

基于轻量化 NDFEDet-SOLOv2 的遥感图像建筑物提取方法

汪 强, 郭来功, 程伟涛

安徽理工大学 电气与信息工程学院, 安徽 淮南 232001

摘要:目的 在地籍测绘和灾害管理等领域中, 建筑物轮廓和位置的自动提取是至关重要的一环。为了解决高分辨率遥感图像建筑物因环境因素导致分割精度不准确等问题, 提出了一种改进的轻量化 SOLOv2 实例分割模型——NDFEDet-SOLOv2。方法 该模型选用双向特征金字塔网络(BiFPN)特征融合方式的轻量级 EfficientDet 网络, 其中将骨干网络部分的 EfficientNet 升级为 EfficientNetv2, EfficientNetv2 中的三层 MBCConv 模块 SE 注意力更换为含有 DropBlock 正则化的轻量级标准化注意力机制(NAM), 构成 NAD-MBCConv 模块。BiFPN 特征融合部分, 向其尾端各特征层并入双水平路由注意视觉变压器(BiFormer), 形成双向水平路由注意特征金字塔网络结构(BiFPN-Former), 从而聚焦微小建筑物轮廓信息, 以实现更高层次的特征融合。结果 NDFEDet-SOLOv2 模型相较于传统轻量级 SOLOv2 实例分割算法, 平均精度 mAP、mAP₅₀ 和 mAP₇₅ 分别提高了 3.9%、3.7% 和 2.5%, 检测帧率(FPS)提高了 2.7 帧/s。结论 轻量化 NDFEDet-SOLOv2 实例分割模型消除了建筑物边角的图像畸变, 在地理环境空间不均等复杂情况下也能准确提取出遥感图像建筑物的基本轮廓, 从而为城市布局更新和建筑变化检测提供理论参考。

关键词: 高分辨率遥感图像; 实例分割; EfficientDet; 标准化注意力机制(NAM); 双水平路由注意视觉变压器(BiFormer)

中图分类号: TP751; TP391 文献标识码: A doi: 10.16055/j.issn.1672-058X.2024.0006.003

A Method for Extracting Buildings from Remote Sensing Images Based on Lightweight NDFEDet-SOLOv2

WANG Qiang, GUO Laigong, CHENG Weitao

School of Electrical and Information Engineering, Anhui University of Science & Technology, Anhui Huainan 232001, China

Abstract: Objective In the fields of cadastral surveying and disaster management, automatic extraction of building contours and positions is crucial. To address the problem of inaccurate segmentation accuracy of buildings in high-resolution remote sensing images due to environmental factors, an improved lightweight SOLOv2 instance segmentation model called NDFEDet-SOLOv2 was proposed. **Methods** This model adopted a lightweight EfficientDet network with a bidirectional feature pyramid network (BiFPN) feature fusion method. The EfficientNet in the backbone network was upgraded to EfficientNetv2, and the three-layer MBCConv module SE attention in EfficientNetv2 was replaced with a lightweight normalized attention mechanism (NAM) containing DropBlock regularization, forming the NAD-MBCConv module. The feature fusion part of BiFPN incorporated a bi-level routing attention visual transformer (BiFormer) into each feature layer at its tail end to form a bi-directional horizontally routing attention feature pyramid network structure (BiFPN-Former), which focused on the contour information of small buildings and achieved higher-level feature fusion.

收稿日期: 2023-09-04 修回日期: 2023-09-11 文章编号: 1672-058X(2024)06-0020-10

基金项目: 国家自然科学基金资助项目(61873004)。

作者简介: 汪强(1997—), 男, 安徽宣城人, 硕士研究生, 从事数字图像处理及 FPGA 开发与应用研究。

通讯作者: 郭来功(1980—), 博士, 副教授, 硕士生导师, 从事数字信号处理、FPGA 的开发与应用、矿山安全监测技术等研究。Email: wq3628123@outlook.com.

引用格式: 汪强, 郭来功, 程伟涛. 基于轻量化 NDFEDet-SOLOv2 的遥感图像建筑物提取方法[J]. 重庆工商大学学报(自然科学版), 2024, 41(6): 20-29.

WANG Qiang, GUO Laigong, CHENG Weitao. A method for extracting buildings from remote sensing images based on lightweight NDFEDet-SOLOv2[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2024, 41(6): 20-29.

Results Compared with traditional lightweight SOLOv2 instance segmentation algorithms, the NDFEDet-SOLOv2 model has improved average accuracies by 3.9%, 3.7%, and 2.5% for mAP, mAP₅₀, and mAP₇₅, respectively, and improved detection frame rate (FPS) by 4.7 frames/s. **Conclusion** The lightweight NDFEDet-SOLOV2 instance segmentation algorithm model eliminates image distortion of building edges and corners, and can accurately extract the basic contours of buildings in remote sensing images even in complex and uneven geographical environments. This provides a theoretical reference for the update of urban layouts and the detection of building changes.

Keywords: high-resolution remote sensing images; instance segmentation; EfficientDet; normalized attention mechanism (NAM); bi-level routing attention visual transformer (BiFormer)

1 引言

地球上,建筑物是一种以人类智慧和为基础,通过人工手段构建而成的实体形态^[1]。在地籍测绘和灾害管理等领域,建筑物轮廓和位置的自动准确提取是至关重要的一环。然而,实现建筑物的自动提取的过程,需要克服高分辨率遥感和无人机数据所带来的巨大挑战^[2]。这是因为建筑物常常受到不同气象变化和地质环境的影响,因此需要特定算法来消除这些复杂信息,以达到高度精准的辨识效果。其中,实例分割是应用最为广泛的一类算法。近年来,随着深度学习和人工智能领域的蓬勃发展,以卷积神经网络为基础的模型已成为广泛采用的工具,从而极大地促进了实例分割的进步^[3]。

目前国内多位学者通过各种算法来实现对目标的实例分割,其中最主流的算法有 Mask-RCNN^[4]、YOLACT^[5]、SOLOv2^[6]等。这几种算法都是针对特定应用场景所提出来的,并分别采用不同的方式进行优化。本文选取 SOLOv2 算法,用户可以根据硬件参数对轻量化网络进行剪枝固化部署,从而实现模型对所在领域的精准分割,SOLOv2 各项分割参数远胜于二阶段网络 Mask-RCNN。从近几年研究成果来看,SOLOv2 算法优化主要集中于 ResNet 骨干网络和特征金字塔两部分。比如,HE 等^[7]在 ResNet50 两个残差网络的部分分别嵌入了无参数注意力机制(SimAM)和可变形卷积(DCN),通过扩大感受野形式来增强模型泛化能力,同时增加了少量参数提升了分割精度,但是该模型仍采用了传统庞大的检测网络,同时未对模型金字塔部分进行优化。于是 Lu 等^[8]向模型嵌入 SE 注意力机制,同时引入双向特征金字塔(BiFPN),该方法专注于信道关系,自适应地重新校准信道方向的特征权值,利用全局信息有选择地强调信息元素,抑制掉有价值的元素,从而提高网络的敏感性,可是 SE 注意力机制只考虑了通道维度上的注意力,无法捕捉到空间维度上的注意力,对于通道数较少的情况可能就无法聚焦关键信息。Yu 等^[9]则引入高效通道注意力(ECA)和 DropBlock 正则化等方法,并换用了 DenseNet-169 网络来构建分割模型。模型随机隐藏了特征块,删除了特征图上的一些连续的语义信息,从而有效解决了模型

梯度消失的问题,提高了抗点蚀能力,然而 Yu 等将方法应用于边缘计算和云计算,发现规模仍较大,在嵌入式开发平台上并不容易使用。Zou 等^[10]则直接将使用轻量级 ShuffleNetv2 网络,利用信道剪枝方法对头部的冗余特征信道进行修剪,有效提高了计算资源的准确性和鲁棒性,成功实现分割网络的轻量化。以上研究均在各自领域取得了较好的分割效果。各种改进措施都在一定程度上提高了模型平均精度和检测速度。然而,ResNet-FPN 虽然可以稳定实现对目标图像的特征提取,但是其庞大的计算量在实现轻量化硬件部署方面仍面临的巨大挑战,这一点在上述研究中已经得到了充分的体现。

综合上述研究成果,本文决定选取轻量级 EfficientNetV2 为骨干网络,BiFPN 为特征金字塔网络,并通过更换标准化注意力机制(NAM)和 DropBlock 正则化以提高网络的泛化能力,从而构成遥感图像建筑物实例分割基本模型。除此之外,为了使模型更有效地学习通道注意力权值,注重对微小建筑轮廓信息的关注,在金字塔各特征层并入一种新型双水平路由注意视觉变压器(BiFormer)。实验表明:该模型提高了遥感图像建筑物的平均分割精度,整体网络参数量符合嵌入式平台部署。

2 SOLOv2 实例分割架构

SOLOv2 网络是 WANG 等^[6]于 2020 年提出的单阶段实例分割网络,由于其良好的速度和精度,非常适合遥感图像建筑物的分割,图 1 所呈现的是该网络的结构。SOLOv2 网络可被划分为三个组成部分:全卷积网络(FCN)、特征金字塔(FPN)和头部网络(掩码核分支、掩码特征分支和语义分支)。网络首先会将输入的遥感图像划分成 $S \times S$ 网格化结构,并经由 FCN 和 FPN 提取其轮廓不同层次的特征图,其中,FCN 部分使用 ResNet50 卷积的残差块来实现残差学习,FPN 部分通过上采样和特征融合逐层向下传递,生成一系列高分辨率的特征金字塔,将低分辨率的高级语义特征与高分辨率的低层次细节特征相结合,使映射到 $S \times S$ 网格上的实例中心进入头部网络^[11]。

头部网络部分,主要由语义分支和掩码分支构成。 $S \times S \times C$ 构成了语义分支的结构,其中 C 代表了该分支所

属类别的数目。掩码卷积核分支将输入的特征 F 动态卷积成 $G:S \times S \times D$ 的形式,最大程度生成 S^2 的实例掩码。掩码特征分支则用来实现对 FPN 的预测,将不同层次的

特征图进行归一化激活操作^[12]。模型的最后采用了非极大值抑制矩阵(NMS)方法。NMS 主要通过计算效率值以快速实现非极大抑制,直至提高网络的训练精度。

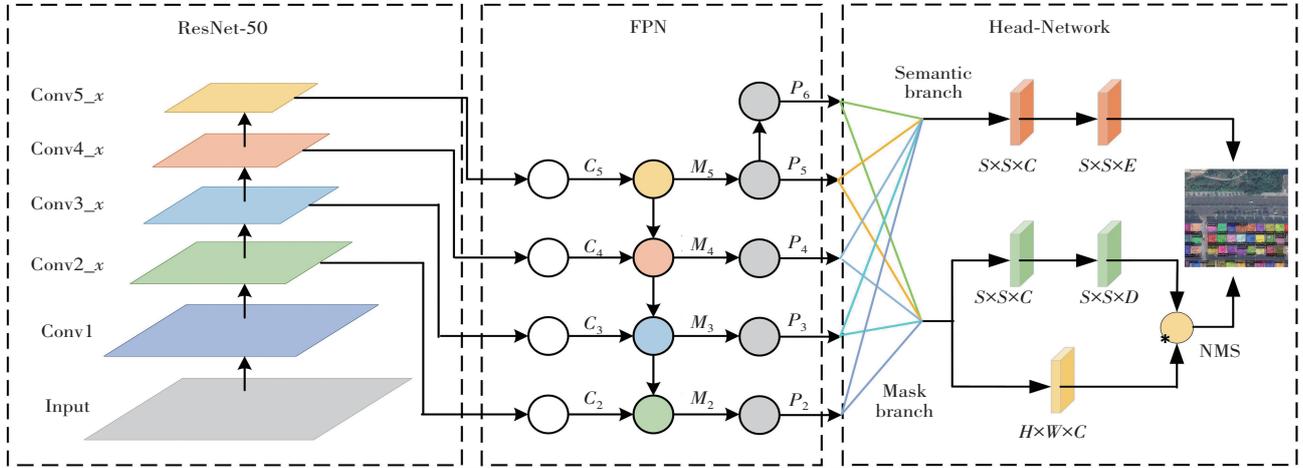


图 1 SOLOv2 实例分割传统模型结构图

Fig. 1 SOLOv2 instance segmentation traditional model structure diagram

3 遥感图像建筑物实例分割架构

3.1 改进轻量化分割网络

EfficientNet 是 Google 提出的一种新型轻量化卷积分类网络,模型具有参数效率高、训练速度快、性能优异等特点,适用应用于建筑物的实例分割。图 2 显示了 EfficientNet 的网络结构,在基础网络(Baseline)前提下,优化出的以固定比例均匀缩放所有 3 个维度的复合缩放方法。2020 年,Tan 等^[13]开发了名为 EfficientDet 的检测网络。该模型使用 ImageNet 数据集的预训练网络,以 EfficientNet 作为 Backbone,以双向特征金字塔(BiFPN)作为特征网络,从 Backbone 的第 3—7 层获取特征并且重复使用自上而下和自下而上的双向特征融合方式。

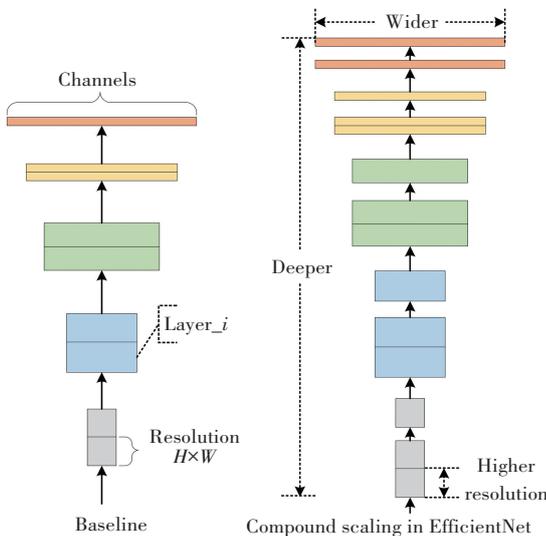


图 2 EfficientNet 网络结构

Fig. 2 EfficientNet network structure

2021 年,Tan 等又将 EfficientNet 网络升级为 EfficientNetV2,在训练速度以及参数数量上都明显优于先前的一代网络,主要修改是在其浅层卷积部分引入了一种新的卷积模块:Fused-MBConv。图 3 显示了 EfficientNetV2 的 MBConv 模块和 Fused-MBConv 模块^[14]。表 1 为 EfficientNetV2-S 网络的结构表。除了保留了 EfficientNet 中使用的复合缩放模型扩展策略外,还结合准确率、参数效率和网络训练速度设置了奖励函数,见式(1):

$$f(a, s, p) = a \times s^w \times p^v \tag{1}$$

式(1)中, a 表示模型的精度, s 表示每一步的训练时间, p 表示模型参数的大小。 w, v 是控制奖励函数程度的两个超参数,需要根据训练结果给出的经验值,本次可设 $w = -0.08, v = -0.06$ 。

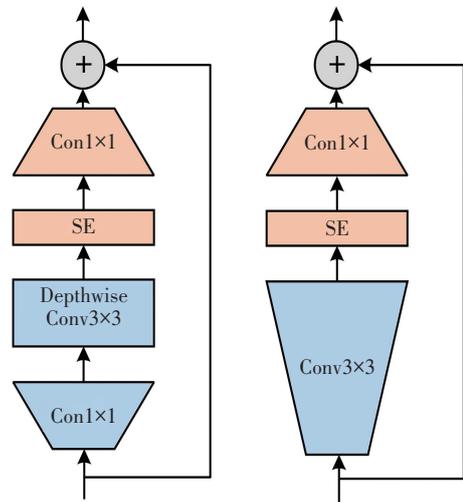


图 3 MBConv 模块和 Fused-MBConv 模块

Fig. 3 MBConv module and Fused-MBConv module

表 1 EfficientNetv2-S 网络结构表
Table 1 EfficientNetv2-S network structure table

阶 段	操作层	步 距	通道数	层 数
1	Conv3×3	2	24	1
2	Fused-MBConv1, k3×3	1	24	2
3	Fused-MBConv4, k3×3	2	48	4
4	Fused-MBConv4, k3×3	2	64	4
5	MBConv4, k3×3, SE0.25	2	128	6
6	MBConv6, k3×3, SE0.25	1	160	9
7	MBConv6, k3×3, SE0.25	2	256	15
8	Conv1×1&Pooling&FC	—	1280	1

3.2 注意力机制的优化

在训练遥感图像实例分割模型时,网络参数越多,网络中存储的特征信息就越多;然而,更大的参数并不能保证更好的结果。注意力机制的作用就是使模型关注与当前任务相关的信息,从而减少训练过程中不相关的信息。NAM 作为一种轻量级标准化的注意力机制 (BN),可以很好解决特征信息过载的问题^[15]。

NAD-MBConv 模块的构造如图 4 所示,此次将 MBConv 模块的 SE 注意力更换为 NAM 的通道注意力和空间注意力,其中两种注意力串联控制。具体操作是通过点卷积来增加输入特征的通道维数,然后通过

逐层卷积提取跨特征点的特征。这种不连续的注意机制,可以将网络的注意力集中在信息量更大的通道和空间位置上,并通过点卷积对通道进行相应调整。模块的最后加入了 DropBlock 正则化,它是 Dropout 的一个改进版本,比 Dropout 更简单地删除语义信息,其适用于整个空间块,通常随机将神经元隐藏在完全连接的层中。

随机设置一些权值为 0,防止残差连接前网络过拟合,点卷积和逐层卷积构成了深度可分卷积运算^[16]。NAM 注意力机制使用比例因子来衡量每个维度的重要性。当比例因子越大,维度数据的方差则越大。方差越大,维度变化越大,所包含的语义信息越丰富,就越值得关注。其中,标准化计算和尺寸权重计算分别见式(2)和式(3):

$$B_{out} = BN(B_{in}) = \gamma \frac{B_{in} - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} + \beta \quad (2)$$

式(2)中, B_{in} 和 B_{out} 分别为 BN 计算的输入量和输出量; μ_B 和 σ_B 分别为输入数据的均值和方差; γ 为比例因子对应的维度方差; β 为维度的平均值。

$$W_i = \frac{\gamma_i}{\sum_{i=1}^L \gamma_i} \quad (3)$$

式(3)中, L 为维度长度, γ_i 为维度 i 的比例因子, W_i 为维度 i 对应的权值。BN 计算的尺寸一般为通道尺寸。

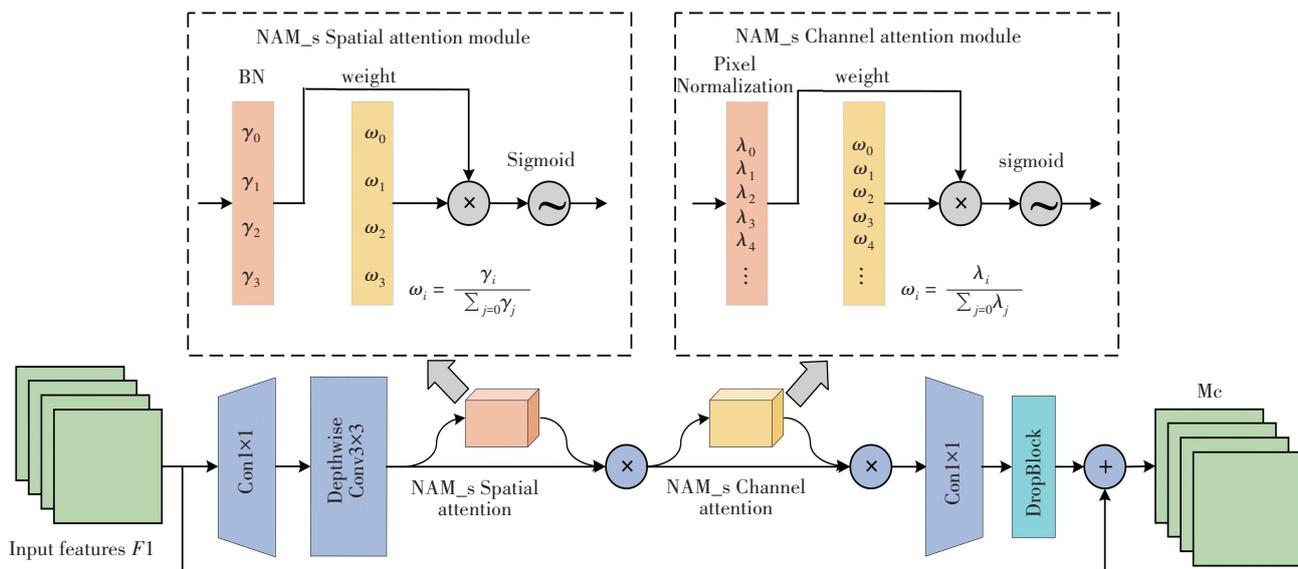


图 4 NAD-MBConv 模块的构造图

Fig. 4 Construction diagram of NAD-MBConv module

3.3 特征金字塔网络的优化

双向特征金字塔 (BiFPN) 结构如图 5 所示,它由含多个特征图的 BiFPN 模块构成,本次采用其中的 3 个模块,而每个特征层来至于不同尺度的特征图,在

PANet 的基础上,去掉了只有一个输入边的节点 (P_3 和 P_7 的中间节点),增加了左边输入节点到右边输出节点的直接跳连,从而减少计算量,更有利于对目标特征进行融合。

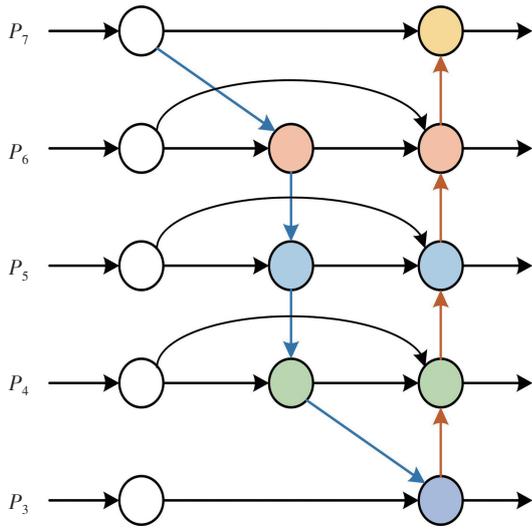


图 5 BiFPN 结构示意图

Fig. 5 Schematic diagram of BiFPN structure

为了特征融合更加注重细小轮廓的能力, ZHU 等^[17]的利用稀疏性来节省计算和内存的思想, 提出的双水平路由注意视觉变压器——BiFormer, 其中双级路由注意机制(BRA)是整个变压器的核心部分, 其整体构造如图 6 所示。BRA 运用微小目标的感知特征和上下文相互关联来增加相关信息, 并利用更广泛场景的上下文信息来推断轮廓位置或类别。该机制主要由区域划分、输入投影、有向图区域到区域路由以及令牌注意等部分构成。可以根据给定的特定目标而动态地调整参数, 其核心思想在于过滤掉那些并不重要的键值对, 以实现细节和稀疏性的高度关注^[18]。

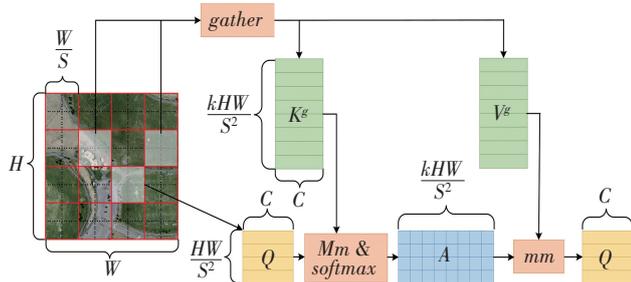


图 6 BRA 整体构造图

Fig. 6 Overall structure of BRA

针对给定的二维特征图 $X \in \mathbf{R}^{H \times W \times C}$ (H, W, C 分别为输入图像的高度、宽度和通道数), 其注意机制分别线性投影到查询向量 $Q \in \mathbf{R}^{N_q \times C}$, 键向量 $K \in \mathbf{R}^{N_k \times C}$ 和值向量 $V \in \mathbf{R}^{N_v \times C}$ 上, 则注意力机制的输入公式见式(4):

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{D}}\right)V \quad (4)$$

式(4)中, \sqrt{D} 为标量因子。而 Softmax 函数会将输入图

像映射到 $(0, 1)$ 空间中。

对于输入特征映射 $X \in \mathbf{R}^{H \times W \times C}$, 将其初始划分为 $S \times S$ 个不重叠的方形区域, 每个区域包含 $\frac{HW}{S^2}$ 个特征向量, 则所得到的查询向量、键向量和值向量的表达式见式(5):

$$Q = X'W^q, K = X'W^k, V = X'W^v \quad (5)$$

式(5)中, Q, K, V 和 $X \in \mathbf{R}^{\frac{S^2 \times HW}{S^2} \times C}$; 而每个线性投影所对应的权值 W^q, W^k 和 $W^v \in \mathbf{R}^{C \times C}$ 。

第二部分“带有有向图的区域到区域的路由”中, 对输入特征映射 X 的圈定区域进行加权, 通过构造有向图来定位给定区域的相关区域, 从而获得区域之间的注意力关系^[19]。首先, 分别计算每个划分区域的 Q 和 K 的平均值, 得到 Q' 和 $K' \in \mathbf{R}^{S^2 \times C}$ 。然后, 计算区域间语义相关性的邻接矩阵 A' , 则 A' 见的表达式见式(6):

$$A' = Q'(K')^T \quad (6)$$

同时, 为了减少每个区域与其他区域的交互次数, 如式(7)所示, BRA 会通过索引矩阵 $I' \in \mathbf{N}^{S^2 \times k}$ 保留每个区域最相关的 k 个查询区域。

$$I' = \text{topkIndex}(A') \quad (7)$$

第三部分“令牌对令牌注意力”中, 如式(8)所示, 需要将这些分散在整个特征映射的键向量和值向量集成到 GPU 操作中。

$$K^g = g(K, I'), V^g = g(V, I') \quad (8)$$

式(8)中, g 为收集张量的操作, $K^g, V^g \in \mathbf{R}^{\frac{S^2 \times HW}{S^2} \times C}$ 。则最终形成的双级路由注意机制(BRA)可表示为式(9)的形式:

$$\text{BRA} = \text{Attention}(Q, K^g, V^g) + \text{LCE}(V) \quad (9)$$

式(9)中, LCE 是深度卷积网络在 V 上对多尺度令牌聚合的局部增强运算。

改进的 Bi-FPN-Former 网络结构图如图 7 所示。针对 BiFPN 的尾端进行了改进, 各特征层并入一个双水平路由注意视觉变压器(BiFormer), 在 SOLOv2 模型的掩码特征分支对 BiFPN 特征进行预测时, 所预测出的锚框主要需要依靠 NMS(非极大值抑制)来过滤掉大量容易遗漏微小目标的低置信边界, 然而 NMS 需要合适的阈值, 否则对目标抑制效果不明显。BiFormer 的自我注意机制的主要目的就是降低对外部信息依赖程度, 尽量使用原特征信息编码不同地点关联信息以达到对小目标的注意力集中。

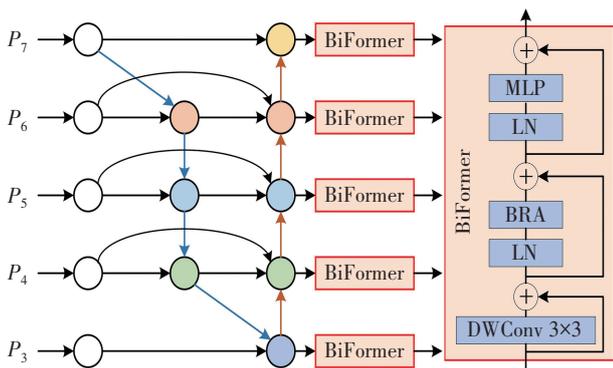


图 7 Bi-FPN-Former 网络结构图

Fig. 7 Bi-FPN-Former network structure diagram

4 NDFEDet-SOLOv2 模型的构建

为了高效、准确地分析出遥感图像建筑物的轮廓特征,所提出基于 NDFEDet-SOLOv2 实例分割算法的遥感图像建筑物轮廓自动提取方法,总结起来,主要有以下改进措施:将传统 SOLOv2 网络 ResNet50-FPN 改

为采用双向特征金字塔 (BiFPN) 特征融合方式的轻量化 EfficientDet 网络,其中将骨干网络部分的 EfficientNet 升级为 EfficientNetV2,分别由 M_2 、 M_3 、 M_4 、 M_6 和 M_7 作为特征层进行输出;考虑存储信息过载、输入信息通道数较少等问题,将 EfficientNetV2 中三层 MBConv 模块的 SE 注意力更换为含有 DropBlock 正则化的轻量级标准化注意力机制 (NAM),其中 NAM 的空间和通道注意力以串联的方式进行插入。在特征重新校准的帮助下,NAM 可以学习利用全局信息有选择地强调信息元素,抑制有价值的元素,从而提高网络的敏感性和整体的泛化能力;在模型颈部部分,向 BiFPN 尾部各特征层分别嵌入一种双水平路由注意视觉变压器 (BiFormer),从而聚焦微小建筑物轮廓信息,以实现更高层次的特征融合,进一步提高平均精度,同时节省计算和内存以提高 GPU 的工作效率。改进的 NDFEDet-SOLOv2 模型主要架构如图 8 所示。

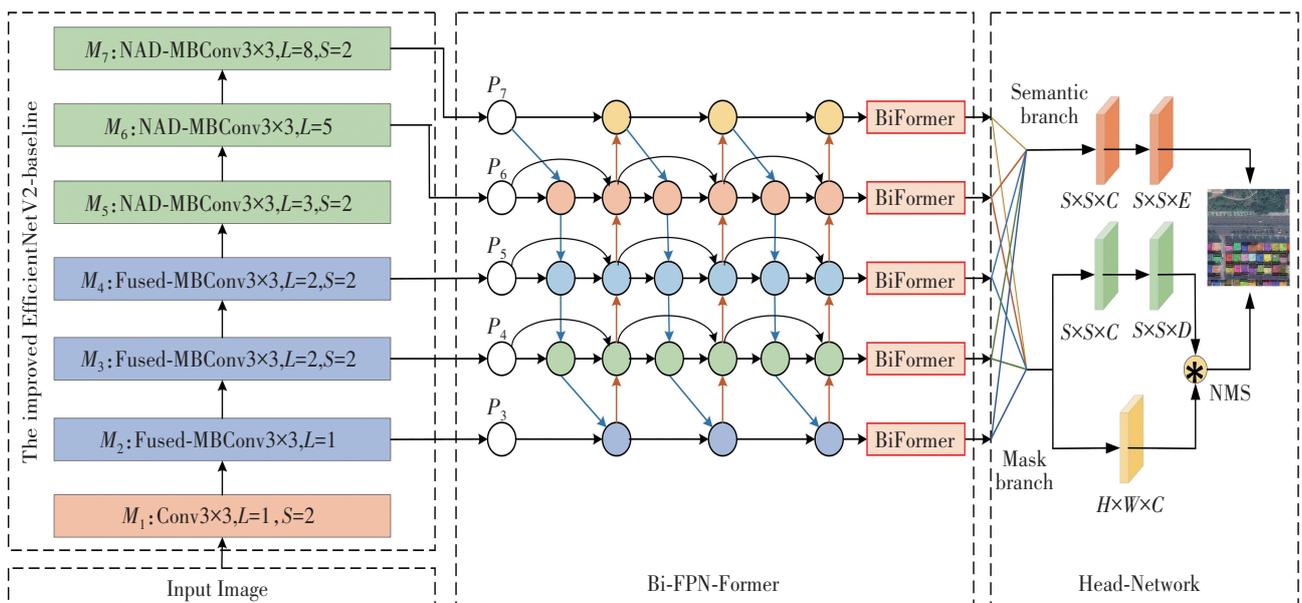


图 8 轻量化 NDFEDet-SOLOv2 模型结构

Fig. 8 Lightweight NDFEDet-SOLOv2 model structure

5 仿真实验与结果分析

5.1 实例分割数据集

实验采用开源 WHU 数据集,部分图例如图 9 所示。此数据集源于中国地质大学(武汉)方芳团队采用人工交互式标注的方式构建的城市建筑物实例数据集。其中包含中国主要城市的房屋密集区域和稀疏区域,同时还存在绿色植被、道路、车辆和运动场等复杂因素,共 7 260 张图片,63 886 栋建筑物,像素大小为 500×500 ,分辨率为 0.29 m。其中,将遥感图像中的建筑物按照 7 : 3 的比例分为训练集和测试集,并以标准的 MS COCO 2017 的数据格式导出。

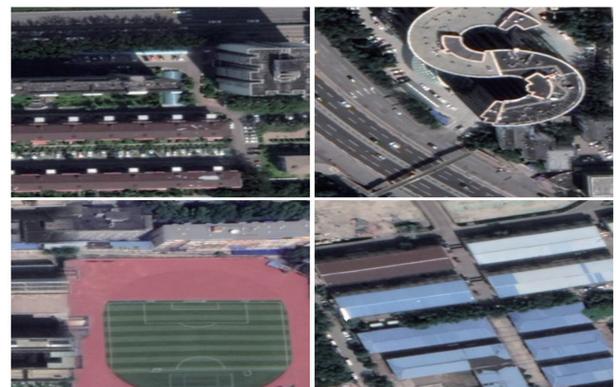


图 9 中国典型城市建筑物实例分割数据集

Fig. 9 Typical urban building example dataset in China

5.2 模型训练设置

本文使用 Windows 10 的台式计算机作为实例分割模型的训练平台,其中 CPU 处理器的型号为 Inter i5-12400F(16 GB),GPU 显卡的型号为英伟达 GeForce RTX 3050(8 GB)。采用 Python3.7,Pytorch1.8.0,Cuda11.1 和 Mmdetection2.28.2 框架建立遥感图像实例分割模型。由于 Batch_size 的数值取决于 GPU 数量。而主机只有一个 GPU,则设置模型每次接受两张图片,即 Batch_size 为 2,训练一个 batch 为一次 iteration,对于全部数据集,进行一次完整的 epoch 训练,总共训练了 100 个 epoch。

5.3 损失函数

实例分割可以看成是目标检测和语义分割的结合体,不仅要精准地识别目标位置,还要对每个目标的像素进行精确分割。SOLOv2 模型的损失函数 L 为 L_{mask} 与 L_{cls} 的总和, L 公式见式(10):

$$L = L_{\text{cls}} + \lambda L_{\text{mask}} \quad (10)$$

式(10)中, L_{mask} 表示实例掩模分割的损失, L_{cls} 表示实例类别预测的损失, λ 为超参数,的此可设 $\lambda = 3$ 。 L_{cls} 表达式见式(11):

$$L_{\text{cls}} = -\alpha_i (1-p_i)^\gamma \log(p_i) \quad (11)$$

式(11)中, α 为类别权重,表示难区分样本权重,设置 $\alpha_i = 0.25$, $\gamma = 2$, p 表示类别的预测概率^[20]。则 L_{mask} 和 d_{mask} 表达式分别见式(12)和式(13):

$$L_{\text{mask}} = \frac{1}{N_{\text{pos}}} \sum_{k=1}^{N_{\text{pos}}} \mathbb{I}\{p_{i,j}^* > 0\} d_{\text{mask}}(m_k, m_k^*) \quad (12)$$

$$d_{\text{mask}}(p, q) = 1 - \frac{2 \sum_{x,y} (p_{x,y} q_{x,y})}{\sum_{x,y} p_{x,y}^2 + \sum_{x,y} q_{x,y}^2} \quad (13)$$

5.4 评价指标

平均精度(mAP)作为实例分割任务性能评估中常用的度量指标,是衡量模型精确度的重要标准。计算时涉及交并比 IoU,精确度 P 和召回率 R 等。相应的评价指标计算公式见式(14)一式(17),IoU 是预测掩码区域和真实掩码区域交集和并集之间的比。其中,采用 mAP_{50} 与 mAP_{75} 代表阈值分别设定在 0.5 与 0.75 的平均精度,而 mAP_L , mAP_M 与 mAP_S 则分别代表了在大,中,小尺度目标的平均精度。这些指标能够在不同的尺度上对模型进行全面的评价。

$$P = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}} \times 100\% \quad (14)$$

$$R = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}} \times 100\% \quad (15)$$

$$f_{\text{AP}} = \int_0^1 P(R) dR \quad (16)$$

$$f_{\text{mAP}} = \frac{1}{N} \sum_{i=1}^N i_{\text{AP}} \quad (17)$$

5.5 结果分析

为验证改进模型有效性和准确性,选择在中国经典城市建筑数据集上进行模型结果分析。图 10 和图 11 分别表示的是改进模型的损失函数变化图和平均精度变化图。损失函数会随着迭代次数的增加而减少,当模型迭代达到 5720 后,损失曲线会趋于平缓。经过多次训练验证,模型损失函数最小可达到 0.48,而 L_{mask} 和 L_{cls} 最低值分别为 3.52% 和 2.06%。至于平均精度(mAP)值在 epoch 超过 33 时才最终稳定下来, mAP_{50} 、 mAP_{75} 和 mAP 的最大值分别为 83.5%、57.9% 和 53.8%。由于优化模型引入了 EfficientDet 检测网络使得模型对建筑物分割具有较强的适应性,NAM 和 BiFormer 的加入,也使得改进模型有了较好的提取效果。

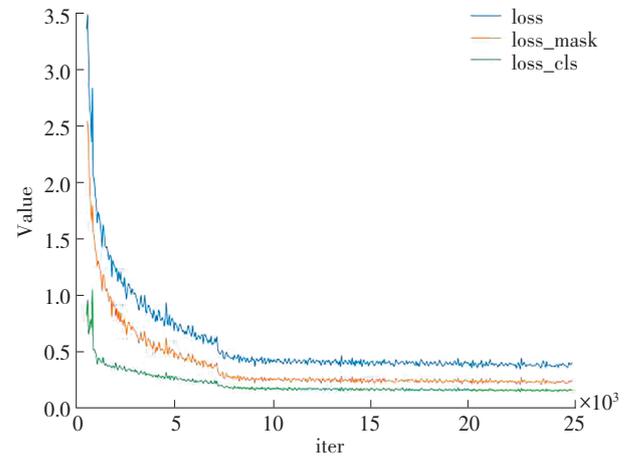


图 10 NDFEDet-SOLOv2 损失函数曲线变化图

Fig. 10 NDFEDet-SOLOv2 loss curve change chart

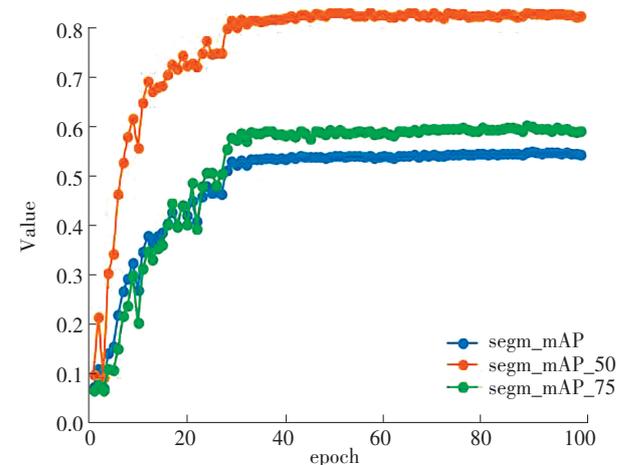


图 11 NDFEDet-SOLOv2 平均精度曲线变化图

Fig. 11 NDFEDet-SOLOv2 mAP curve variation diagram

表 2 显示不同网络组合进行融合实验的评价指标对比。其中以传统的轻量级 SOLOv2 模型作为基准。结果可以看出,在更换 NAM 后,模型的 mAP 增加了 1.4%,特别是在高阈值区域,精度提升较为明显。

BiFPN 单独加入 BiFormer 视觉变压器, mAP 增加了 2.4%, 检测帧率相应的达到了 7.1 帧/s。最终优化后

的模型的 mAP 达到 53.8%, mAP₅₀ 高达 83.5%, 符合融合实验的基本要求。

表 2 改进实例分割模型消融实验评价指标对比

Table 2 Comparison of evaluation indicators for improved instance segmentation model ablation experiments

模 型	主干网络	NAM	BiFormer	mAP/%	mAP ₅₀ /%	mAP ₇₅ /%	FPS
SOLOv2(Light)	ResNet50-FPN	—	—	48.9	79.8	55.4	4.7
SOLOv2(improve1)	EfficientNetV2-BiFPN	—	—	49.3	79.1	54.8	5.6
SOLOv2(improve2)	EfficientNetV2-BiFPN	✓	—	50.7	79.8	55.9	6.4
SOLOv2(improve3)	EfficientNetV2-BiFPN	—	✓	51.4	81.3	56.3	7.1
NDFEDet-SOLOv2	EfficientNetV2-BiFPN	✓	✓	53.8	83.5	57.9	7.4

图 12 显示的是在不同场景下, 遥感建筑物实例分割对比图。从图 12 中可以看出, 传统算法虽然可以训练出模型的基本轮廓, 但是训练结果发现, 部分建筑出现了边角错误分割的问题, 有的边角甚至完全脱离了建筑俯视面的真实轮廓。改进后的实例分割模型就可

以很好的解决这部分建筑分割的局部失真问题, 所训练的图像结果基本符合 Mask 标记视图的图像轮廓信息。综合各方面数据来看, 改进网络的提取效果明显更好; 漏检率降低, 黏附问题得到改善, 边缘分割也比较完整, 整体精度更高。

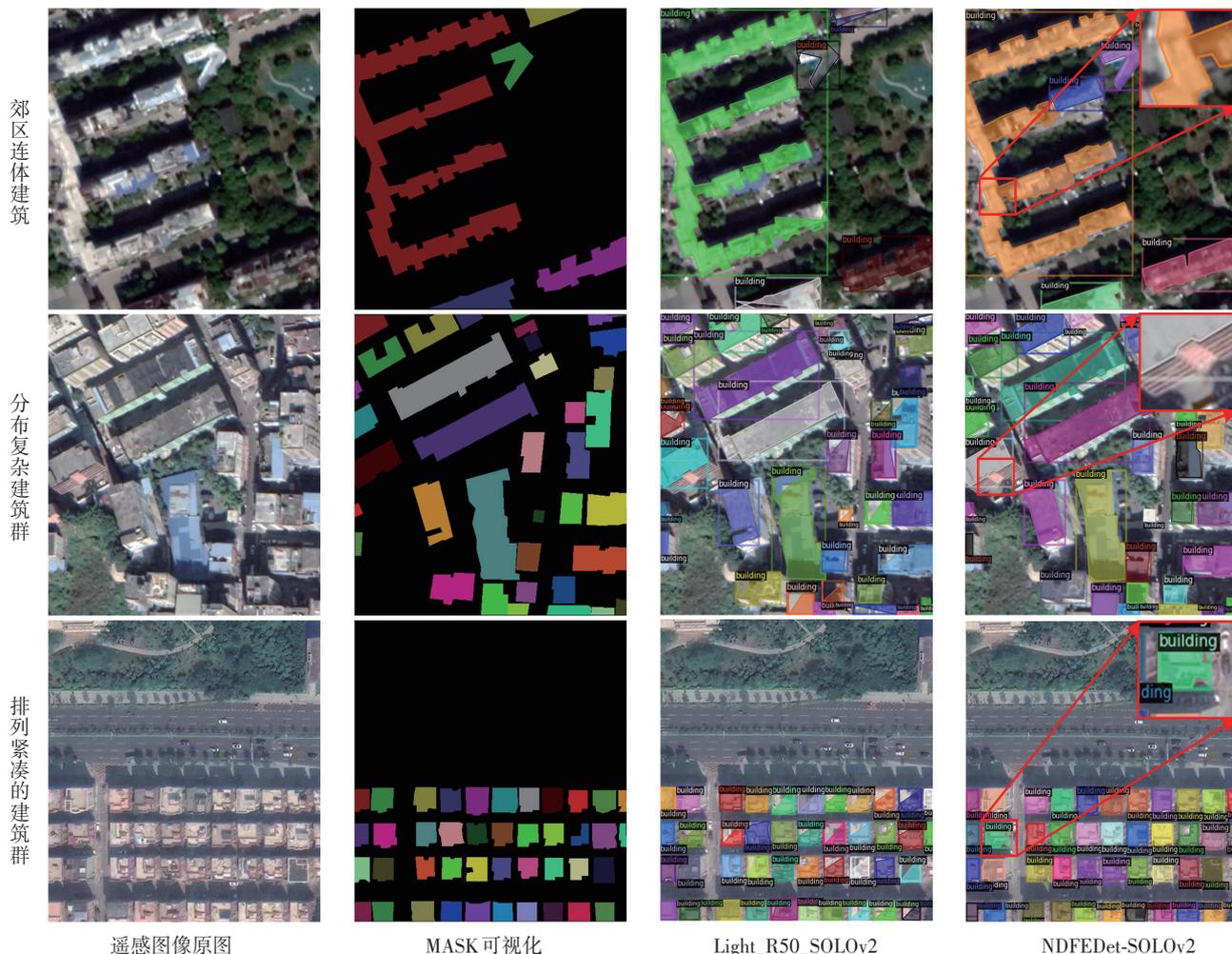


图 12 传统 SOLOv2 与改进模型实例分割模型效果

Fig. 12 Traditional SOLOv2 and improved model instance segmentation model performance

5.6 对比实验

为了更加凸显出改进方案的最佳性,此次针对 NAM 注意力在 EfficientNetv2 骨干网络中不同更换位置作了对比试验,各参数对比见表 3。其中,分别以 EfficientNetV2 网络的 MBCConv 模块、Fused-MBCConv 模块和两个卷积模块共同更换 NAM 注意力作为分组。最终实验表明,NAM 更换在 3 个不同层次的 Fused-MBCConv 位置时,整体精度虽较传统模型有所提高,但是与其他两个方案相比,尤其在小目标分割平均精度

上,效果并不明显。而将两种卷积模块同时更换 NAM 注意力机制时,对大、中、小目标的分割精度都有了提高,但是发现整个网络的参数量也随之发生了显著变化,高达 31.3 M,这对轻量化模型部署是一个大问题。同时加入 3 块不同层次 MBCConv 模块中不仅提升了对建筑物整体分割精度,网络参数基本也基本与加入 Fused-MBCConv 模块的情况下持平,所以最终决定通过将 NAM 更换在 3 个不同层次 MBCConv 模块的位置来优化实例分割模型。

表 3 NAM 注意力更换位置各参数对比表

Table 3 Comparison of NAM attention different replacement position parameters

NAM 添加位置	mAP/%	mAP ₅₀ /%	mAP ₇₅ /%	mAP _L /%	mAP _M /%	mAP _S /%	FPS	Param/M
SOLOv2(EfficientNetV2)	49.3	79.1	54.8	51.3	49.1	43.1	5.6	24.4
Fused-MBCConv	52.1	81.4	56.1	52.3	53.6	43.2	8.3	26.2
MBCConv+Fused-MBCConv	56.1	83.1	58.6	56.3	55.1	46.3	9.4	31.3
NDFEDet-SOLOv2	53.8	83.5	57.9	55.1	54.8	44.7	7.5	26.7

6 结 论

改进的遥感图像建筑物轮廓轻量化提取模型——NDFEDet-SOLOv2,在传统 Light-SOLOv2 算法的基础上,将 ResNet50-FPN 网络换用升级的 EfficientDet (EfficientNetV2-BiFPN) 网络,并分别向网络的骨干网络和颈部网络加入了标准化注意力机制(NAM)和双水平路由注意视觉变压器(BiFormer)。经过训练,NDFEDet-SOLOv2 将评价指标 mAP、mAP₅₀ 和 mAP₇₅ 分别提高了 4.9%、3.7% 和 2.5%,检测帧率 FPS 也提高了 2.7 帧/s,尤其对一些大、中目标的分割,提升效果明显。从各方面的数据来看,该优化方案基本消除了建筑物边角的图像畸变,尤其在各地地理环境空间分布不均等复杂情况下也能较为准确地提取出遥感图像建筑物的基本轮廓。

同时,为满足现实生活的需要,该模型经剪枝固化后期会通过 Xilinx 深度学习 FPGA 开发板进行轻量化部署。但是,本文改进的网络在提取密集和小规模的建筑群方面的平均精度仍需要提高。在日后的研究中,将不断对模型的结构(如激活函数和 NMS 等)进行

局部调整和优化,以致达到更好的提取效果,从而为城市布局更新和建筑变化检测提供理论参考。

参考文献(References):

- [1] JIANG K, WANG Z, YI P, et al. Edge-enhanced GAN for remote sensing image superresolution[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(8): 5799-5812.
- [2] SHI F, ZHANG T. An anchor-free network with box refinement and saliency supplement for instance segmentation in remote sensing images[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19(10): 21-25.
- [3] 丁飞, 石颀, 吴宏杰. 改进 YOLOv4 的轻量级遥感图像建筑物检测模型[J]. 计算机工程与应用, 2023, 59(10): 213-220.
DING Fei, SHI Jie, WU Hong-jie. Light weight building detection model based on YOLOv4 optimization for remote sensing images[J]. Computer Engineering and Applications, 2023, 59(10): 213-220.
- [4] CHEN S, OGAWA Y, ZHAO C, et al. Large-scale individual building extraction from open-source satellite imagery via

- super-resolution-based instance segmentation approach[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2023, 195: 129–152.
- [5] BOLYA D, ZHOU C, XIAO F, et al. YOLACT: real-time instance segmentation[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. Piscataway: IEEE Press, 2019: 9156–9165.
- [6] WANG X, ZHANG R, SHEN C, et al. SOLO: a simple framework for instance segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(11): 8587–8601.
- [7] 何佳琦, 周思艺, 唐晓萌, 等. 基于改进的轻量版 SOLOv2 红鳍东方鲀实例分割方法[J]. *华中农业大学学报*, 2023, 42(3): 71–79.
- HE Jia-qi, ZHOU Si-yi, TANG Xiao-meng, et al. Instance segmentation method of Takifugu rubripes based on improved light SOLOv2[J]. *Journal of Huazhong Agricultural University*, 2023, 42(3): 71–79.
- [8] LU Y, WEI W, LI P, et al. A deep learning method for building facade parsing utilizing improved SOLOv2 instance segmentation[J]. *Energy and Buildings*, 2023, 295: 113275.
- [9] YU Z, LIU L, JIAO H, et al. Leveraging SOLOv2 model to detect heat stress of poultry in complex environments[J]. *Frontiers in Veterinary Science*, 2023, 9(1): 10–23.
- [10] ZOU Y, ZENG G. Light-weight segmentation network based on SOLOv2 for weld seam feature extraction[J]. *Measurement*, 2023, 208(28): 263–276.
- [11] 吴忧, 袁雪. 基于改进 SOLOv2 的复杂场景下智能机器人巡检识别算法[J]. *北京交通大学学报*, 2022, 46(5): 95–106.
- WU You, YUAN Xue. Inspection and identification algorithm based on improved SOLOv2 of intelligent robot in complex environment[J]. *Journal of Beijing Jiaotong University*, 2022, 46(5): 95–106.
- [12] KOH E J Y, AMINI E, MCLACHLAN G J, et al. Utilising convolutional neural networks to perform fast automated modal mineralogy analysis for thin-section optical microscopy[J]. *Minerals Engineering*, 2021, 173(11): 1–11.
- [13] TAN M, PANG R, LE Q V. EfficientDet: scalable and efficient object detection[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE Press, 2020: 10778–10787.
- [14] LIU L, FANG J, HUANG H, et al. Mask defect detection algorithm based on improved EfficientNetV2[C]//*Proceedings of the International Symposium on Advances in Informatics, Electronics and Education*. Piscataway: IEEE Press, 2022: 507–510.
- [15] WANG R, GAO R, LI Q, et al. Pig face recognition based on metric learning by combining a residual network and attention mechanism[J]. *Agriculture*, 2023, 13(1): 144–151.
- [16] LIU Q, LIU W, LIU Y, et al. Rice grains and grain impurity segmentation method based on a deep learning algorithm-NAM-EfficientNetV2[J]. *Computers and Electronics in Agriculture*, 2023, 209: 7824–7834.
- [17] ZHU L, WANG X, KE Z, et al. BiFormer: vision transformer with bi-level routing attention[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE Press, 2023: 10323–10333.
- [18] GAO X, TANG Z, DENG Y, et al. HSSNet: a end-to-end network for detecting tiny targets of apple leaf diseases in complex backgrounds[J]. *Plants*, 2023, 12(15): 2806.
- [19] WANG G, CHEN Y, AN P, et al. UAV-YOLOv8: a small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios[J]. *Sensors*, 2023, 23(16): 7190–7217.
- [20] LUO X, ZENG W, CHEN Y, et al. SOLOv2-cable: a power cable segmentation algorithm in complex scenarios[C]//*Proceedings of the 3rd International Conference on Computer Vision, Image and Deep Learning & International Conference on Computer Engineering and Applications*. Piscataway: IEEE Press, 2022: 1156–1161.

责任编辑:陈芳