

文章编号: 2096-1618(2022)05-0520-07

# 基于瓦片重叠法的在线高分遥感图像 目标智能提取方法研究

禹政阳<sup>1</sup>, 陈军<sup>1</sup>, 江明桦<sup>1</sup>, 马世岩<sup>1</sup>, 秦鑫<sup>2</sup>

(1. 成都信息工程大学资源与环境学院, 四川 成都 610225; 2. 四川省水利科学研究院, 四川 成都 610072)

**摘要:**在线遥感影像上目标智能提取是地表机器智能感知的重要途径。然而,如何利用对单个图像大小限制的 Mask R-CNN 模型在线遥感影像上指定空间范围内有效提取空间目标,是亟待解决的关键问题。为此,提出了一种瓦片重叠法以实现在线高分遥感影像上大范围目标进行智能提取。首先,依据高分遥感影像上空间目标大小,设定检测单元的宽度和高度均为 512 个像素,并按一定的步长将相邻瓦片拼接成一系列相互重叠的检测单元;然后,利用 Mask R-CNN 提取局部范围的空间目标,并将其转换为空间坐标;最后,通过空间重叠融合相邻检测单元的重复目标。实验发现,当重叠率为 50% 时,能以最优的性能达到较高的提取精度。以在线地图上的运动场、城市路口、跨河大桥为例,开展了中国部分城市的在线高分遥感影像目标提取实验,平均目标提取率为 80% 左右,为地表机器智能感知提供了新途径。

**关键词:**人工智能;Mask R-CNN;目标提取;在线遥感影像;地表机器智能感知

**中图分类号:**TP751

**文献标志码:**A

**doi:**10.16836/j.cnki.jcuit.2022.05.006

## 0 引言

从海量的、有噪声和模糊随机的遥感影像中提取出有用信息,一直是遥感信息提取和识别领域的研究难点与热点。随着人工智能技术,尤其是深度学习的突破,基于人工智能的遥感信息提取已成为重要的研究方向<sup>[1-2]</sup>。随着研究的深入,大量人工智能模型应用于图像处理领域,如 AlexNet<sup>[3]</sup>, VGGNet<sup>[4]</sup>, GoogleNet<sup>[5]</sup>, UNET<sup>[6]</sup>, Mask R-CNN<sup>[7]</sup> 等。Mask R-CNN 模型属于一种实例分割的图像智能算法,不仅能准确识别出感兴趣目标在图像上的位置,而且还能以掩膜方式标记目标<sup>[8-15]</sup>。将其应用于遥感图像的目标提取,是地表机器智能感知的重要手段。

然而,遥感目标智能提取的研究目前主要集中于本地遥感影像<sup>[16-17]</sup>。作为一种宝贵的在线资源,在线遥感影像的空间目标智能提取具有重大的理论意义和实践价值,但该领域目前仍鲜有相关研究。受限于显卡的存储和计算能力限制,Mask R-CNN 模型的输入图像大小受一定限制。区别于本地影像,在线遥感影像覆盖全球范围。如何从大量在线遥感瓦片中有效提取空间目标,是亟待解决的关键问题。

本文以 18 级在线遥感影像为数据源,开展了在线

遥感影像的目标提取研究,提出了一种瓦片重叠法的在线高分遥感图像目标智能提取方法。该方法在保证目标提取效率的同时,提高了目标检测精度,为地表机器智能感知提供了新的途径。

## 1 空间目标与数据源选择

### 1.1 空间目标选择

在高分遥感影像上,人工构筑物一般具有相对的稳定形态特征和光谱特征,是 Mask R-CNN 模型容易提取的目标。图 1 展示了几种典型的空间目标(运动场、城市路口、跨河大桥)的形态特征和光谱特征。从形态特征上看,运动场主要由球场和周围的塑胶跑道构成,一般呈现为椭圆形或矩形;城市路口是城市道路交叉的地方,道路交叉是其典型的形态特征,具体分为十字型路口、三叉型路口等,城市路口往往可见人工绘制的斑马线;跨河大桥为跨越河流的道路部分,呈现为条形。从光谱特征上看,由人造草坪或者塑胶构成的球场区域在真彩色图像上呈现浅绿色,由塑胶构成的跑道一般呈现橙色;城市路口、跨河大桥光谱特征主要取决于铺面材料,水泥沙路反射率较高,在干爽状态下呈灰白色,沥青路反射率较低,呈暗色调。本文以运动场、城市路口、跨河大桥为例,探讨在线遥感影像的空间目标智能提取方法。

收稿日期:2021-12-24

基金项目:国家重点研发计划地球观测与导航重点专项资助项目(2018YFB0505300);四川省科技计划项目重点研发资助项目(2020YFG0146)



图1 典型样本的空间目标标记

## 1.2 数据源选择

运动场、城市路口、跨河大桥等属于微观地物。18级在线遥感影像的空间分辨率约为0.6 m,运动场、城市路口、跨河大桥清晰可见。因此,以在线遥感影像18级为数据源,尝试提取指定空间范围内的空间目标。

## 2 基于Mask R-CNN的高分遥感影像空间目标提取模型的建立

### 2.1 Mask R-CNN模型概述

Mask R-CNN模型是经典的实例分割模型,从Faster R-CNN发展而来。Faster R-CNN将特征提取、区域建议、目标框定、目标分类整合到一个网络中,实现了图像中目标的识别及框定。由于Faster R-CNN仅能预测出目标在图像上的范围,He等<sup>[7]</sup>将Faster R-CNN中的RoI Pooling层替换成RoIAlign层,提升了目标框的预测精度。同时,在网络中添加了预测目标掩膜(Mask)分支,构建了Mask R-CNN模型。

Mask R-CNN模型所返回的目标数据有两部分,即目标的预测框和掩膜数据。预测框包含目标的分布范围、目标类别和概率。其中,分布范围由矩形的4个角点的图像坐标确定,即像素点在栅格图像上的序号;目标类别以编码方式表示,来源于训练时的目标类别编码;目标概率是指该目标框属于该目标类别的概率。掩膜数据为一个大小为28×28的二维数据,它将预测框划分为28行、28列具有相同大小的单元,每一个单元给出目标在该位置上存在的概率。通过设定掩膜的概率阈值,即可得到目标在预测框中的分布区域。

### 2.2 训练与测试样本数据集的建立

Mask R-CNN是一种监督学习型神经网络,需要准备训练数据集和测试数据集。训练数据集和测试数

据集的每一张图片均为18级在线遥感影像的某一个局部区域,且需事先勾绘出其上的所有运动场、城市路口和跨河大桥。将球场的外边界封闭成的规则区域作为运动场的勾绘区域;将人行斑马线封闭形成区域作为城市路口的勾绘区域;跨河大桥则勾绘出横跨河流的桥梁部分,见图1,红色方框为目标标绘轮廓。

为提高模型的鲁棒性,采集不同光照条件、不同拍摄角度和不同区域的遥感影像作为训练样本。为提高模型训练精度,应正确标绘每一个样本包含的所有空间目标。以全国为范围,随机采集一定数量的样本,构建训练数据集和测试数据集。其中,训练数据集共1000个样本,包含1258个目标;测试数据集共200个样本,包含266个目标,见表1。

表1 样本数据集各目标数量及测试精度

目标	训练样本数	测试样本数	类别精度/%		掩膜精度/%
			召回率	精确率	
运动场	360	82	91.7	98.5	85.6
城市路口	486	96	80.6	93.5	82.4
跨河大桥	412	88	81.6	96.9	82.6
合计	1258	266	84.5	96.4	83.5

### 2.3 模型训练与测试

将数据集代入模型中学习。采用随机梯度下降法(SGD)进行学习,学习率为0.001,权重衰减设置为0.0005;训练共计800轮,损失值达到最低并趋于稳定。训练完成后,利用测试数据集进行测试,并设定目标类别概率阈值和掩膜概率阈值提取图像上的空间目标及掩膜。实验发现,类别概率阈值越低,召回率越高,但精确率越低,这说明较低的概率阈值将导致在没有目标的区域预测出大量无效的目标。为降低模型的错误率,需要较高的类别概率阈值,但又会出现部分目标无法预测。通过反复实验,将运动场、城市路口和跨河大桥的类别概率阈值分别设为0.98、0.97和

0.97,测试数据集总体召回率为84.5%,精确率为96.4%;当掩膜概率阈值设定为0.5时,预测的掩膜区域与目标分布区域拟合效果最佳,掩膜精度为83.5%(表1)。

### 3 基于瓦片重叠法的在线遥感影像空间目标提取

以在线遥感影像为数据源提取空间目标,需要以瓦片为基本单元,利用前述模型进行目标检测与提取。基于瓦片重叠法的在线遥感影像空间目标提取方法,是将若干瓦片组合成一个检测单元,通过相互重叠的检测单元来实现在指定空间范围内的空间目标提取。

#### 3.1 瓦片序列的求取

在线遥感影像以瓦片金字塔进行组织,将全球范围的遥感影像,按照固定的若干个比例尺(瓦片级别)和指定图片尺寸切成若干行及列的正方形图片,以指定的格式保存为图像文件,按一定的命名规则和组织方式存储到目录系统或数据库系统中,形成金字塔模型的静态地图缓存。以 google 和天地图在线遥感影像为例,它们以 Web 墨卡托投影组织影像金字塔数据。每一个瓦片的宽度和高度均为 256 个像素。

在线遥感影像上提取空间目标,需要先指定提取的空间范围,依据空间范围获取瓦片序列。设地图上某一点的经纬度为 $(g_x, g_y)$ ,求取在层级为 $L$ 时该点的像素坐标:

$$\begin{cases} P_x = \frac{g_x + 180}{360} \cdot 2^L \\ P_y = 0.5 + \frac{\lg((1 + \sin l(1 - \sin l)) \cdot 2^L)}{4\pi} \end{cases} \quad (1)$$

式中, $P_x, P_y$ 为该点在层级 $L$ 上的像素坐标。其中:

$$\sin l = \sin(\pi g_y / 180) \quad (2)$$

根据像素坐标,求取该点所处的瓦片在给定层级 $L$ 的瓦片行列位置:

$$\begin{cases} T_x = \text{int}(P_x / 256) \\ T_y = \text{int}(P_y / 256) \end{cases} \quad (3)$$

式中, $T_x, T_y$ 为瓦片在层级 $L$ 下的行列位置,int为取整函数。

设待提取空间目标的经纬度范围的左上角和右上角分别为 $(g_{x1}, g_{y1}), (g_{x2}, g_{y2})$ ,利用式(1)~(3)可求得待提取的瓦片序列:

$$T_s = \{t_{xy} \mid T_{x1} \leq x \leq T_{x2}, T_{y1} \leq y \leq T_{y2}\} \quad (4)$$

式中, $T_s$ 为瓦片序列, $t_{xy}$ 为序列中的每一个瓦片, $(T_{x1}, T_{y1}), (T_{x2}, T_{y2})$ 分别表示待提取的空间范围左上角和右下角的瓦片行列位置。

#### 3.2 在线遥感影像目标检测单元的空间目标提取

##### 3.2.1 检测单元的定义

在18级高分遥感影像上,瓦片的一个像元的分辨率约为0.6 m,单个瓦片在水平和垂直方向上代表实地距离约154 m。这种空间尺度,对于空间尺寸稍大的空间目标,一个瓦片难以完整地表示。本文以单个瓦片的宽度和高度的2倍作为一个检测单元,其宽度和高度均为512个像素。

##### 3.2.2 检测单元上空间目标的矢量化与空间坐标的求取

首先,将检测单元的局部遥感影像代入前述模型,提取其空间目标预测框及掩膜;然后,利用空间目标的预测框,将空间目标的掩膜重采样至原图像分辨率,并以0.5为阈值,对空间目标掩膜逐像素处理。当掩膜值 $\geq 0.5$ 赋值为1,否则赋值为0;在此基础上,将掩膜二值图值为1的像素块矢量化得到空间目标的外边界。

空间目标的掩膜按双线性插值方式进行重采样。设待插值位置坐标为 $(x, y), (x_0, y_0), (x_1, y_0), (x_0, y_1), (x_1, y_1)$ 分别为待插值位置左上、右上、左下和右下最邻近掩膜像元的坐标,其概率值分别为 $z_{00}, z_{10}, z_{01}$ 和 $z_{11}$ 。首先,在 $x$ 方向进行内插,得到 $z_0$ 和 $z_1$ :

$$\begin{cases} z_0 = \frac{(x-x_0)(z_{10}-z_{00})}{x_1-x_0} + z_{00} \\ z_1 = \frac{(x-x_0)(z_{11}-z_{01})}{x_1-x_0} + z_{01} \end{cases} \quad (5)$$

然后,在 $y$ 方向插值得到待插值位置的掩膜概率 $z$ :

$$z = \frac{(y-y_0)(z_1-z_0)}{y_1-y_0} + z_0 \quad (6)$$

设检测单元的左上角空间坐标为 $(L, T)$ ,空间分辨率为 $(C_x, C_y)$ ,将空间目标的矢量转换为空间坐标:

$$\begin{cases} X = L + C_x x \\ Y = T - C_y y \end{cases} \quad (7)$$

式中, $(x, y)$ 和 $(X, Y)$ 为分别为空间目标的像素坐标和空间坐标。

#### 3.3 基于瓦片重叠法的空间目标提取

##### 3.3.1 无缝不重叠瓦片拼接法及其问题

无缝不重叠瓦片拼接法是在线遥感影像上提取空间目标的最简单直接的方式。从瓦片序列的左上角第一个瓦片开始,步长为2,按从左向右、从上到下的顺序依次遍历每一个瓦片,与右侧和下侧的3个相邻瓦片合并为一个检测单元,构成 $512 \times 512$ 的局部遥感影像,利用模型提取其上的空间目标。当所有检测单元的空间目标提取完成后,汇总得到最终的空间目标序列。

由于瓦片的划分是固定的和静态的,不可避免出

现空间目标跨越瓦片的情况,如图2所示(图中绿线划分的每一个方形格网为一个检测单元)。这些跨越瓦片的空间目标,在局部遥感影像上表现为目标“碎片”。虽然在训练数据集中,添加了部分区域的目标碎片作为训练样本,使模型具有空间目标“碎片”的提取能力。但在局部遥感影像范围的“碎片”过小,超出了模型的识别能力,将导致目标提取不完整。如图2(a)所示,图中最大的运动场跨越了4个相邻瓦片,由于左侧的运动场“碎片”比重过小,导致模型无法识别,该运动场最终只提取到右侧的两个“碎片”。在运动场左侧的城市路口,被划分为上下2个“碎片”,这两个“碎片”均超出了模型的提取能力,导致该路口未被识别。



(a) 无缝叠加方法空间目标提取效果



(b) 重叠率50%时的空间目标提取效果

图2 无缝不重叠瓦片拼接法与瓦片重叠法的空间目标提取效果对比

### 3.3.2 瓦片重叠法的算法描述

瓦片重叠法的核心是瓦片重叠,即检测单元之间存在一定的重叠率。通过相互覆盖的瓦片以提高跨越瓦片的目标提取率和空间完整度。设瓦片重叠率为 $o$ ,则目标检测单元的移动步长为

$$\lambda = 512(1-o) \quad (0 < o < 1) \quad (8)$$

设待提取区域有 $m$ 列、 $n$ 行在线瓦片,则待提取空间目标的检测单元数 $N$ 为

$$N = 256^2 mn / \lambda^2 \quad (9)$$

瓦片重叠法遍历瓦片序列的示意图如图3所示。

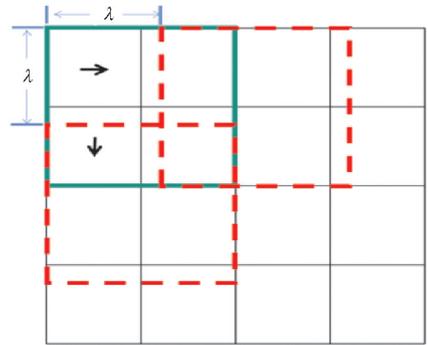


图3 瓦片重叠法示意

使用瓦片重叠法提取空间目标时,重叠区域的空间目标存在重复提取的情况。需要通过融合得到最终的空间目标集。首先,分别提取每一个检测单元的目标集;然后,对于所有检测单元的每一个目标,搜索与之空间相互重叠并且类别编码相同的所有目标,这些目标被认为是同一个空间目标在不同检测单元重复提取的目标。利用空间融合算法将它们融合为一个空间目标。

图2(b)为重叠率50%时空间目标的提取效果。对比图2(a)发现,瓦片重叠法在一定程度上解决了在线地图上无缝瓦片法的空间目标“碎片”难以提取的问题,提高了空间目标的提取率和空间完整度。

### 3.3.3 不同重叠率的空间目标提取性能评价

以成都市区部分范围(面积为 $33.4 \text{ km}^2$ )为实验区域,设定不同的重叠率进行空间目标提取,统计不同瓦片重叠率提取的性能和精度。实验计算机的CPU为Inter(R) Core(TM) i7-7700HQ,显卡为NVIDIA GeForce GTX 1070。为降低网络传输不稳定对性能评价的影响,研究区域内的所有瓦片事先下载到本地。同时,通过目视判断方法统计各重叠率对应的空间目标提取数、漏提数和误提数,并计算提取精度:

$$p = 1 - \frac{N_e + N_a}{N_t + N_a} \quad (10)$$

式中, $p$ 表示提取精度, $N_t$ 表示总提取数, $N_e$ 表示误提数, $N_a$ 表示漏提数。

见表2,当重叠率为0时,采用无缝不重叠瓦片拼接法提取空间目标,其提取精度仅为58.45%,因此需要利用瓦片重叠法提高提取精度。随着瓦片重叠率越高,漏提数不断减少,提取精度不断提高。当重叠率 $\geq 50\%$ 时,提取精度在80%以上。这说明,较高的重叠率,增加了相同位置从不同的“片段”反复识别目标的机会,从而减少了漏提率。但反复识别同一位置,也增加了错误识别的比率,即误提数也相应增加。若仅考虑提取精度,其重叠率在75%左右达到最佳。

表2 不同重叠率的空间目标提取性能对比

重叠率/%	检测单元数	空间目标提取数		漏提数	误提数	精度/%	总耗时/s	检测单元耗时/s
		融合前	融合后					
0	360	125	122	85	1	58.45	48	0.133
20	598	195	145	62	1	69.57	81	0.135
25	648	231	146	62	2	69.23	87	0.134
50	1517	481	171	38	3	80.38	199	0.131
75	5760	1893	193	18	5	89.10	793	0.138
80	9191	3003	201	16	10	88.82	1258	0.137

从性能上分析,忽略瓦片下载时间,单个检测单元目标提取平均耗时约为0.13 s,这意味着每秒可处理7个左右的检测单元。对于同一空间范围,瓦片重叠率越高,待提取空间目标的检测单元越多,总耗时也随之增加。总耗时包括检测单元数据准备时间、目标提取时间和空间目标融合时间。当重叠率为0和50%时,单个检测单元仅需拼接相邻4个瓦片;其余重叠率的单个检测单元则需拼接相邻9个瓦片。因此,无论是网络传输还是瓦片拼接,当重叠率为0和50%时,其性能为最佳;随着重叠率的增加,被检测出来的总目标数越大,其融合所需时间越多。综合以上分析,若仅考虑提取时间性能,其重叠率为0和50%时达到最佳。

若要求提取精度在80%左右,并要求时间性能最优,则重叠率50%为最优参数,其流程为,将步长设定为1个瓦片。对于瓦片序列的每一个瓦片,将其与右

侧、下侧及右下侧的相邻瓦片组合成一个 $512 \times 512$ 大小的局部遥感影像,代入模型中进行空间目标的提取。

#### 4 在线高分遥感影像空间目标提取实验

以中国部分城市为例,利用18级天地图影像按瓦片重叠率50%提取市域某范围内的运动场、城市路口、跨河大桥,共计提取5595个空间目标。通过目视判断方法获取各空间目标的漏提数和误提数,并采用式(9)计算提取精度,实验结果见表3。由表3可知,本文方法的目标提取率约为80%,为地表机器智能感知奠定了基础。

图4为实验城市某局部范围对应的空间目标提取效果。由图4可见,虽然存在1处误提取(A)和2处漏提取(B),但绝大部分空间目标均被正确地提取出来。

表3 中国部分城市在线遥感影像空间目标提取结果

城市	面积/ $\text{km}^2$	提取数量			漏提数			误提数			提取精度/%		
		1	2	3	1	2	3	1	2	3	1	2	3
成都	421.17	56	182	8	7	3	8	2	11	0	85.8	92.5	50
杭州	306.33	38	304	89	4	11	30	3	14	5	83.4	92.1	70.6
武汉	776.52	179	129	132	11	9	3	7	8	14	90.6	87.7	87.5
合计	1504.02	273	615	229	22	23	41	12	33	19	88.5	91.3	77.8

注:1代表运动场,2代表城市路口,3代表跨河大桥。

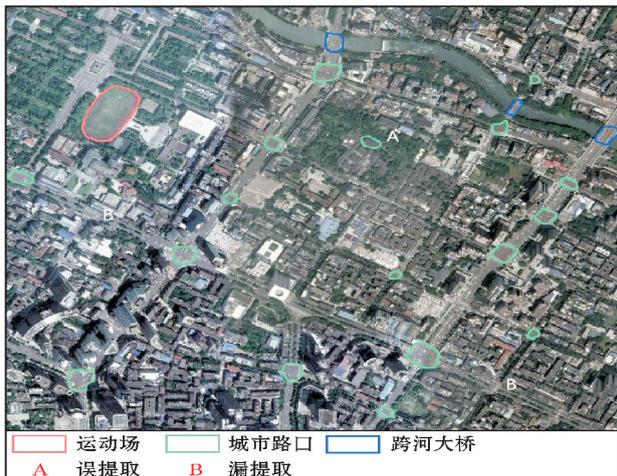


图4 在线遥感影像空间目标提取效果

#### 5 结束语

为提高机器智能感知地表的能力,以在线遥感影像为研究对象,提出了一种基于瓦片重叠法的在线遥感影像的空间目标智能提取方法。本文不仅验证了瓦片重叠法能提高目标提取精度,还通过不同的重叠率对空间目标提取精度和性能进行了分析,发现了重叠率为50%时,在线遥感影像的目标提取性能最优,且具有较高的提取精度。其研究成果对于在线遥感影像空间目标提取的深入研究具有一定的指导意义。

本文不足之处在于仅以三类空间目标为例,进行了高分遥感影像上空间目标提取的研究。若要进一步提高机器智能感知地表领域的的能力,还需要增加空间目标的类别,并利用样本数据集训练模型。

## 参考文献:

- [1] Kulasekara D, Harshanath S. Detecting changed areas in images from different view points[C]. International Conference on Computer Science & Education. IEEE, 2013: 525-528.
- [2] Marmanis D, Datcu M, Esch T, et al. Deep Learning Earth Observation Classification Using ImageNet Pretrained Networks[J]. IEEE Geoscience & Remote Sensing Letters, 2016, 13(1): 105-109.
- [3] Krizhevsky A, Sutskever I, Hinton G. ImageNet Classification with Deep Convolutional Neural Networks[J]. Communications of the ACM, 2012, 60(6): 84-90.
- [4] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[C]. Computational and Biological Learning Society, 2015: 1-14.
- [5] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 1-9.
- [6] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[C]. Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 234-241.
- [7] He Kaiming, Georgia Gkioxari, Piotr Dollar, et al. Mask R-CNN[C]. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, 42(2): 2961-2969.
- [8] Yayla R, Sen B. A New Classification Approach with Deep Mask R-CNN for Synthetic Aperture Radar Image Segmentation[J]. Elektronika ir Elektrotechnika, 2020, 26(6): 52-57.
- [9] Xu Yingying, Li Dawei, Xie Qian, et al. Automatic defect detection and segmentation of tunnel surface using modified Mask R-CNN[J]. Measurement, 2021, 178(4): 109316.
- [10] Kassim Y M, Byrne M E, Burch C, et al. Small Object Bird Detection in Infrared Drone Videos Using Mask R-CNN Deep Learning[C]. Imaging and Multimedia Analytics in a Web and Mobile World, 2021, 2020(8): 85-1-85-8.
- [11] Naufal A, Faticah C, Suciati N. Preprocessed Mask RCNN for Parking Space Detection in Smart Parking Systems[J]. International Journal of Intelligent Engineering and Systems, 2020, 13(6): 255-265.
- [12] Wong CC, Yeh LY, Liu CC, et al. Manipulation Planning for Object Re-Orientation Based on Semantic Segmentation Keypoint Detection[J]. Sensors (Basel). 2021, 21(7): 2280-2286.
- [13] Shi J, Zhou Y, Zhang W X Q. Target Detection Based on Improved Mask Rcnm in Service Robot[C]. 2019 Chinese Control Conference (CCC), 2019: 8519-8524.
- [14] Lu Chuhan, Kong Yang, Guan Zhaoyong. A mask R-CNN model for reidentifying extratropical cyclones based on quasi-supervised thought[J]. Scientific Reports, 2020, 10(1): 15011-15017.
- [15] Lv Y, Zhang C, Yun W, et al. The Delineation and Grading of Actual Crop Production Units in Modern Smallholder Areas Using RS Data and Mask R-CNN[J]. Remote Sensing, 2020, 12(7): 1074-1080.
- [16] Ji Chao, Tang Hong. Number of Building Stories Estimation from Monocular Satellite Image Using a Modified Mask R-CNN[J]. Remote Sensing, 2020, 12(22): 3833-3890.
- [17] Mahmoud A, Yushen He, Qiang Xu, et al. Object Detection Using Adaptive Mask RCNN in Optical Remote Sensing Images[J]. International Journal of Intelligent Engineering and Systems, 2020, 13(1): 65-76.

# Research on Intelligent Extraction Method of Spatial Objects from Online High-Resolution Remote Sensing Images based on Tile-overlapping Strategy

YU Zhengyang<sup>1</sup>, CHEN Jun<sup>1</sup>, JIANG Minghua<sup>1</sup>, MA Shiyan<sup>1</sup>, QIN Xin<sup>2</sup>

(1. College of Resources and Environment, Chengdu University of Information Technology, Chengdu 610225, China; 2. Sichuan water conservancy Research Institute, Chengdu 610072, China)

**Abstract:** The intelligent extraction of spatial objects from online remote sensing images is an important approach for the machine intelligence perception of earth surface. However, how to use the Mask R-CNN model which limits the size of a single image to effectively extract the spatial objects in the specified spatial range on the online remote sensing images is a key problem to be solved. In this paper, an extraction method of spatial objects based on tile-overlapping strategy is proposed to realize the intelligent extraction of spatial objects from online high-resolution remote sensing images. Firstly, according to the size of the spatial objects on the high-resolution remote sensing images, the width and height of the detection unit are set to 512 pixels, and the adjacent tiles are spliced into a series of overlapping detection units according to a certain overlap rate. Then, Mask R-CNN is used to extract the spatial objects from each of the detection units. The local coordinates of each extracted spatial object are converted to spatial coordinates through the coordinates and spatial resolution of the upper left corner of the prediction box. Finally, the repeated objects of adjacent detection units are fused through spatial overlap analysis to obtain the final spatial object sequence. Taking part of Chengdu urban area (with an area of 33.4 km<sup>2</sup>) as the experimental area, set different overlap rates for spatial object extraction, and evaluate the performance and accuracy of different tile overlap rates. The experimental result indicated that the higher the overlap rate, the more opportunities there are to repeatedly identify the same object from different "fragments" at the same location, so as to reduce the leakage. On the other hand, repeated identification of the same location also increases the number of false identification. When the overlap rate is 50%, it can achieve high extraction accuracy with the best performance. Taking the playground, urban intersection and river crossing bridge on the online map as examples, the spatial objects of some cities in China were extracted from online remote sensing images with the the overlap rate set to 50%. The results show that the average object extraction rate is about 80%, which provides a new way for the intelligent perception of earth surface.

**Keywords:** artificial intelligence; Mask R-CNN; object extraction; online remote sensing images; machine intelligent perception of earth surface