ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 41,No. 1,Jan. 2026,pp. 147－159
DOI:10. 16337/j. 1004-9037. 2026. 01. 010

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

基于调和函数理论的二阶段遥感目标实例分割算法

李泽坤，史振威，邹征夏

(北京航空航天大学宇航学院,北京 100191)

摘　要：本文提出了一种基于调和背景建模的二阶段实例分割方法,可实现复杂遥感图像背景下目标的快速且精细的实例分割。方法包括2个阶段:第1阶段采用可灵活替换的目标检测器,如YOLOv10(You only look once v10)或DINO(DETR with improved denoising anchor boxes),获取候选目标框;第2阶段设计为“即插即用”的掩膜计算模块,无需额外训练即可基于调和函数模型对背景进行快速回归,并计算前景掩膜,从而提升掩膜计算的精度与鲁棒性。本文方法以调和函数理论及复分析中的相关定理为数学基础,以Dirichlet问题为核心框架,创新性地提出利用局部边界信息推断全局背景的实例掩膜生成策略。通过将Dirichlet问题转化为最小二乘回归形式,算法兼具可实现性与灵活性。在NWPU VHR-10数据集上的实验结果表明,与典型方法相比,本文方法在包围框平均精度(Average precision of boxes, AP-Box)和掩膜平均精度(Average precision of masks, AP-Mask)指标上均取得更优表现,其中AP-Mask指标可以在设定交并比(Intersection over union, IoU)指标为50%时达到92.1%,较现有最佳结果提升2.5个百分点。结果验证了该方法在遥感目标分割任务中的有效性与应用潜力。

关键词：实例分割;背景建模;调和多项式;Dirichlet问题;遥感图像

中图分类号：TP751.1　　**文献标志码**:A

引用格式：李泽坤,史振威,邹征夏.基于调和函数理论的二阶段遥感目标实例分割算法[J].数据采集与处理,2026,41(1):147-159. LI Zekun,SHI Zhenwei,ZOU Zhengxia. Two-stage remote sensing object instance segmentation based on harmonic function theory[J].Journal of Data Acquisition and Processing,2026,41(1):147-159.

引　　言

图像实例分割是计算机视觉领域的重要研究方向,旨在同时实现对图像中目标的识别与像素级分割[1-3]。与仅输出目标边界框的目标检测任务不同,实例分割不仅需要精确定位目标位置,还需划分每个目标的像素区域[4-5]。相比仅区分语义类别的语义分割任务[6],实例分割的粒度更细,对模型结构、特征表达与计算效率均提出了更高要求[7-8]。近年来,随着深度学习算法及并行计算硬件的不断进步,实例分割在医学影像分析、遥感监测与视频目标追踪等多个领域得到广泛应用。同时,MS COCO[9]、PASCAL VOC[10]等公共数据集与挑战赛的建立,为该方向提供了标准化的评测基准,促进了算法性能的持续提升。

现有的实例分割方法主要可分为两类:先检测后分割与先分割后聚合[2,4,11]。前者以Mask R-CNN[4]为代表,沿用目标检测的两阶段范式,即通过候选区域网络(Region proposal network, RPN)[12]或特征金字塔网络(Feature pyramid network, FPN)[13]生成候选框,再在各候选区域内执行像素级掩膜预测。此类方法依托成熟的检测框架,能够充分利用多尺度特征并生成较为精确的掩膜。后一类“先

基金项目:国家自然科学基金(62471014, 62125102, U24B20177)。
收稿日期:2025-02-24;**修订日期**:2025-10-15

分割后聚合”方法，如 CondInst[11]、SOLO[14]等，则通过像素分组或可变形卷积直接将语义分割结果聚合为实例掩膜，不再依赖候选框的生成与回归，适用于多目标并行与复杂场景。近年来，还出现了融合两类思路的混合式算法，通过联合建模语义与实例信息，实现了对图像全局特征的更高效利用。

随着多尺度特征融合与自注意力机制的发展，基于 Transformer 架构[15]的实例分割方法成为研究热点。以 Mask2Former[16]为代表的算法，将多尺度特征编码、跨层注意力与掩膜解码模块深度结合，在复杂场景中显著提升了分割的精度与鲁棒性[17]。此外，BoxInst[18]、SCNet[19]及 CATNet[20]等方法分别在检测头设计、语义分割头优化与多任务协同方面提出了创新方案。已有研究表明，这些方法在 NWPU VHR-10 等遥感数据集上的包围框平均精度（Average precision of boxes, AP-Box）和掩膜平均精度（Average precision of masks, AP-Mask）指标均有明显提升，但在多目标或目标形状复杂的场景下，仍存在精度与推理速度之间的权衡问题。

然而，在遥感图像中，诸如飞机、舰船等具有复杂边界的目标仍给实例分割带来挑战。部分方法生成的掩膜轮廓与真实边界存在显著差距，尤其在背景复杂或目标形状细长的情况下[21-22]。造成这一问题的主要原因包括：(1)部分算法过度依赖高层语义特征而忽视低层边缘信息，导致前景与背景区分能力下降；(2)遥感图像中背景纹理常与目标存在颜色或亮度相似性，使模型在小目标与半遮挡区域的识别上出现困难；(3)现有方法多依赖语义特征区分前景，而忽略了人类视觉更依赖的颜色与纹理差异[23-24]。这些因素在一定程度上限制了实例分割模型在遥感监测等应用中的准确性与稳定性。

针对上述问题，本文提出一种基于调和背景建模的二阶段实例分割方法。该方法充分利用前景与背景之间的色彩与结构差异，以提升边界细节的精确性与掩膜的整体一致性。具体而言，算法的第 1 阶段采用可替换的目标检测框架，如 YOLOv10（You only look once v10）或 DINO（DETR with improved denoising anchor boxes），以满足不同场景对实时性或精度的需求；第 2 阶段基于调和函数模型，在无需额外训练的条件下对背景进行快速回归，通过背景抑制获得更贴合真实边界的前景掩膜。在 NWPU VHR-10 数据集上的实验表明，与典型实例分割方法相比，本文方法在 AP-Box 和 AP-Mask 指标上均取得更优结果，且推理速度与原检测模型相当。实验结果表明，本文方法在 NWPU VHR-10 与 iSAID-mini 数据集上均取得稳定提升，其中在 NWPU VHR-10 数据集上，当设定交并比（Intersection over union, IoU）为 0.5 时，AP-Mask 可以达到 92.1%，较现有最佳结果提升 2.5 个百分点。

1 本文方法

1.1 问题建模

数字灰度图像的通用模型为一个尺寸为 $m \times n$ 的实值矩阵 $\boldsymbol{U} \in \mathbf{R}^{m \times n}$，其中，矩阵 $\boldsymbol{U}$ 中第 i 行、第 j 列的元素记为 $u_{i,j}$（$u_{i,j} \in \mathbf{R}$），且下角标 i 和 j 的取值范围分别为：$i \in \{1, 2, \cdots, m\}$，$j \in \{1, 2, \cdots, n\}$。从成像原理的角度出发，传感器输出的连续信号经空间采样与幅值量化后得到数字灰度图像 $\boldsymbol{U}$[25]。

数字彩色图像的通用模型为 1 个尺寸为 $m \times n \times 3$ 的实值张量 $\boldsymbol{W} \in \mathbf{R}^{m \times n \times 3}$，其中，张量 $\boldsymbol{W}$ 中第 i 行、第 j 列的元素记为 $\boldsymbol{w}_{i,j}$，是 1 个三维实欧几里得空间 $\mathbf{R}^3$ 中的向量，表达式为

$$\boldsymbol{w}_{i,j} = (r_{i,j}, g_{i,j}, b_{i,j})^{\mathrm{T}} \tag{1}$$

式中 $r_{i,j}$、$g_{i,j}$、$b_{i,j}$ 分别为红、绿、蓝 3 个波段的可见光传感器在像素坐标 (i,j) 处采集到的量化灰度值。因此，数字彩色图像 $\boldsymbol{W}$ 等价于 3 个重叠放置的，尺寸均为 $m \times n$ 的灰度图像。所以本文的算法主要基于数字灰度图像 $\boldsymbol{U} \in \mathbf{R}^{m \times n}$ 展开。算法针对彩色图像 $\boldsymbol{W}$ 的应用，是通过分别处理 $\boldsymbol{W}$ 中的红、绿、蓝 3 个波段，然后再将结果重新叠加来实现的。因此本节仅针对灰度图像展开特征分析。

为了便于在分析学框架下描述数字灰度图像,可以将传感器信号描述为1个连续函数f,其定义域为二维实欧几里得空间$\mathbf{R}^2$,值域为$\mathbf{R}$。对于$\mathbf{R}^2$中的任意一个坐标(x,y),函数f给出1个实函数值$f(x,y)$。因此,数字图像$\boldsymbol{U}$可以表述为

$$\boldsymbol{U}=\begin{bmatrix} f(1,1) & f(1,2) & \cdots & f(1,n-1) & f(1,n) \\ f(2,1) & f(2,2) & \cdots & f(2,n-1) & f(2,n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f(m-1,1) & f(m-1,2) & \cdots & f(m-1,n-1) & f(m-1,n) \\ f(m,1) & f(m,2) & \cdots & f(m,n-1) & f(m,n) \end{bmatrix} \tag{2}$$

也即:$u_{i,j}=f(i,j)$。因此直接讨论连续函数f的数学性质也就等价于在讨论灰度图像$\boldsymbol{U}$的数学性质,而且便于引入与连续函数有关的微积分工具来完成更深入的图像特征建模。

在弱噪声假设下,如果对飞机、离岸舰船、汽车等典型遥感目标周围的地面背景成像,那么数字灰度图像$\boldsymbol{U}$对应的传感器信号函数f满足以下两点特征:(1)平滑性:邻域内像素变化缓慢,可由拉普拉斯算子度量,近似满足$\Delta f(x,y)=0,\forall(x,y)\in\Omega$,其中闭集$\Omega$是1个包含数字图像$\boldsymbol{U}$所需全部取值点$\{(x,y):1\leqslant x\leqslant m,1\leqslant y\leqslant n\}$的单连通域;(2)边界唯一延拓:$f$的内部取值完全由边界$\partial\Omega$上的取值唯一决定。

当f在Ω上严格满足Laplace方程$\Delta f(x,y)=0$时,f称为调和函数。由调和函数的经典性质可知,此时上述两点性质可以自洽成立[26-28]。因此,本文以调和函数作为背景灰度图像的数学模型,并在两阶段框架中利用一阶段方法针对候选实例给出的边界框作为局部边界$\partial\Omega$,用以完成背景回归、背景抑制和掩膜提取,完成两阶段的实例分割。

1.2 调和背景函数的理论基础

调和背景函数可由经典的Dirichlet边值问题给出,即

$$\begin{cases} \Delta f(x,y)=0 & (x,y)\in\Omega \\ f(x,y)=g(x,y) & (x,y)\in\partial\Omega \end{cases} \tag{3}$$

式中函数$g(x,y)$在单连通域Ω的边界$\partial\Omega$上连续,表示对于连续背景函数$f(x,y)$在边界$\partial\Omega$上施加的函数值约束。其解的存在唯一性可由标准结果保证[29]。当区域为圆盘或半平面时,可由Poisson积分给出显式表达;一般边界下属于经典Dirichlet问题[27-28],虽然可借助Green函数等工具证明存在性,但直接求解开销较大。

另一方面,调和多项式展开为可计算建模提供了可行基底。对具有分段光滑边界的单连通区域内的任意调和函数,存在以调和多项式为基的局部一致收敛展开[30],表达式为

$$f(x,y)=\sum_{n=0}^{\infty}c_n h_n(x-x_0,y-y_0) \tag{4}$$

式中:h_n为一组线性无关且完备的调和多项式(例如由解析函数的实/虚部构造而得),在以(x_0,y_0)为中心的最大开圆盘内对任意紧子集绝对一致收敛;c_n为对应于调和多项式h_n的展开系数。典型的低阶调和基函数可取为

$$h_0=1,\quad h_1=x,\quad h_2=y,\quad h_3=x^2-y^2,\quad h_4=xy,h_5=x^3-3xy^2,h_6=-y^3+3x^2y,\cdots \tag{5}$$

式(5)为后续的数值近似提供了可实现的有限维逼近空间。关于解析函数存在性与由Taylor级数诱导的调和多项式构造证明过程,可见附录A。

1.3 基于调和多项式的最小二乘背景回归

在以实例候选框为中心、半径不越界的局部区域内，用式(4)的前$k+1$项近似f(泰勒展开及其收敛性可参考文献[31])，表达式为

$$f(x,y)\approx\sum_{i=0}^{k}c_i h_i(x-x_0,y-y_0) \tag{6}$$

本文将函数f和$h_i(i=0,1,\cdots,k)$展开为向量，然后使用最小二乘法求出这$k+1$个系数$c_0,c_1,\cdots,c_k$的值。具体而言，若将式(3)中的$\partial\Omega$取为集合$\{(x,y):x=1,1\leqslant y\leqslant n,$ 或$x=m,1\leqslant y\leqslant n,$或$1\leqslant x\leqslant m,y=1,$或$1\leqslant x\leqslant m,y=n\}$，那么可以利用式(2)中的$2m+2(n-2)$个边界值

$$\begin{aligned}&f(1,1),f(1,2),\cdots,f(1,n),f(2,n),\cdots,\\&f(m,n),f(m,n-1),\cdots,f(m,1),f(m-1,1),\cdots,f(2,1)\end{aligned} \tag{7}$$

来展开连续函数f和h_i，即

$$\begin{cases}\boldsymbol{f}=(f_1,f_2,\cdots,f_{2m+2n-4})^{\mathrm{T}}\\ \boldsymbol{h}_0=(h_{0,1},h_{0,2},\cdots,h_{0,2m+2n-4})^{\mathrm{T}}\\ \boldsymbol{h}_1=(h_{1,1},h_{1,2},\cdots,h_{1,2m+2n-4})^{\mathrm{T}}\\ \vdots\\ \boldsymbol{h}_k=(h_{k,1},h_{k,2},\cdots,h_{k,2m+2n-4})^{\mathrm{T}}\end{cases} \tag{8}$$

$$\begin{cases}f_1=f(1,1),f_2=(1,2),\cdots,f_{2m+2n-4}=f(2,1)\\ h_{0,1}=h_0(1-x_0,1-y_0),h_{0,2}=h_0(1-x_0,2-y_0),\cdots,h_{0,2m+2n-4}=h_0(2-x_0,1-y_0)\\ h_{1,1}=h_1(1-x_0,1-y_0),h_{1,2}=h_1(1-x_0,2-y_0),\cdots,h_{1,2m+2n-4}=h_1(2-x_0,1-y_0)\\ \qquad\qquad\vdots\\ h_{k,1}=h_k(1-x_0,1-y_0),h_{k,2}=h_k(1-x_0,2-y_0),\cdots,h_{k,2m+2n-4}=h_k(2-x_0,1-y_0)\end{cases} \tag{9}$$

式中平移坐标(x_0,y_0)可以取为图像中心点$(m/2,n/2)$。

可以定义矩阵$\boldsymbol{H}=\begin{bmatrix}h_{0,1}&h_{1,1}&\cdots&h_{k,1}\\ h_{0,2}&h_{1,2}&\cdots&h_{k,2}\\ \vdots&\vdots&\ddots&\vdots\\ h_{0,2m+2n-4}&h_{2,2m+2n-4}&\cdots&h_{k,2m+2n-4}\end{bmatrix}$，并将所有组合系数整合为向量$\boldsymbol{c}=(c_0,c_1,\cdots,c_k)^{\mathrm{T}}$，于是式(6)可以表述为标准的最小二乘问题[32]，即

$$\min_{c\in\mathbf{R}^{k+1}}\|\boldsymbol{H}\boldsymbol{c}-\boldsymbol{f}\|_2^2 \tag{10}$$

其闭式解为

$$\boldsymbol{c}=(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-\dagger}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{f} \tag{11}$$

式中$(\cdot)^{-\dagger}$表示Moore-Penrose伪逆，以应对边界点过少或几何分布退化导致的不可逆情形。最终的调和背景估计为

$$f_{\mathrm{harmonic}}(x,y)=c_0h_0(x-x_0,y-y_0)+\cdots+c_kh_k(x-x_0,y-y_0) \tag{12}$$

上述过程由于完全使用了数字图像$\boldsymbol{U}$的边界像素值(见式(7))，因此可称为边界框约束值选取策略，在后文的实验对比中简记为bbox(Bounding box)策略；在具体实践中，还可以随机选用数字图像$\boldsymbol{U}$中的其他灰度值，并允许人为设定展开向量$\boldsymbol{f}$和$\boldsymbol{h}_i(i=1,2,\cdots,k)$的长度为$r$，这种策略可称为随机约束

值选取策略,在后文的实验对比中简记为rand(random)策略。

1.4 调和背景建模模块与两阶段框架集成

本文提出的调和背景建模(Harmonic background modelling, HBM)模块作为即插即用的第2阶段掩膜生成器,与第1阶段检测器解耦,步骤为:

(1)候选框生成(第1阶段):由检测器输出候选框集合,每个候选框对应1个局部实例区域及其边界像素集合。

(2)调和背景回归(第2阶段):对每个候选框,按第1.3节构建$\boldsymbol{H}$、$\boldsymbol{f}$并求解组合系数$\boldsymbol{c}$,回归得到局部调和背景f_{harmonic}。然后通过式(2)的采样策略对f_{harmonic}进行采样和量化,得到调和背景图像$\boldsymbol{U}_{\text{harmonic}}$。

(3)前景抑扬与二值化:计算$\boldsymbol{U}$和$\boldsymbol{U}_{\text{harmonic}}$的通道内逐像素差,并取绝对值,送入tanh非线性增强对比度,随后采用大津法完成二值分割,得到实例掩膜为

$$\boldsymbol{M}=\text{binary}\left(\tanh\left(\left|\boldsymbol{U}-\boldsymbol{U}_{\text{harmonic}}\right|\right)\right) \tag{13}$$

式中binary(•)、tanh(•)和|•|等函数操作都是对矩阵逐元素进行。

(4)实例后处理:当单框内包含多实例(密集或回归不准)时,保留面积最大的连通域作为有效实例,其余噪声剔除。

图1给出了HBM模块的工作流程示意。由于HBM与检测头解耦,整体推理速度主要由第1阶段检测器决定;HBM的自身运行时间与复杂度在第2节消融实验中以频率,单位FPS(Frames per second)与参数对比形式报告。

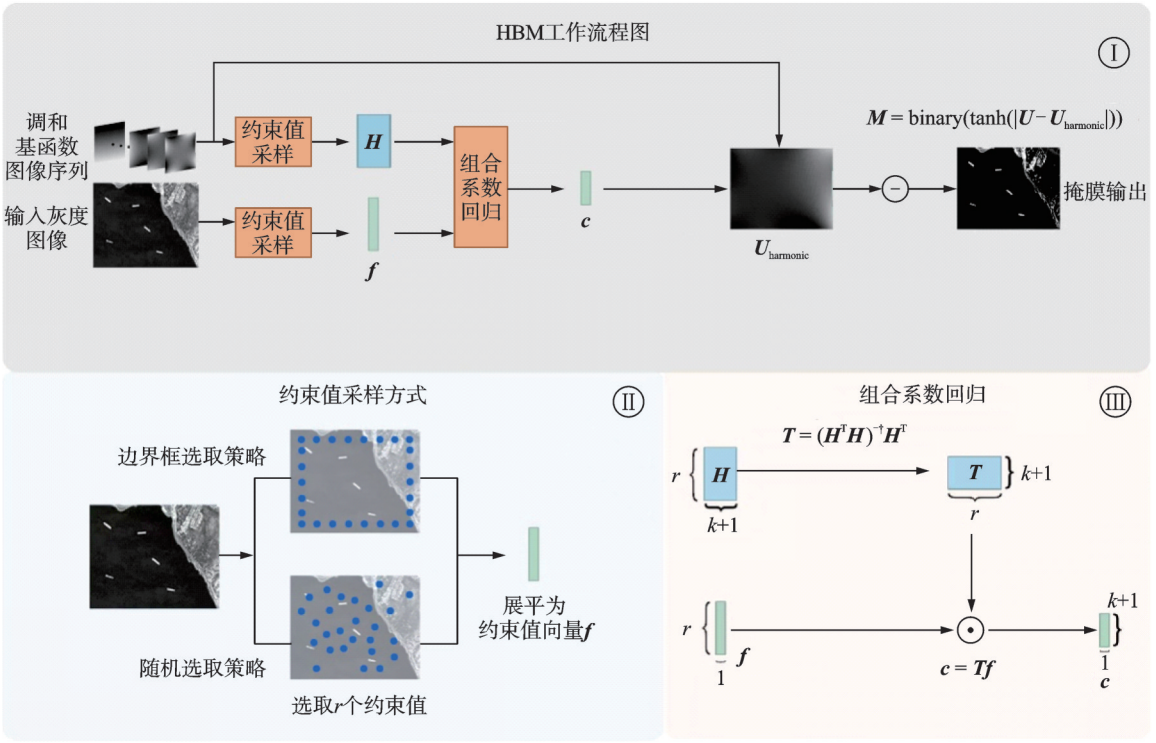


图1 HBM模块的全程工作原理示意图

Fig.1 Schematic diagram of the entire working principle of the HBM module

1.5 实现细节与复杂度分析

对RGB图像按通道独立回归背景并求差；令每个像素为三元组$(r,g,b)^{\mathrm{T}}$，将其以二范数$\sqrt{r^2+g^2+b^2}$汇聚为单标量后送入tanh，再以大津法阈值化得到前景掩膜。

有限项截断阶数k需在表达能力与数值稳定性间折中：k过小会欠拟合，过大会引入高阶项数值放大与病态性。Moore-Penrose伪逆在边界点稀疏或几何退化时提供稳定解。实际实现中，k与边界采样数r一并调节，经验上$r \gg k$可缓解病态。

对单候选框，构造$\boldsymbol{H}$的代价与$r(k+1)$成正比；最小二乘可通过奇异值分解算法[33]实现，时间复杂度约为$O(r(k+1)^2)$；重建$\boldsymbol{U}_{\mathrm{harmonic}}$代价为$O(r(k+1))$。因此，HBM的主导开销与边界采样数$r$和截断阶数$k$相关，而与全图大小弱相关，便于在多尺度场景下保持高效。

HBM不参与第1阶段检测器的回归与损失计算，属于独立的第2阶段掩膜回归模块，不是对检测器的结构性改动。整体吞吐与第1阶段模型强相关，HBM的自有代价在消融实验中单列呈现。

2 实验结果与分析

2.1 数据集与评价指标

本文采用NWPU VHR-10和iSAID两个遥感图像数据集作为测试基准。NWPU VHR-10共含800幅VHR光学遥感图像，其中715幅来自Google Earth的彩色图像（空间分辨率为0.5～2 m），85幅为自Vaihingen数据集锐化得到的彩色红外图像（空间分辨率为0.08 m），包含10个目标类别（见表1）。

表1 NWPU VHR-10数据集实例数量统计

Table 1 Statistics of the number of instances in NWPU VHR-10 dataset

实例类型	数量	实例类型	数量
飞机	757	篮球场	159
船只	302	田径场	163
储油罐	655	港口	224
棒球场	390	桥梁	124
网球场	524	车辆	477

iSAID数据集[34]包含2 806幅高清光学遥感图像与15类目标，共655 451个实例标注。本文按照iSAID数据集原始的各类别数据比例，进行等比例抽样，从完整的iSAID训练集和测试集中随机选取800幅图像构建iSAID-mini，用于第2轮对比评测。

采用平均精度（Average precision，AP）作为核心指标，并区分检测框与掩膜两类度量：AP-Box与AP-Mask；同时报告IoU阈值为50%与75%时的分段指标AP^{50}与AP^{75}。推理效率以FPS度量。

2.2 实验设置

两个数据集均按训练/测试比例7∶3划分。所有实验在单卡NVIDIA RTX 4090上完成，框架为PyTorch。对比方法选取具有代表性的实例分割模型：Mask R-CNN[4]、SOLOv2[14]、BoxInst[18]、Mask2Former[16]、SCNet[19]与CATNet[20]。上述模型均加载MMDetection[35]基于COCO2017[9]的预训练权重后在本文数据上微调。

第1阶段检测器选用YOLOv10[36]与DINO[37]，分别代表轻量级与基于Transformer[15]的检测框架，均使用COCO2017预训练权重。HBM模块默认参数设为$k=11$，$r=200$，候选框缩放至40×60的固定尺寸作为局部输入。

2.3 算法消融实验

为评估HBM中关键超参数对性能的影响，选取基函数阶数k、关键点数量r与采样策略（bbox和

rand)三项变量,在NWPU VHR-10上进行消融,指标为AP-Mask和FPS。候选框统一缩放至40×60;其最外层边界像素数为$2\times(40+60)-4=196$,作为"单层边界覆盖"的参考数量级。

NWPU VHR-10上的消融实验结果如表2所示。由表2可见,当r处于与单层边界像素数大致相同的规模(100~400)时,AP-Mask的变化不敏感,说明局部边界信息的有限采样(约占局部像素的10%)已足以在调和约束下重建背景。另一方面,随着k增大,性能在$k\approx10$附近达到最大值;继续增大易引入高阶项导致的数值不稳定或过拟合。综合精度与效率,$k=10$, $r=200$为稳健选择。图2展示了不同设定下的可视化结果。由图2可见,随机采样可能落在前景上,导致背景回归混入前景色,且在r值不变的情况下,k值越大越容易出现过拟合。

表2 NWPU VHR-10上的消融实验结果

Table 2 Results of ablation study on the NWPU VHR-10 dataset

k	r	约束值选取策略	AP-Mask/%	频率/FPS
6	100	rand	52.4	285.65
6	100	bbox	53.7	257.00
6	200	rand	55.2	247.26
6	200	bbox	57.3	221.29
6	400	rand	52.4	215.67
6	400	bbox	53.7	193.55
8	200	rand	64.8	190.97
8	200	bbox	65.3	182.86
10	200	rand	65.2	136.97
10	200	bbox	**66.3**	176.94
12	200	rand	54.7	164.56
12	200	bbox	66.1	152.08

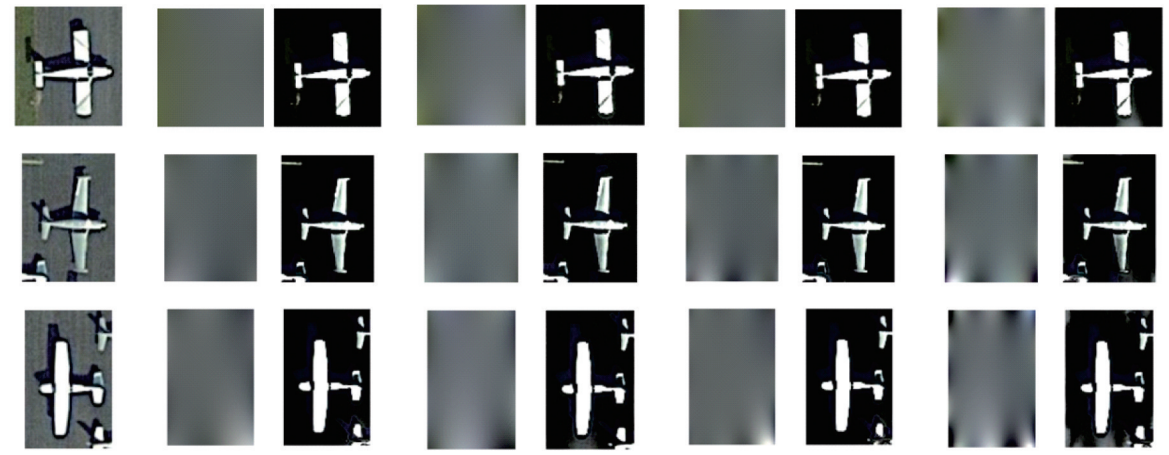


图2 消融实验对比可视化结果示例

Fig.2 Visualization examples of the ablation study

2.4 NWPU VHR-10数据集上的对比实验

表3给出在NWPU VHR-10数据集上的对比结果。HBM与两种检测器集成后在AP-Box和AP-Mask上均取得优势;其中DINO+HBM的AP^{50}_{mask}达到92.1%,较对比方法最佳结果提升2.5个百分点。图3显示,在背景与前景对比弱或形状细长的实例上,HBM能维持更完整的掩膜边界。

2.5 iSAID-mini数据集上的对比实验

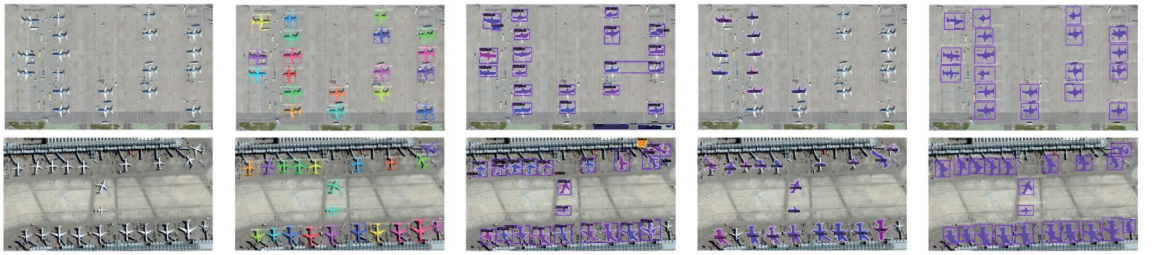
表4为iSAID-mini数据集上的对比实验结果。HBM仍带来整体性提升,但因该数据集背景复杂度更高,绝对数值相对NWPU VHR-10略低;其中DINO+HBM的AP^{50}_{Mask}为72.1%,较最佳对比方法提升2.5个百分点。图4、5展示了典型的可视化对比结果。

表3 NWPU VHR-10数据集上的对比实验结果

Table 3 Results of comparative experiments on the NWPU VHR-10 dataset %

模型	AP_{Box}	AP^{50}_{Box}	AP^{75}_{Box}	AP_{Mask}	AP^{50}_{Mask}	AP^{75}_{Mask}
Mask R-CNN	62.3	87.5	**75.3**	59.4	89.2	66.2
SOLOv2	—	—	—	51.2	77.5	54.1
BoxInst	64.5	89.2	73.3	47.6	77.2	51.3
Mask2Former	57.5	75.5	63.7	58.8	83.1	63.5
SCNet	60.0	87.5	69.2	58.1	87.4	62.0
CATNet	63.2	89.0	73.8	60.4	89.6	64.1
YOLOv10 + HBM (Ours 1)	68.2	91.5	72.9	65.2	91.5	72.2
DINO + HBM (Ours 2)	**69.3**	**92.2**	73.4	**66.3**	**92.1**	**73.3**

注：表中加黑数值为最优值；下划线数值为次优值；—表示未测该指标。



(a) 原始图像 (a) Original image　(b) 地面真实值 (b) Ground truth　(c) Mask2Former模型 (c) Mask2Former model　(d) SOLOv2模型 (d) SOLOv2 model　(e) DINO+HBM(本文方法2) (e) DINO+HBM (Ours 2)

图3 NWPU VHR-10数据集实例分割可视化结果示例

Fig.3 Visualization examples of the instance segmentation results on the NWPU VHR-10 dataset

表4 iSAID-mini数据集上的对比实验结果

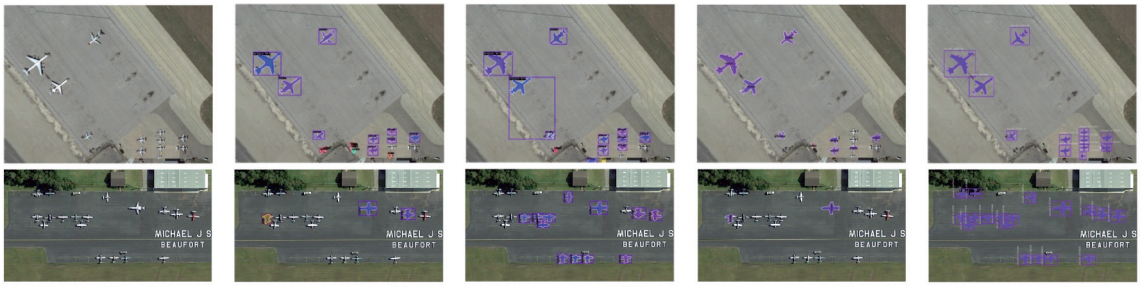
Table 4 Results of comparative experiments on the iSAID-mini dataset %

模型	AP_{Box}	AP^{50}_{Box}	AP^{75}_{Box}	AP_{Mask}	AP^{50}_{Mask}	AP^{75}_{Mask}
Mask R-CNN	48.3	67.5	55.3	39.4	68.2	43.5
SOLOv2	—	—	—	31.2	57.5	34.1
BoxInst	50.5	65.5	53.3	35.6	57.2	43.3
Mask2Former	47.5	**69.2**	**59.2**	38.8	63.1	46.2
SCNet	45.0	67.5	49.2	38.1	67.4	42.0
CATNet	50.2	64.0	53.8	40.4	69.6	44.1
YOLOv10 + HBM (Ours 1)	54.2	68.3	52.9	41.2	71.5	52.2
DINO + HBM (Ours 2)	**55.3**	68.7	53.4	**42.3**	**72.1**	**53.3**

注：表中加黑数值为最优值；下划线数值为次优值；—表示未测该指标。

2.6 结果分析与讨论

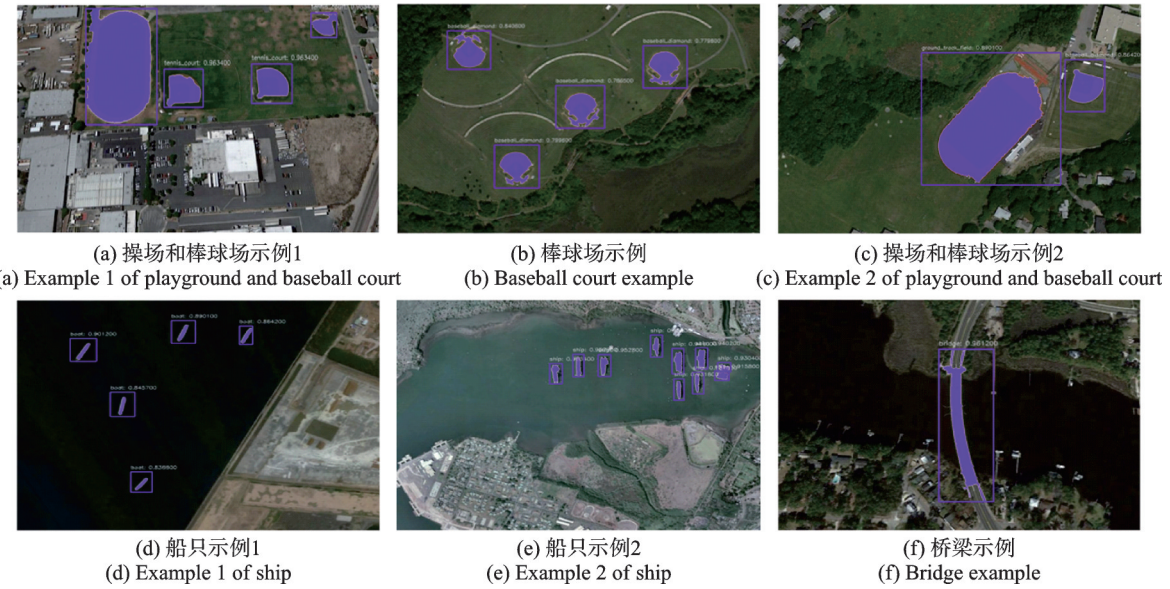
综合两数据集的结果，HBM在不同背景复杂度下均带来稳定的性能增益：在NWPU VHR-10上，AP^{50}_{Mask}为92.1%，较最佳对比方法CATNet提升2.5个百分点；在iSAID-mini上，AP^{50}_{Mask}为72.1%，提升2.5个百分点，同时AP^{75}_{Mask}提升7.1个百分点(46.2%提升至53.3%)。这表明基于局部边界信息的调和背景回归能有效增强前景/背景对比，从而改善实例掩膜的一致性与边界完整性。



(a) 原始图像 (a) Original image
(b)Mask R-CNN模型 (b) Mask R-CNN model
(c) Mask2Former模型 (c) Mask2Former model
(d) SOLOv2模型 (d) SOLOv2 model
(e) DINO+HBM(本文方法2) (e) DINO+HBM (Ours 2)

图4 iSAID-mini数据集可视化对比结果

Fig.4 Visualization results on the iSAID-mini dataset



(a) 操场和棒球场示例1 (a) Example 1 of playground and baseball court
(b) 棒球场示例 (b) Baseball court example
(c) 操场和棒球场示例2 (c) Example 2 of playground and baseball court
(d) 船只示例1 (d) Example 1 of ship
(e) 船只示例2 (e) Example 2 of ship
(f) 桥梁示例 (f) Bridge example

图5 针对操场、棒球场、船只、桥梁等其他实例的可视化分割结果示例

Fig.5 Examples of visualization segmentation results for other instances such as playgrounds, baseball fields, ships, bridges, etc

性能差异的主要限制因素包括:(1)当前超参数k、r采用全局固定设定,未对实例尺度与纹理复杂度进行自适应;(2)边界采样策略仍以候选框外缘为主,未严格贴合真实实例外侧的背景带。未来工作将借助HBM的可微性,引入可学习的注意力式边界采样与自适应阶数控制,以进一步提升复杂背景场景下的泛化与稳健性。

3 结束语

本文提出了一种新的调和背景建模HBM算法,用于解决实例分割任务中的背景复杂性问题。通过在NWPU VHR-10和iSAID-mini数据集上的实验结果可以看出,本文算法在不同复杂度的背景下均表现出显著的性能提升。在NWPU VHR-10数据集上,算法在AP^{50}_{Mask}指标上达到了92.1%,比其他算法的最高值提高了2.5个百分点;在iSAID-mini数据集上,尽管背景复杂度较高,本文算法仍然在AP^{50}_{Mask}指标上达到了72.1%,比其他算法的最高值提高了2.0个百分点。这验证了本文提出的算法在处

理复杂背景时的有效性和鲁棒性。未来的研究工作可以进一步优化调和背景建模算法,使其在更多不同类型的数据集上表现出更好的性能。此外,还可以探索该算法在其他计算机视觉任务中的应用,如自然图像的目标检测和语义分割等。

附录A 调和背景建模算法的理论支撑

本附录补充正文中不便展开的定理证明与必要推导。

A.1 解析函数存在性与调和共轭

定理1(解析函数存在性—调和共轭) 设$\Omega\subseteq\mathbf{R}^2$为具有分段光滑边界的单连通区域,$u\in C^2(\Omega)\cap C(\bar{\Omega})$在$\Omega$上调和,即$\Delta u=0$。则存在$v\in C^2(\Omega)$使得

$$\frac{\partial v}{\partial x}=-\frac{\partial u}{\partial y},\ \frac{\partial v}{\partial y}=\frac{\partial u}{\partial x} \tag{A-1}$$

从而$f(z)=u(x,y)+\mathrm{i}v(x,y)$在$\Omega$上解析,且$u=\mathrm{Re}(f)$。这里$C^2(\Omega)$表示在定义域$\Omega$上,二阶可导,并且二阶导数连续的函数集合;$C(\bar{\Omega})$表示在定义域$\Omega$的闭包(也即$\Omega\cup\partial\Omega$)上连续的函数集合;$\mathrm{Re}(f)$表示对复数值函数$f$提取实部,亦即,$\mathrm{Re}(f)$表示函数值$f$的实部;$\mathrm{Im}(f)$表示对复数值函数$f$提取虚部,亦即,$\mathrm{Im}(f)$表示函数值$f$的虚部。

证明 记一阶微分形式$w=-(\partial u/\partial y)\mathrm{d}x+(\partial u/\partial x)\mathrm{d}y$。由Green定理与$\Delta u=0$得$\oint_{\partial\Gamma}w=0$,对任意分片光滑闭曲线$\partial\Gamma\subseteq\Omega$成立,故$w$为对应保守场(积分满足路径无关)。单连通性保证存在势函数v使$w=\mathrm{d}v$,即给出Cauchy-Riemann方程组。令$f=u+\mathrm{i}v$,则f解析[28]。

A.2 由Taylor级数诱导的调和多项式展开

定理2(Taylor级数) 若f在Ω上解析,$z_0\in\Omega$,则在以z_0为中心的最大开圆盘内,有

$$f(z)=\sum_{n=0}^{\infty}a_n(z-z_0)^n,\quad a_n=\frac{f^{(n)}(z_0)}{n!} \tag{A-2}$$

命题1(调和函数的局部多项式展开) 在定理1的条件下,任意$u=\mathrm{Re}(f)$可在以(x_0,y_0)为中心的最大开圆盘$B((x_0,y_0),r_{\max})$内表示为

$$u(x,y)=\sum_{n=0}^{\infty}c_n h_n(x-x_0,y-y_0) \tag{A-3}$$

式中h_n为调和多项式基函数,级数对任意紧子集绝对一致收敛。

证明 由定理2,$f(z)=\sum_{n=0}^{\infty}a_n(z-z_0)^n$。设$z-z_0=(x-x_0)+\mathrm{i}(y-y_0)$,则$(z-z_0)^n$可展开为实部与虚部的齐次多项式线性组合,记$p_n+\mathrm{i}q_n$,且$p_n$、$q_n$为调和齐次多项式(因解析函数的实部与虚部均为调和函数)。于是

$$u(x,y)=\mathrm{Re}(f)=\mathrm{Re}(a_0)+\sum_{n=1}^{\infty}\Big(\mathrm{Re}(a_n)p_n-\mathrm{Im}(a_n)q_n\Big) \tag{A-4}$$

令$h_0=1$, $h_{2n-1}=p_n$, $h_{2n}=q_n$, $c_0=\mathrm{Re}(a_0)$, $c_{2n-1}=\mathrm{Re}(a_n)$, $c_{2n}=-\mathrm{Im}(a_n)$,即得所需展开与一致收敛性[28,30]。

一组显式的低阶调和基 由$(x+\mathrm{i}y)^n$的实虚部可得常用基(与式(5)一致),有

$$h_0=1,\quad h_1=x,\quad h_2=y,\quad h_3=x^2-y^2,\quad h_4=xy,\quad h_5=x^3-3xy^2,\quad h_6=-y^3+3x^2y,\quad \cdots \tag{A-5}$$

它们均为齐次调和多项式,线性无关且在局部盘内完备,适合作为有限维逼近空间的基底。

A.3 Dirichlet问题与可计算逼近

定理3(Dirichlet问题解的存在唯一性) 设Ω为有界区域且边界分段光滑,给定$g\in C(\partial\Omega)$,则边值问题(3)的解存在且唯一[19,31]。

当边界为一般形状时虽然存在Green函数和势理论工具,但直接进行数值实现开销大。结合命题1的局部完

备性,用有限阶调和多项式子空间作逼近可显著降低复杂度,并便于与候选框(局部边界)结合。

A.4 与经典工具的关系

当区域为圆盘、半平面等标准域时,可用Poisson核给出显式解[28]。在一般边界下,Green函数与层位势理论可证明存在唯一性[29],但数值代价较大。本文采用的有限维调和多项式子空间+最小二乘路径,可视为对经典势论的一种可计算替代,尤其适合与局部候选框边界的现代视觉流程相衔接。

参考文献:

[1] ZHANG C, LIU L, CUI Y, et al. A comprehensive survey on segment anything model for vision and beyond[EB/OL]. (2023-05-14). https://doi.org/10.48550/arXiv.2305.08196.

[2] HAFIZ A M, BHAT G M. A survey on instance segmentation: State of the art[J]. International Journal of Multimedia Information Retrieval, 2020, 9(3):171-189.

[3] MOLINA J M, LLERENA J P, USERO L, et al. Advances in instance segmentation: Technologies, metrics and applications in computer vision[J]. Neurocomputing, 2025, 625: 129584.

[4] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision(ICCV). Venice, Italy: IEEE, 2017: 2980-2988.

[5] 李烨, 周生翠, 张驰. MSDAB-DETR: 一种多尺度遥感目标检测算法[J]. 数据采集与处理, 2024, 39(6): 1455-1469.
LI Ye, ZHOU Shengcui, ZHANG Chi. MSDAB-DETR: A multi scale remote sensing target detection algorithm[J]. Journal of Data Acquisition and Processing, 2024, 39(6): 1455-1469.

[6] 张鹏, 彭宗举, 张文瑞, 等. 多级注意力特征优化的道路场景实时语义分割[J]. 数据采集与处理, 2024, 39(6):1505-1516.
ZHANG Peng, PENG Zongju, ZHANG Wenrui, et al. Real time semantic segmentation of road scene based on multi level attention feature optimization[J]. Journal of Data Acquisition and Processing, 2024, 39(6): 1505-1516.

[7] IAN G, YOSHUA B, AARON C. Deep learning[M]. Massachusetts, USA: The MIT Press, 2016.

[8] CHARISIS C, ARGYROPOULOS D. Deep learning-based instance segmentation architectures in agriculture: A review of the scopes and challenges[J]. Smart Agricultural Technology, 2024, 8: 100448.

[9] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: Common objects in context[C]//Proceedings of the 13th European Conference on Computer Vision (ECCV). Zurich, Switzerland: Springer, 2014: 740-755.

[10] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The pascal visual object classes (VOC) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303-338.

[11] TIAN Z, SHEN C H, CHEN H. Conditional convolutions for instance segmentation[C]//Proceedings of the Computer Vision-ECCV 2020 16th European Conference. Glasgow, UK: [s.n.], 2020: 282-298.

[12] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[13] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017: 936-944.

[14] WANG X L, KONG T, SHEN C H, et al. SOLO: Segmenting objects by locations[C]//Proceedings of the Computer Vision-ECCV 2020 16th European Conference. Glasgow, UK: [s.n.], 2020: 649-665.

[15] ASHISH V, NOAM S, NIKI P, et al. Attention is all you need[EB/OL]. (2017-06-12). https://doi. org/10.48550/arXiv.1706.03762.

[16] CHENG B, MISRA I, SCHWING A G, et al. Masked-attention mask transformer for universal image segmentation[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE, 2022: 1280-1289.

[17] KHAN S, NASEER M, HAYAT M, et al. Transformers in vision: A survey[EB/OL]. (2021-01-04). https://doi. org/10.48550/arXiv.2101.01169.

[18] TIAN Z, SHEN C, WANG X, et al. BoxInst: High-performance instance segmentation with box annotations[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Nashville, USA:

IEEE, 2021: 5439-5448.

[19] VU T, KANG H, YOO C D. SCNet: Training inference sample consistency for instance segmentation[C]//Proceedings of the AAAI Conference on Artificial Intelligence. [S.l.]: AAAI, 2021: 2701-2709.

[20] LIU Y, LI H, HU C, et al. CATNet: Context aggregation network for instance segmentation in remote sensing images[EB/OL]. (2021-11-22). https://doi.org/10.48550/arXiv.2111.11057.

[21] WANG S, CHEN W, XIE S M, et al. Weakly supervised deep learning for segmentation of remote sensing imagery[J]. Remote Sensing, 2020, 12(2): 207.

[22] SU H, WEI S, YAN M, et al. Object detection and instance segmentation in remote sensing imagery based on precise mask R-CNN[C]//Proceedings of the IGARSS 2019 IEEE International Geoscience and Remote Sensing Symposium. Yokohama, Japan: IEEE, 2019: 1454-1457.

[23] HUERTA I, AMATO A, ROCA X, et al. Exploiting multiple cues in motion segmentation based on background subtraction [J]. Neurocomputing, 2013, 100: 183-196.

[24] PAPALE P, LEO A, CECCHETTI L, et al. Foreground-background segmentation revealed during natural image viewing[J]. eNeuro, 2018, 5(3): ENEURO0075-18.

[25] GONZALEZ R C, WOODS R E. Digital image processing[M]. 4th ed. New York: Pearson, 2018.

[26] ROGER F H. Time-harmonic electromagnetic fields[M]. New York: McGraw-Hill, 1961.

[27] TRISTAN N. Visual complex analysis[M]. Oxford: Oxford University Press, 1999.

[28] LARS V A. Complex analysis[M]. 3rd ed. New York: McGraw-Hill, 1979.

[29] LAWRENCE C E. Partial differential equations[M]. [S.l.]: American Mathematical Society, 2010.

[30] SHELDON A, PAUL B, WADE R. Harmonic function theory[M]. Beijing: World Publishing Corporation, 2004.

[31] WALSH J L. On the expansion of harmonic functions in terms of harmonic polynomials[J]. Proceedings of the National Academy of Sciences of the United States of America, 1927, 13(4): 175-180.

[32] BOYD S P, VANDENBERGHE L. Convex optimization[M]. Cambridge: Cambridge University Press, 2004.

[33] RICHARD L B, DOUGLAS F J, ANNETTE M B. Numerical analysis[M]. 10th ed. Boston: Cengage Learning, 2016.

[34] WAQAS Z S, ARORA A, GUPTA A, et al. iSAID: A large-scale dataset for instance segmentation in aerial images[EB/OL]. (2019-05-30). https://doi.org/10.48550/arXiv.1905.12886.

[35] CHEN K, WANG J Q, PANG J M, et al. MMDetection: Open MMLab detection toolbox and benchmark[EB/OL]. (2019-06-17). https://doi.org/10.48550/arXiv.1906.07155.

[36] WANG A, CHEN H, LIU L H, et al. YOLOv10: Real-time end-to-end object detection[EB/OL]. (2024-05-23). https://doi.org/10.48550/arXiv.2405.14458.

[37] CARON M, TOUVRON H, MISRA I, et al. Emerging properties in self-supervised vision transformers[C]//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision(ICCV). [S.l.]: IEEE, 2021: 9630-9640.

作者简介:

李泽坤(2002-),男,硕士研究生,研究方向:遥感图像智能处理、实例分割、目标检测与识别,E-mail:ZachLee@buaa.edu.cn。

史振威(1977-),男,教授,研究方向:遥感图像处理、模式识别、机器学习,E-mail:shizhenwei@buaa.edu.cn。

邹征夏(1991-),**通信作者**,男,教授,研究方向:遥感图像处理、人工智能、计算机视觉、深度学习,E-mail:zhengxiazou@buaa.edu.cn。

(编辑:张黄群)

Two-Stage Remote Sensing Object Instance Segmentation Based on Harmonic Function Theory

LI Zekun, SHI Zhenwei, ZOU Zhengxia*
(School of Astronautics, Beihang University, Beijing 100191, China)

Abstract: Remote-sensing instance segmentation often suffers from ambiguous object boundaries and cluttered backgrounds, while adding heavy mask heads can increase computational cost and reduce deployment flexibility. This paper aims to develop a fast, accurate, and detector-agnostic mask-generation scheme that can be integrated into existing detection pipelines with minimal engineering overhead and without extra training. We propose a two-stage framework that couples a replaceable object detector (e.g., YOLOv10 or DINO) with a plug-and-play harmonic background modelling (HBM) module. For each detected bounding box, HBM treats the local background as a harmonic function and reconstructs it by least-squares fitting of a truncated harmonic-polynomial basis. Boundary constraints are formed by sampling pixel values along the bounding-box boundary, and the coefficients are solved efficiently via the Moore-Penrose pseudoinverse. The foreground mask is then derived from the channel-wise residual between the original image and the reconstructed background, followed by a contrast-enhancing nonlinearity, Otsu thresholding, and connected-component filtering to suppress spurious fragments. The overall pipeline is fully decoupled from the detector: the detector is not modified or retrained, and the additional computation mainly comes from solving a small least-squares problem per proposal rather than processing full-resolution feature maps with a learned segmentation head. Extensive experiments on NWPU VHR-10 and iSAID-mini datasets demonstrate consistent gains in both box and mask metrics, while maintaining high throughput. With DINO as the proposal generator, DINO+HBM achieves AP-Box and AP-Mask of 69.3% and 66.3% on NWPU VHR-10 and reaches AP-Mask-50 of 92.1%, improving the previous best result by 2.5 percentage points. On iSAID-mini, DINO+HBM obtains AP-Box and AP-Mask of 55.3% and 42.3% with AP-Mask-50 and AP-Mask-75 of 72.1% and 53.3%, showing clear benefits under more complex scenes. Ablation studies further verify the roles of truncation order, constraint-point number, and sampling strategy, and indicate that bounding-box boundary sampling is more stable than random sampling for background regression and mask extraction without sacrificing speed. The proposed training-free harmonic background suppression provides an efficient way to obtain boundary-faithful instance masks in remote-sensing images and offers a practical, modular add-on to detector-based pipelines when rapid inference and easy deployment are required.

Highlights:

1. A training-free, plug-and-play harmonic background modelling (HBM) module is introduced to generate instance masks from detector proposals without modifying the detector.
2. Local background reconstruction is cast as a Dirichlet-type harmonic regression problem and solved efficiently via a truncated harmonic-polynomial basis and least-squares fitting under boundary constraints.

Key words: instance segmentation; background modelling; harmonic polynomials; Dirichlet problem; remote sensing imagery

---

Foundation items: National Natural Science Foundation of China (Nos.62471014, 62125102, U24B20177).
Received: 2025-02-24; **Revised**: 2025-10-15
***Corresponding author**, **E-mail**: zhengxiazou@buaa.edu.cn.