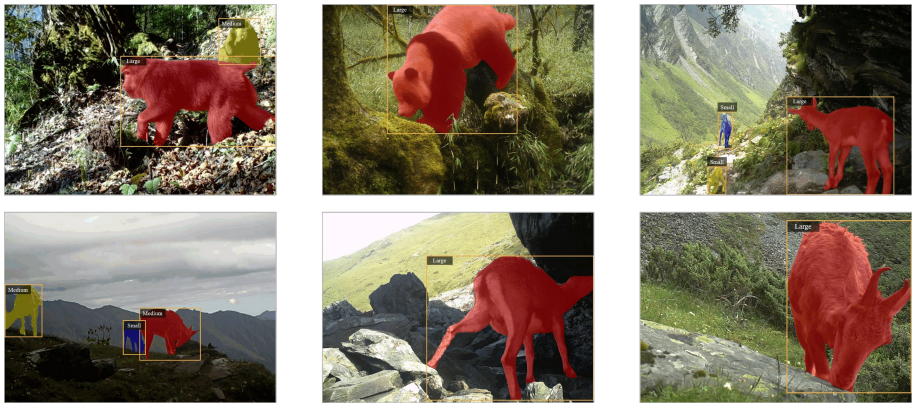


图 9 自适应分辨率掩膜效果可视化

Fig. 9 Visualization of adaptive-resolution mask effects

景视觉特征相似性而难以有效区分目标与干扰物。相较之下，本方法通过集成先进的特征提取与细节增强技术，可准确区分树枝与动物躯体，避免此类误判并获得更精确的分割结果。这一优势不仅体现在分割精度较基线方法 Mask R-CNN 的提升(见表 1，在 ResNet-101-FPN 主干网络下，本方法 mAP 值为 53.7%，较 Mask R-CNN 的 42.5% 提升 11.2 个百分点)，更保证了分割结果的一致性与可靠性。

此外，在动物躯体相互遮挡场景下(图 10 第 2 列)，现有方法普遍存在识别困难。TensorMask 常将不同个体的躯干部分错误关联，导致归属判定失效；Mask R-CNN 和 CondInst 等模型亦难以准确判定遮挡区域的躯干归属，产生错误分割。本方法通过通道注意力机制聚合全局上下文信息，有效区分重叠动物实体，增强了模型识别遮挡区域内边界的能力，显著提高了分割精度。

针对野生动物隐匿时躯干局部暴露且受植被干扰的场景(图 10 第 3 列)，本方法展现出显著优势。现有方法因缺乏处理局部遮挡与复杂纹理交互的先进机制，常将背景植被误判为动物躯体或使动物特征融入背景。通过集成通道注意力机制与可变形卷积机制，增强特征图中边缘细节的表征能力，使模型对复杂背景下局部暴露动物躯干的分割精度提升明显(相较基准方法)，尤其在枝叶纹理与动物皮毛的区分任务中与现有最优方法分割结果对比强烈。

在自然环境下小尺度目标的检测场景中(图 10 第 4 列)，现有方法因目标尺寸过小常导致特征提取失效，产生错误分割，且无法正确分离微小目标。相比之下，本研究的方法通过动态调整掩膜分辨率，并通过 PDCO 模块进一步处理细节，确保即使是小尺寸实体也能准确描绘其形态特征，从而使本研究的方法能够有效区分复杂背景中的小实体，并对其进行精确分割。

本方法在红外图像分割任务中同样表现卓越

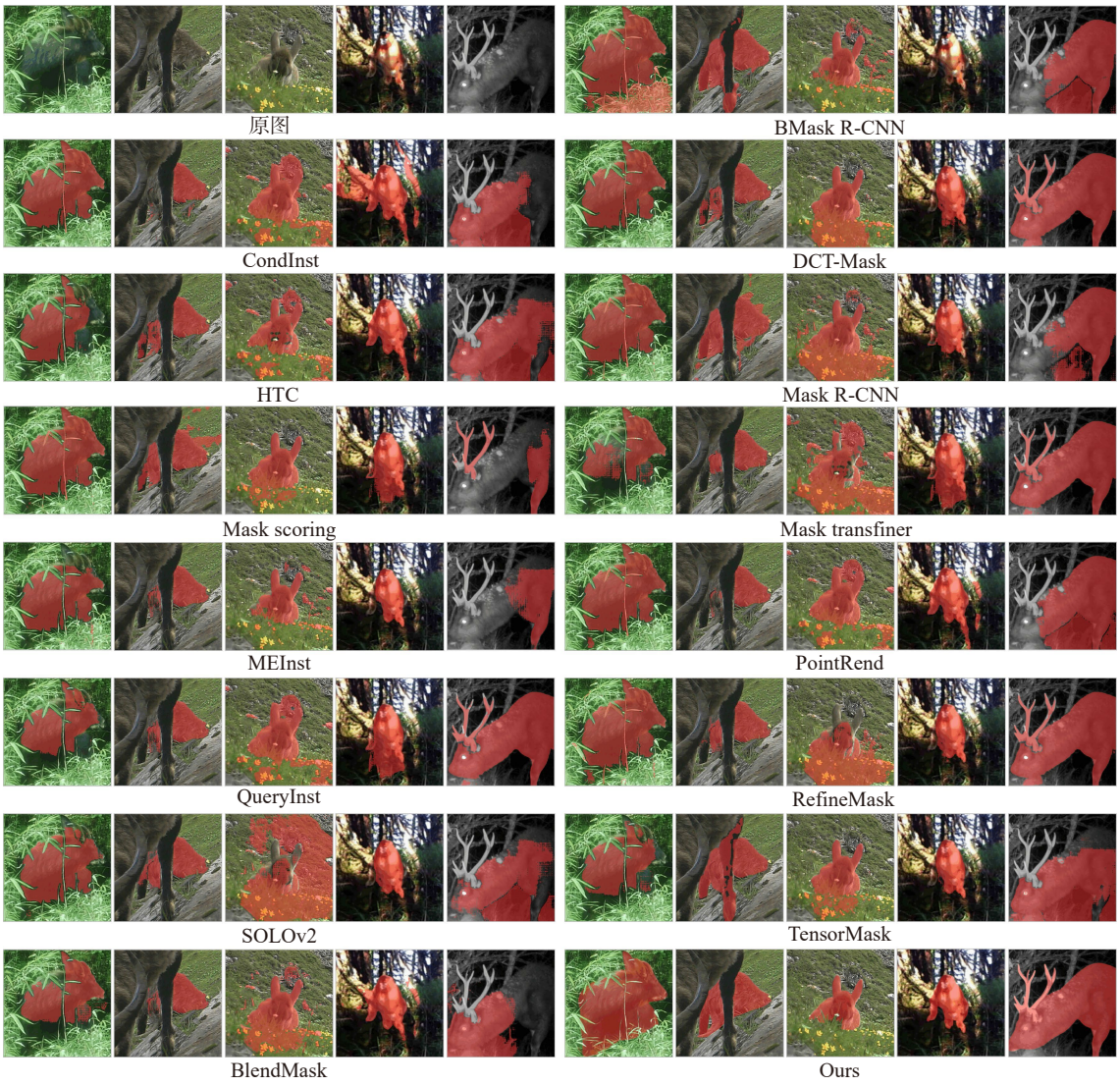


图 10　实例分割方法分割效果对比图

Fig. 10　Comparison of segmentation results by different instance segmentation methods

(如图 10 第 5 列所示)。对比方法如 Mask Scoring[31] 和 MEInst 在此类场景常表现不佳,尤其在红外图像高对比度低照度条件下,面临高背景噪声与环境复杂性时难以有效识别动物目标。本方法通过 SE-DAM 模块解决红外图像空间错位问题,并结合 PDCO 模块增强边界特征提取,显著提升复杂噪声场景的分割精度。该策略对密集个体遮挡(扭角羚群重叠)和小目标分割尤为有效(如图 10 所示),在消融实验中验证了其关键作用(如表 5 所示):结合 CRF 时,模型 mAP 提升 2.6 个百分点(见表 5)。相较于 Mask Transfiner 依赖计算密集的四叉树 Transformer 优化局部区域,PDCO 在实现相当甚至更优边缘精度的同时(见表 1,在 Swin-B 下,本方法 mAP 为 54.8%,Mask Transfiner mAP 为 50.8%),保持了更高的处理效率(见表 1,在 Swin-B 下,本方法 FPS 为 10.9 帧/s,高于 Mask Transfiner 的 6.5 帧/s)。

在可见光场景下,由于可变形卷积能够动态校正空间偏移,有效补偿动物运动导致的特征错位,同时通道注意力机制增强了毛发纹理等高频特征的响应,使模型能精准区分动物轮廓与相似背景。如图 11 可见光部分所示,在第 4 列中目标个体面部被枝叶遮挡的情况下,模型仍可有效将非躯干部分的枝叶排除在分割掩膜部分之外。而对于红外图像,通道注意力机制强化了热辐射梯度差异,能够有效区分动物躯体与遮挡物的边界。然而,当密集枝叶的温度与动物体温相近时(如图 11 红外部分第 3 列所示),当密集枝叶遮挡动物躯体时,由于红外图像缺乏可见光的色彩和纹理差异,ARM 模块易将枝叶误判为动物的低热区;同时,枝叶与皮毛的热融合导致 DCT 高频分量趋于一致,使色彩敏感性失效,最终难以排除对枝叶的掩膜分配。

图像中的特征分布和模型的边缘提取能力比较(图 12)显示:本研究提出的模型能有效区分这些动物的边缘,并通过整合掩膜类型生成精细的分割掩膜。

4 结 论

本工作针对野生动物监测中动物隐藏、伪装和



图 11 可见光与红外图像分割效果对比图

Fig. 11 Comparison of segmentation results between visible-light and infrared images

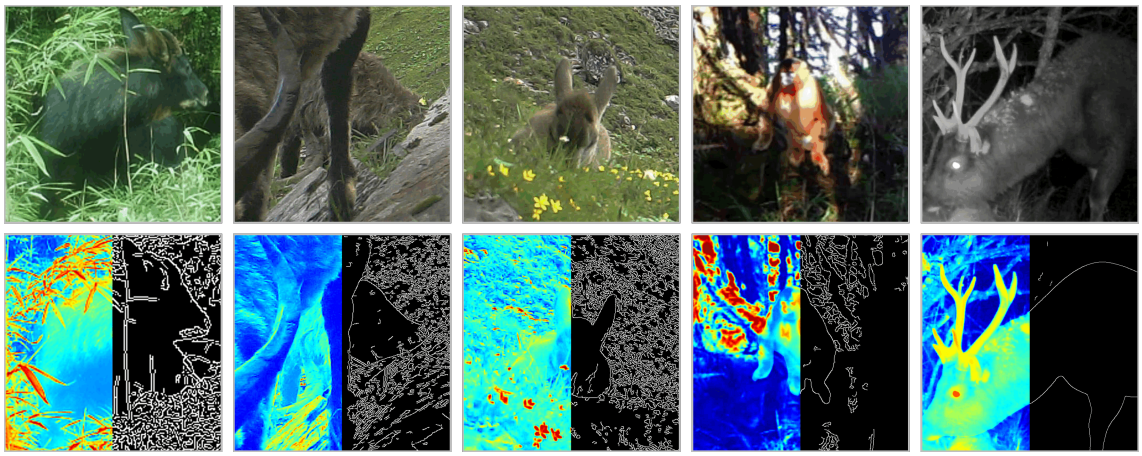


图 12 特征分布与边缘提取效果展示图

Fig. 12 Comparative visualization of feature distribution and edge extraction

复杂地形带来的挑战，提出了一种集成 SE-DAM、ARM 和 PDCO 模块的野生动物图像实例分割方法。该方法通过动态掩膜分辨率分配与高效边缘处理机制，在显著提升分割精度的同时，最大限度地降低了计算资源消耗。具体而言，SE-DAM 模块利用通道注意力机制突出关键特征，并结合可变形卷积调整空间错位，优化特征表示精度；ARM 模块通过通道注意与 Gumbel-Softmax 实现动态分辨率掩膜分配，在极小计算开销下显著增强模型捕捉细节的能力；PDCO 模块则通过融合混合斑块分类与条件随机场（CRF）重点优化边缘结构，提升分割边界的清晰度与准确性。大量实验结果表明，所提方法在降低计算开销的同时显著提升了性能，在 ResNet-101-FPN 主干网络下，mAP 与 mI^2oU 分别达到 53.7% 和 86.71%。此外，帧率稳定在 12.4 帧/s，实现了精度与效率的有效平衡。

尽管本框架在复杂野生动物场景中表现出较高的分割准确率，其性能仍受限于极端光照变化与强背景干扰下的特征鲁棒性。当前模型在跨环境特征泛化能力方面的不足，是野外生态监测系统面临的共性挑战。未来研究将探索引入自适应图像增强模块，以抑制环境噪声并提升模型在多变光照条件下的稳定性。为进一步验证本方法的广泛适用性，后续工作将开展在鸟类种群上的系统性评估，并推动其在无人机移动监测平台与边缘计算节点中的部署应用，最终服务于实时、高效且保护隐私的生物多样性监测实践。

参　考　文　献

[1] Burton A C, Neilson E, Moreira D, et al. Review: wildlife camera trapping: a review and recommendations for linking surveys to ecological processes [J]. Journal of Applied Ecology, 2015, 52(3): 675−685.

[2] 齐建东, 马鐘添, 郑尚姿. 基于 YOLOv7 的红外相机野生动物图像筛选 [J]. 北京林业大学学报, 2024, 46(2): 143−154.
Qi J D, Ma Z T, Zheng S Z. Wildlife image screening for infrared cameras based on YOLOv7 [J]. Journal of Beijing Forestry University, 2024, 46(2): 143−154.

[3] 刘文定, 李安琪, 张军国, 等. 基于 ROI-CNN 的赛罕乌拉国家级自然保护区陆生野生动物自动识别 [J]. 北京林业大学学报, 2018, 40(8): 123−131.
Liu W D, Li A Q, Zhang J G, et al. Automatic identification method for terrestrial wildlife in Saihanwula National Nature Reserve in Inner Mongolia of northern China based on ROI-CNN [J]. Journal of Beijing Forestry University, 2018, 40(8): 123−131.

[4] Pettorelli N, Laurance W F, O'Brien T G, et al. Satellite remote sensing for applied ecologists: Opportunities and challenges [J]. Journal of Applied Ecology, 2014, 51(4): 839−848.

[5] 张彪, 康峰, 许舒婷. 基于改进 RRT 与 GA 的多目标路径规划——以无人机林区巡检为例 [J]. 北京林业大学学报, 2025, 47(4): 129−141.
Zhang B, Kang F, Xu S T. Multi-objective path planning based on improved RRT and GA: Taking UAV forest inspection as an example [J]. Journal of Beijing Forestry University, 2025, 47(4): 129−141.

[6] Bárbulo Barrios D, Valente J, Van Langevelde F. Monitoring mammalian herbivores via convolutional neural networks implemented on thermal UAV imagery [J]. Computers and Electronics in Agriculture, 2024, 218: 108713.

[7] Wang S, Zhao C, Su Y, et al. A framework for measuring tree rings based on panchromatic images and deep learning [J]. Plant, Cell & Environment, 2025, 48(1): 189−198.

[8] Beery S, Wu G, Rathod V, et al. Context R-CNN: Long term temporal context for per-camera object detection[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 13072−13082.

[9] Li Y, Qi H, Dai J, et al. Fully convolutional instance-aware semantic segmentation[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 4438−4446.

[10] Bos J M, Bovenkerk B, Feindt P H, et al. The quantified animal: Precision livestock farming and the ethical implications of objectification [J]. Food Ethics, 2018, 2(1): 77−92.

[11] Zin T T, Kobayashi I, Tin P, et al. A general video surveillance framework for animal behavior analysis[C]//Proceedings of the 3rd International Conference on Computing Measurement Control and Sensor Network. Matsue: IEEE, 2016: 130−133.

[12] Xu B, Wang W, Falzon G, et al. Automated cattle counting using Mask R-CNN in quadcopter vision system [J]. Computers and Electronics in Agriculture, 2020, 171: 105300.

[13] Tian Z, Shen C, Chen H. Conditional convolutions for instance segmentation[C]//Proceedings of the 16th European Conference on Computer Vision: Vol. 12346. Glasgow: Springer International Publishing, 2020: 282−298.

[14] Wang X, Zhang R, Kong T, et al. SOLOv2: Dynamic and fast instance segmentation[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc. , 2020: 17721−17732.

[15] Ke L, Danelljan M, Li X, et al. Mask transfiner for high-quality instance segmentation[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 4402−4411.

[16] Chen K, Pang J, Wang J, et al. Hybrid task cascade for instance segmentation[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 4969−4978.

[17] Raizman E A, Rasmussen H B, King L E, et al. Feasibility study on the spatial and temporal movement of Samburu's cattle and wildlife in Kenya using GPS radio-tracking, remote sensing and GIS [J]. Preventive Veterinary Medicine, 2013, 111(1-2): 76−80.

[18] Fang Y, Yang S, Wang X, et al. Instances as queries[C]//

Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021: 6890−6899.

[19] Chen X, Girshick R, He K, et al. TensorMask: A foundation for dense object segmentation[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 2061−2069.

[20] Jin X, Xie Y, Wei X S, et al. Delving deep into spatial pooling for squeeze-and-excitation networks [J]. Pattern Recognition, 2022, 121: 108159.

[21] Wen Q, Yang J, Yang X, et al. PatchDCT: Patch refinement for high quality instance segmentation[C]//Proceedings of the 11th International Conference on Learning Representations. Kigali: OpenReview, 2023.

[22] Chen X, Zhang H, Shankar A, et al. Multi-target detection and tracking based on CRF network and spatio-temporal attention for sports videos [J]. Scientific Reports, 2025, 15(1): 6808.

[23] Liu D, Hou J, Huang S, et al. LoTE-Animal: A long time-span dataset for endangered animal behavior understanding[C]//Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision. Paris: IEEE, 2023: 20007−20018.

[24] Chen K, Pang J, Wang J, et al. Hybrid task cascade for instance segmentation[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 4974−4983.

[25] Chen H, Sun K, Tian Z, et al. BlendMask: Top-down meets bottom-up for instance segmentation[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 8570−8578.

[26] Shen X, Yang J, Wei C, et al. DCT-Mask: Discrete cosine transform mask representation for instance segmentation[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 8716−8725.

[27] Zhang G, Lu X, Tan J, et al. RefineMask: Towards high-quality instance segmentation with fine-grained features[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 6857−6865.

[28] Kang Q, Ye F, Li Q, et al. Bubble boundary R-CNN: A multitask model for segmenting and reconstructing overlapping bubbles [J]. Separation and Purification Technology, 2025, 358: 130300.

[29] Zhang R, Tian Z, Shen C, et al. Mask encoding for single shot instance segmentation[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 10223−10232.

[30] Kirillov A, Wu Y, He K, et al. PointRend: Image segmentation as rendering[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 9796−9805.

[31] Huang Z, Huang L, Gong Y, et al. Mask scoring R-CNN[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 6402−6411.

（责任编辑 吴 娟
责任编委 陈志泊）