

DOI: 10.11686/cyxb2024060

http://cyxb.magtech.com.cn

闫储淇, 黄建强. 基于YOLOv5和改进DeeplabV3+的青藏高原植被提取算法. 草业学报, 2025, 34(1): 41-54.

YAN Chu-qi, HUANG Jian-qiang. Vegetation extraction algorithm for the Tibetan Plateau based on YOLOv5 and improved DeeplabV3+. Acta Prataculturae Sinica, 2025, 34(1): 41-54.

基于YOLOv5和改进DeeplabV3+的 青藏高原植被提取算法

闫储淇, 黄建强*

(青海大学计算机技术与应用学院, 青海省智能计算与应用实验室, 青海 西宁 810016)

摘要: 青藏高原的植被覆盖度是生态研究和环境监测的重要指标。传统的植被覆盖度检测方法在地形简单且植被分布集中的区域效果较好,但在复杂地形下由于成本高、调查范围受限、耗时长等问题,导致植被提取精度受限。近年来,计算机视觉和深度学习技术的飞速发展,为青藏高原复杂地形下的植被精准提取开辟了新的可能性。本研究提出一种结合YOLOv5和改进DeeplabV3+的双阶段植被提取算法。算法引入基于YOLOv5的植被目标检测模型,以减少背景对第二阶段植被分割任务的干扰;设计新型的DeeplabV3+语义分割模型,以实现精准的植被分割提取。改进的模型引入了轻量级主干网络MobileNetV2,优化了ASPP模块膨胀卷积参数,并集成EMA和CloAttention注意力机制。在青藏高原无人机航拍数据集上的实验结果显示,本算法在交并比(IoU)和像素准确率(PA)上分别达到了90.40%和96.32%,显著超过现有技术,且大幅降低了模型参数。本算法在多种环境条件下均展示了高精度的植被提取能力,可以为青藏高原植被覆盖度的快速、精准测定提供有效的技术支持。

关键词: 青藏高原; 植被提取; 深度学习; YOLOv5; DeeplabV3+

Vegetation extraction algorithm for the Tibetan Plateau based on YOLOv5 and improved DeeplabV3+

YAN Chu-qi, HUANG Jian-qiang*

College of Computer Technology and Application, Qinghai University, Intelligent Computing and Application Laboratory of Qinghai Province, Xining 810016, China

Abstract: Vegetation coverage on the Qinghai-Tibet Plateau is a crucial metric for ecological studies and environmental monitoring. Traditional methods to detect vegetation coverage are effective in regions with simple terrains and concentrated vegetation. However, in complex terrains, issues such as high costs, restricted survey areas, and extended time intervals reduce the accuracy of the results obtained using such traditional methods. In recent years, rapid advancements in computer vision and deep learning have created new opportunities for precise vegetation extraction in the complex terrains of the Qinghai-Tibet Plateau. Here, we introduce a two-stage vegetation extraction algorithm that integrates YOLOv5 and an improved DeeplabV3+. The algorithm utilizes a vegetation detection model based on YOLOv5 to minimize background interference during the second stage of vegetation segmentation; and a newly designed DeeplabV3+ semantic segmentation model for accurate vegetation

收稿日期: 2024-02-27; 改回日期: 2024-05-16

基金项目: 青海省重点研发计划: 地球系统模式公共软件平台在青藏高原气候诊断评估的应用与推广(2023-QY-208)资助。

作者简介: 闫储淇(2002-), 女, 天津人, 在读本科。E-mail: 210809010426@qhu.edu.cn

* 通信作者 Corresponding author. E-mail: hjqxaly@163.com

segmentation and extraction. The improved model incorporates the lightweight backbone network MobileNetV2, optimizes the dilated convolution parameters of the ASPP module, and integrates EMA and CloAttention mechanisms. The experimental results on the unmanned aerial vehicle dataset of the Qinghai-Tibet Plateau demonstrate that the algorithm attains an intersection over union (IoU) of 90.40% and a pixel accuracy (PA) of 96.32%, significantly outperforming other current technologies and greatly reducing the model's parameters. Under various environmental conditions, the algorithm exhibits high-precision capabilities for vegetation extraction, offering effective technical support for the rapid and precise measurement of vegetation cover on the Qinghai-Tibet Plateau.

Key words: Tibetan Plateau; vegetation extraction; deep learning; YOLOv5; DeeplabV3+

青藏高原以其独特的地理环境和丰富的生态系统而闻名于世。草地构成了青藏高原最广阔的生态系统,在青藏高原的土地覆盖中,草地超过60%,占中国草地总面积约1/3^[1]。植被是草地生态系统的核心部分,构成了生态系统的根基,对陆地表层的能量交换、生物地球化学反应以及水文循环过程具有不可或缺的作用。草地植被的覆盖程度不仅反映了草原的生长状况,而且是评估生态系统健康状况的关键指标。在遭受荒漠化威胁的干旱及半干旱区域,草地植被覆盖度能敏锐地指示土地退化和沙漠化程度,对维持区域生态稳定和应对环境变化起着至关重要的作用。青藏高原地形复杂多样,涵盖高寒草甸、高寒草原、高寒荒漠草原以及高山冻原等多种草地生态类型^[2],使得植被的精确提取和分析成为生态研究和环境保护领域的一大挑战。因此,针对青藏高原特有的草地生态系统,开发快速、精确、成本效益高的新型植被提取技术,以提高植被覆盖度监测的准确性,不仅对于该地区的生态学研究 and 环境保护具有重大意义,也将为其可持续发展奠定坚实的科学基础。

传统的植被覆盖度检测主要依赖人工现场调查,此方法在地形简单、植被分布较少的区域能够取得较好的效果。然而,人工调查受地理环境限制较大,面临诸如调查范围局限、成本高昂、时间消耗长以及调查结果易受调查人员经验影响等问题。目估法依赖于观测者直观感受和个人经验来评估样方中的植被覆盖度。然而,这种方法具有很大的主观性,其准确度依赖于测量者的经验,误差率可能高达40%^[3]。点框架法、样点法和样线法通过比较样方中的小孔、点或线所占比例与总数来计算植被覆盖度^[4]。其中,点框架法不仅精度较低,而且耗时^[5],而样线法则要求操作者具备相应的专业知识^[6],且仅适用于特定的草地类型^[5,7]。空间定量计法(spatial quantitative squares, SQS)和移动光量计法(translational quantitative squares, TQS)需配备专用的传感器设备,这些设备价格昂贵且在户外操作不便^[4]。照相法则要求操作人员掌握图像分析的专业能力,以便从图像中准确提取所需数据。鉴于青藏高原地形的复杂性及其生态系统的多样性,这些方法往往无法满足高效且精确的植被信息提取需求。

自20世纪60年代以来,随着遥感技术的快速发展,利用卫星和航空遥感技术进行植被信息提取成为主流^[8],该技术显著提高了调查的效率和准确性。众多国内外学者投身于探索多源遥感数据融合技术在植被提取中的应用,广泛使用的技术包括主成分分析(principal component analysis, PCA)^[9]、小波变换(wavelet transform, WT)^[10]、高通滤波(high pass filter, HPF)^[11]等。这些多源遥感数据融合技术有效克服了依靠单一数据源和分类技术时存在的局限性,提高了数据的丰富性和分类的准确性。尽管多源遥感技术在植被信息提取方面取得了显著进步,但卫星遥感存在分辨率较低、易受云层遮挡、重访周期长以及成本较高等局限,这些因素在一定程度上阻碍了植被提取精度的提升,并限制了其实际应用中的普适性和实用性^[12]。

近年来,随着深度学习的迅速发展,目标检测技术在图像信息识别任务中得到了广泛应用。该技术依赖于计算机视觉和算法,精确地识别和定位图像中的特定对象。相较于传统的手工图像处理方法,它能够学习图像的低级至高级特征,展现出更高的检测精度和泛化能力。在众多目标检测算法中,YOLO框架^[13]因其速度与精度的平衡而著称。此外,语义分割作为图像识别提取的重要技术,通过对每个像素进行密集预测和标注,实现深度的像素级推理,确保每个像素都能被准确地分类。网络模型如AlexNet^[14]、DeepLab^[15-18]、GoogleNet^[19]、U-Net^[20]、ResNet^[21]、PSPnet^[22]及HRnet^[23]等,在多个领域均展示了优异的应用性能。当前,众多国内外学者基于深度学习技术,在植被提取领域开展了广泛研究。例如,周欣昕等^[24]针对传统植被提取方法在跨季节时存在的不完整问

题,提出了一种采用特征分离机制的改进版Desenet语义分割网络,其交并比指标达到了85.10%,实现了多源遥感数据下的高精度植被提取。然而,由于多源遥感数据的光谱和空间分辨率限制,该方法在城市区域高层建筑阴影下的植被提取仍面临挑战。Zhang等^[25]结合DeepLabv3+模型和双注意力机制CBAM提出的轻量级冬小麦(*Triticum aestivum*)种植面积提取模型,能够快速准确地提取冬小麦种植位点。da Silva Mendes等^[26]对DeepLabv3+模型和YOLOv5模型在植被目标检测方面进行了对比研究,但对二者结合应用的深入探讨尚缺乏。尽管利用深度学习技术进行植被提取的研究已取得显著进展,但当前的方法大多仅依赖于单一的语义分割技术,未能充分发掘结合目标检测和语义分割的综合应用潜力。

基于深度学习技术的植被提取研究在城市与农业用地领域已取得显著进展,但在处理青藏高原等地形复杂地区的任务时,现有研究多依赖于传统遥感技术或基本的机器学习技术。例如,胡宜娜等^[27]利用无人机高光谱数据,采用空间-光谱特征识别方法,实现了基于特征挖掘的稀疏表示分类方法的草种识别,总体精度提高了4%~5%,但此方法在广泛的草地区域应用上仍有局限性;张宇鹏等^[28]基于遥感信息技术对黄河源青海片区的退化草地进行了识别,采用计算机自动分类与人机交互目视解译的方式,对不同类型退化草地展开内业判读和解译工作;文铜等^[29]则运用随机森林(random forest, RF)、支持向量机(support vector machine, SVM)、K最近邻算法(K-nearest neighbors, KNN)3种机器学习算法构建了高寒草地植物光谱识别模型,虽然模型精确度普遍超过85%,但在特殊情况下依然存在误判问题。这些研究成果为青藏高原的植被提取提供了一定的技术支持,但依靠深度学习技术进行该地区植被提取的研究还相对较少。当前深度学习技术在多个领域已广泛应用并取得显著成效,结合目标检测和语义分割等先进的深度学习技术,将为青藏高原复杂地形下的植被提取研究开辟新的可能性。

针对传统植被覆盖度检测技术的局限性以及深度学习在青藏高原植被提取研究中的不足,本研究提出了一个结合YOLOv5目标检测和改进DeeplabV3+语义分割技术的植被提取算法,以充分发挥多模型结合的优势。此外,针对多源遥感数据在植被识别方面的不足,本研究采用无人机航拍数据进行研究。算法引入YOLOv5的高效目标检测功能准确定位植被区域,以减少背景对第二阶段植被分割任务的干扰;设计新型的DeeplabV3+模型进行青藏高原植被的精细分割提取。针对青藏高原这种地形复杂且对植被提取速度和精确度要求高的任务,本研究对DeeplabV3+模型进行了4项关键改进:引入轻量级主干网络,优化ASPP模块空洞卷积比例,集成EMA注意力机制以及CloAttention注意力机制。实验结果显示,本算法相比传统语义分割方法,在提高青藏高原特有植被类型提取精度的同时,也显著减轻了计算负担。本算法可以为青藏高原植被覆盖度测定提供高效、精确的技术支持,对青藏高原这一生态脆弱区域的植被恢复和生态保护工作具有重要意义。

1 材料与方法

1.1 数据处理

1.1.1 数据获取与筛选 本研究采用了2020和2021年青藏高原无人机航拍的栅格数据集,数据来源于青藏高原国家科学数据中心^[30-31]。该数据集收录了2020年8月及2021年5月底到6月,在青藏高原野外考察期间通过无人机航拍获得的影像资料。拍摄范围广泛,包括西藏的拉萨、山南、日喀则、林芝等地区,以及云南的大理、怒江和四川的甘孜、阿坝、凉山等州市。这批航拍影像详细记录了当地的土地使用状况、植被覆盖类型和草地覆盖度等信息,为本研究植被识别与提取的研究提供了宝贵的验证和参考资料(图1)。

在初步筛选阶段,本研究淘汰了数据集中不满足识别样本要求的图像,例如覆盖率极低的高原裸露地貌、河床侵蚀区以及其周围难以识别的植被等。同时,排除了那些包含大量空白区域或存在其他质量问题的图像样本,以确保选用的数据集在植被识别与提取研究中具有高度的可靠性和代表性。

1.1.2 数据标注与预处理 在数据处理阶段,本研究使用Labelme软件对筛选得到的数据进行了标注。为防止模型过拟合,丰富训练数据集,提升模型的泛化能力及鲁棒性,本研究使用90°和180°图像旋转、亮度调节、对比度调节、色彩调整以及图像水平与垂直翻转共7种数据增强技巧,将这些技术应用于原始图像及其对应的标签图上,以扩增样本的多样性和数量,最终获得共1455张样本数据集。在图像标注与增强工作完成之后,本研究将整

个数据集依照8:1:1的比例随机分配成训练集、验证集和测试集,确保深度学习模型能够得到有效的训练与评估。



图1 青藏高原无人机航拍图

Fig. 1 Unmanned aerial vehicle aerial images of the Tibetan Plateau

1.2 青藏高原植被提取算法框架

针对青藏高原无人机航拍图像的植被提取问题,本研究提出一种双阶段提取算法框架,包括植被的识别定位与分割两个主要环节,具体流程见图2。在算法的初始阶段,利用YOLOv5目标检测技术精准定位植被位置,并对这些区域进行剪裁处理,以此降低背景信息对后续处理的干扰,提升算法对小目标植被的提取能力。随后,将经过初步处理的植被图像输入本研究改进的DeeplabV3+语义分割网络中,以实现青藏高原航拍图植被区域的精确分割和高效提取。

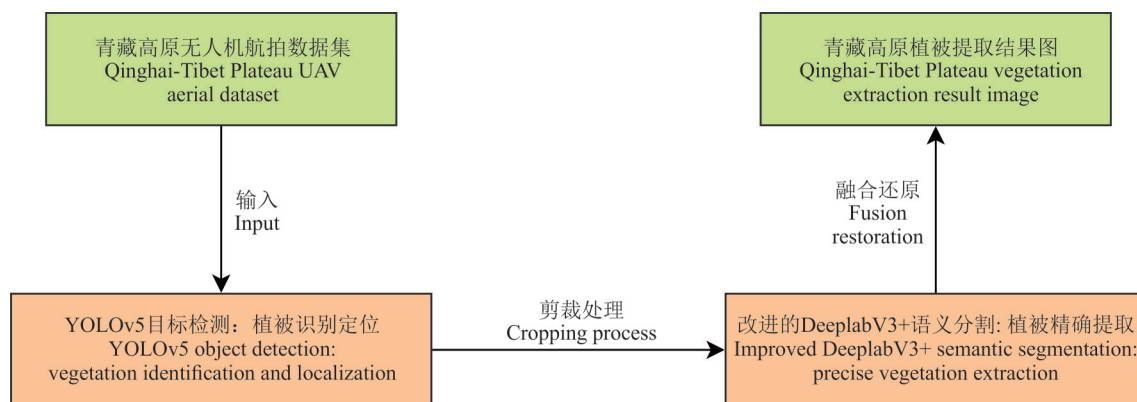


图2 青藏高原植被提取算法流程

Fig. 2 Vegetation extraction algorithm process for the Tibetan Plateau

1.3 基于YOLOv5的植被识别定位

YOLOv5模型由UltraAnalytics团队于2020年6月提出^[13],共发布了4个版本:YOLOv5s、YOLOv5m、YOLOv5l和YOLOv5x。该模型的网络架构主要由输入端(input)、主干网络(backbone)、颈部网络(neck)和预测层(prediction)4个部分组成。在输入端,YOLOv5应用Mosaic数据增强技术和自适应图像缩放机制,提升训练样本的多样性和小尺寸目标的检测能力,这对应青藏高原的复杂植被场景至关重要。主干网络采用跨阶段部分

(cross stage partial, CSP)模块和 Focus 模块,有效增加感受野和网络容量,同时空间金字塔池化(spatial pyramid pooling, SPP)模块结合卷积层提升特征提取能力。颈部结构结合特征金字塔网络(feature pyramid network, FPN)和路径聚合网络(path aggregation network, PAN),优化特征整合和信息传递,增强网络对植被的语义理解。预测层利用 GIOU Loss 和非极大值抑制技术优化边界框预测,提高检测准确性。

本研究选择 YOLOv5 的轻量化版本——YOLOv5s,对青藏高原无人机航拍图中的植被进行识别定位。YOLOv5s 在保持原有高效性和准确度的同时,其较小的模型尺寸更适合计算资源受限的应用环境。此外,YOLOv5s 的高度灵活性和适应性使其能够有效识别和精确定位不同大小的植被。因此,本研究选用 YOLOv5s 作为第一阶段的目标检测模型,以适用于青藏高原复杂的地形和多样化的植被,确保算法的有效性和适用性。

1.4 基于改进 DeeplabV3+ 的植被提取

1.4.1 DeeplabV3+ 模型结构 DeeplabV3+ 模型,由 Chen 等^[15]于 2018 年提出,目前已广泛应用于各类图像分割任务中。该模型包括编码器(encoder)和解码器(decoder)两部分,并整合了深度卷积神经网络(deep convolutional neural networks, DCNN)和空洞空间卷积池化金字塔(atrous spatial pyramid pooling, ASPP)模块。在编码器部分,DeeplabV3+ 采用 Xception 结构作为骨架网络,借助 DCNN 提取出高阶和低阶特征层。ASPP 模块通过空洞卷积(atrous convolution)增加卷积核的有效尺寸并扩展感受野,组合 1×1 卷积、3 层不同空洞率的 3×3 卷积,及全局平均池化层,提高特征提取的效率和范围。解码器部分使用 1×1 卷积压缩低阶特征,并与高阶特征有效融合。通过 3×3 卷积进一步细化特征信息,并执行 4 倍线性上采样以恢复图像原始尺寸,输出高精度的分割结果。解码部分不仅提升了模型对边缘细节的处理能力,也确保了特征的综合利用,从而全面理解和处理图像细节和上下文信息。DeeplabV3+ 模型的结构如图 3 所示。

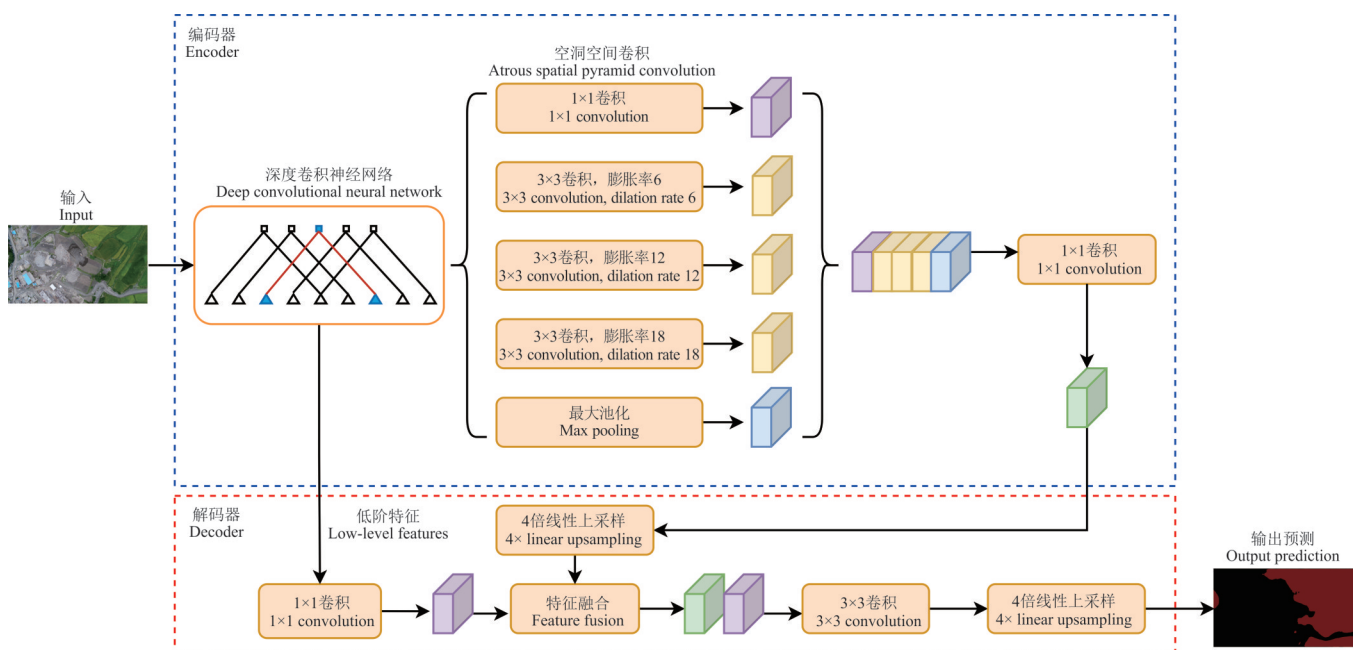


图 3 DeeplabV3+ 模型结构

Fig. 3 DeeplabV3+ model architecture

1.4.2 改进 DeeplabV3+ 的模型结构 为了适应青藏高原复杂地形的需求,并充分利用多模型植被识别的优势,本研究对 DeeplabV3+ 模型进行了优化改进,以实现青藏高原无人机航拍图像中植被信息的高精度提取。经过 YOLOv5 识别定位后的青藏高原植被图像集中于特定植被区域,这为后续的植被精确提取带来了新的挑战:DeeplabV3+ 模型需要在保持大范围区域识别精度的同时,对局部细节特征实现更精细的语义分割。此外,由于识别定位后图像的尺寸通常更小,原有的全局上下文信息可能受到影响,因此,需要调整模型对上下文信息

的处理方式,以适应这些更加专注的图像。同时,较小的图像尺寸也为提高计算效率提供了可能性。

针对上述挑战,本研究从4个方面对DeeplabV3+网络模型进行优化。首先,将模型的主干网络由Xception更换为更轻量的MobileNetV2,旨在降低计算成本和缩短运算时间,使模型更适合处理复杂且尺寸较小的青藏高原植被图像;其次,编码区改进ASPP空洞卷积率组合为4、12、36、48,提高网络对不同大小分割目标的分割能力;此外,在经过空洞卷积的有效特征层之后引入EMA注意力模块,增强模型对聚焦图像的微观特征的捕捉能力,提升植被提取的准确性和鲁棒性;最后,在DeeplabV3+的主干网络输出的低特征层之后引入CloAttention注意力机制,进一步优化了模型对于局部纹理和形状等细节的处理能力,同时保留了模型对全局景观特征的认识,以提升模型在区分植被与非植被区域时的效率和准确度。改进后的DeeplabV3+的网络结构如图4所示。

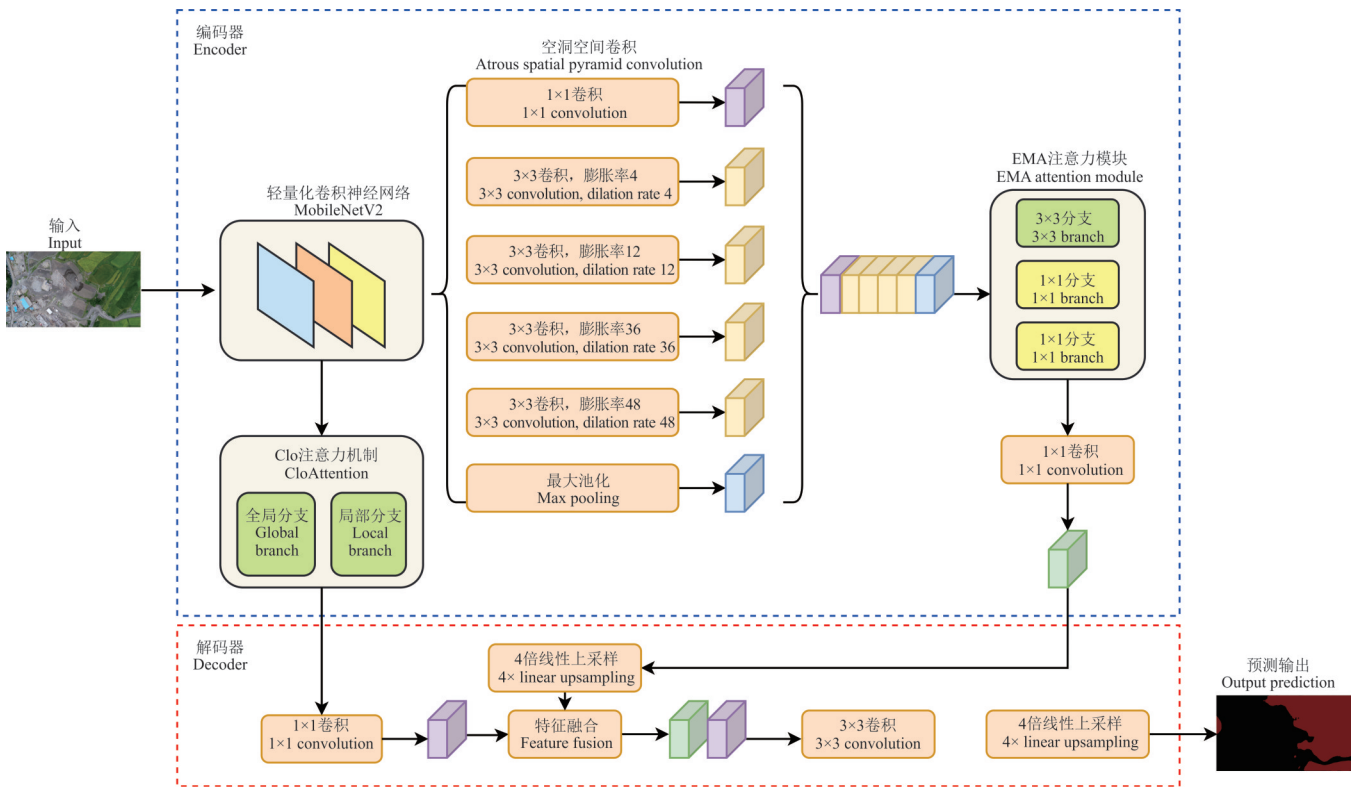


图4 改进DeeplabV3+模型结构

Fig. 4 Improved DeeplabV3+ model architecture

1.4.3 替换主干网络 原DeeplabV3+模型以Xception为主干网络,其高计算需求和大参数量在处理青藏高原航拍图像时可能导致提取效率低下及精度不足等问题。此外,YOLOv5识别后图像通常尺寸较小,对计算资源需求相对降低。因此,本研究选用具有扩张卷积的轻量级网络MobileNetV2作为主干网络。相比于Xception,MobileNetV2大幅减少了参数量,并降低了模型的权重大小及训练时间,更适合本研究需求。

MobileNetV2的结构包括线性瓶颈结构(linear bottlenecks)、深度可分离卷积(depthwise separable convolution)和倒残差网络(inverted residuals)3部分,其结构如图5所示。每个线性瓶颈层通过 1×1 卷积减少通道数,执行 3×3 深度可分离卷积后通过 1×1 卷积恢复通道数。倒残差结构在 1×1 卷积扩展维度后,进行深度可分离卷积并通过 1×1 卷积降维。此外,残差连接在输入输出维度一致时直接跳跃连接。这种设计既保持了MobileNetV2的轻量化特性,又提高了网络在提取复杂特征时的效能。中间层的高维度空间有利于学习复杂的特征,而输出的紧凑性则有效降低了计算负荷。

1.4.4 编码区改进空洞空间卷积池化金字塔模块 空洞空间卷积池化金字塔(ASPP)模块通过不同膨胀率

力。青藏高原地形的复杂性且植被的广泛分布,原有的 ASPP 模块使用 6、8、12 空洞膨胀率组合,无法完全覆盖各种尺度的植被特征。为更有效地提取多分辨率特征,提高模型对不同大小分割目标的分割能力,本研究对 ASPP 的膨胀率进行优化,选择 4、12、36、48 作为新的组合。较大的膨胀率使得模型能够覆盖广阔的植被区域,而较小的膨胀率则精准捕捉局部细节,两者结合保证了植被特征的全面提取。经过 ASPP 模块的 1×1 卷积、不同空洞率的 3×3 卷积以及最终的池化操作,不同尺寸分割目标的特征被逐一提取。将输出的特征图进行融合后,得到了优化后的 ASPP 空洞膨胀率特征图,为模型提供了更丰富的信息。

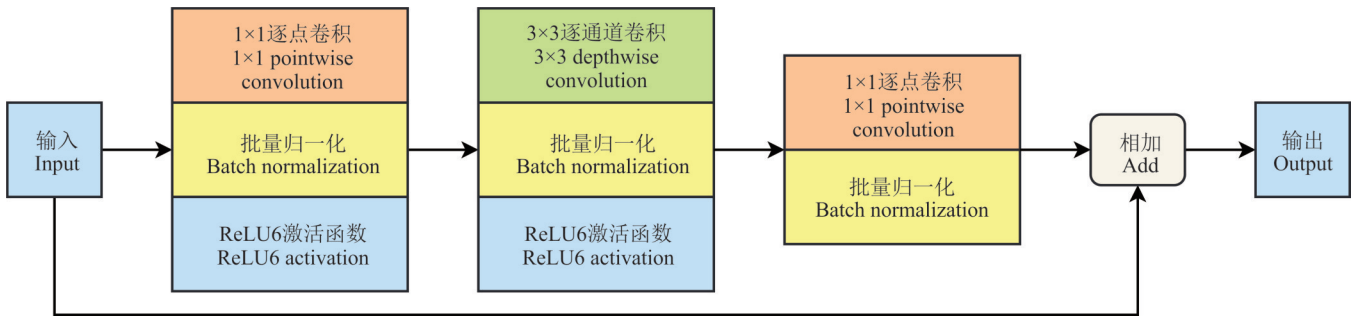


图 5 MobileNetV2 网络结构

Fig. 5 MobileNetV2 network architecture

1.4.5 EMA 注意力机制 EMA (efficient multi-scale attention) 模块是一种高效的多尺度空间特征学习注意力机制^[32],它采用分组策略而无需降维,通过跨维度交互作用有效保留每个通道的信息。通过将通道维度划分为若干子集,实现空间语义特征在各个特征图中的均匀分布。EMA 注意力机制由一个 1×1 分支、一个 3×3 分支和一个跨空间学习模块构成,其结构如图 6 所示。在 1×1 分支中,EMA 通过两个一维全局平均池化操作沿高度和宽度方向对通道进行编码,捕获跨通道交互信息,编码后的通道特征使用 1×1 卷积处理,通过 Sigmoid 激活实现特征选择。 3×3 分支则使用 3×3 卷积提取多尺度特征。跨空间学习模块采用二维全局平均池化,并与 3×3 分支输出进行点积操作,生成空间注意力图,以提高特征的空间精确度。

为进一步增强模型捕获青藏高原植被航拍图像中细微特征的能力,本研究在 ASPP 空洞卷积处理后的特征层引入 EMA 模块。通过 EMA 模块的 1×1 和 3×3 分支建立更加细致的短期和长期依赖关系,为模型的上采样过程提供更丰富的空间和通道信息。尤其针对青藏高原这一地理环境复杂的区域,EMA 模块的高效多尺度注意力机制可有效捕捉从宏观到微观的多层次空间特征,极大提升了模型在植被提取任务上的准确性及鲁棒性。

1.4.6 CloAttention 注意力机制 CloAttention 注意力机制由清华大学团队在其开发的 CloFormer 模型中提出^[33]。该注意力机制的核心在于能够同时处理高频局部信息和保留全局信息。在航拍图像的植被区域分析中,植被的纹理和形状等高频局部信息至关重要,而对于更广阔的景观特征而言,高效区分植被与非植被区域则是关键所在。因此,本研究在原始 DeeplabV3+ 模型的基础上,将 CloAttention 注意力机制融入浅层特征输入解码器之前,以增强模型对青藏高原航拍图像中复杂植被信息的局部细节与全局上下文的捕捉能力。

CloAttention 通过全局分支和局部分支的结合来处理信息(图 7)。全局分支通过对 Key(K) 和 Value(V) 的下采样以及传统注意力操作,有效捕捉全局低频信息,减少计算负担并拓展全局感知能力。局部分支引入结合了注意力机制与卷积操作的 AttnConv 模块,能够有效捕捉高频局部信息。AttnConv 通过深度可分离卷积(DWConv)对 V 进行信息聚合,然后采用一种强化的非线性方法生成上下文感知权重。这种方法相较于传统卷积提供了更精准的局部特征适应性,通过共享权重提升对高频信息的捕捉效率,从而提升模型性能。

2 结果与分析

2.1 算法评价体系

为了全面评估本研究所提出的“基于 YOLOv5 和改进 DeeplabV3+ 青藏高原植被提取算法”的性能,实验选

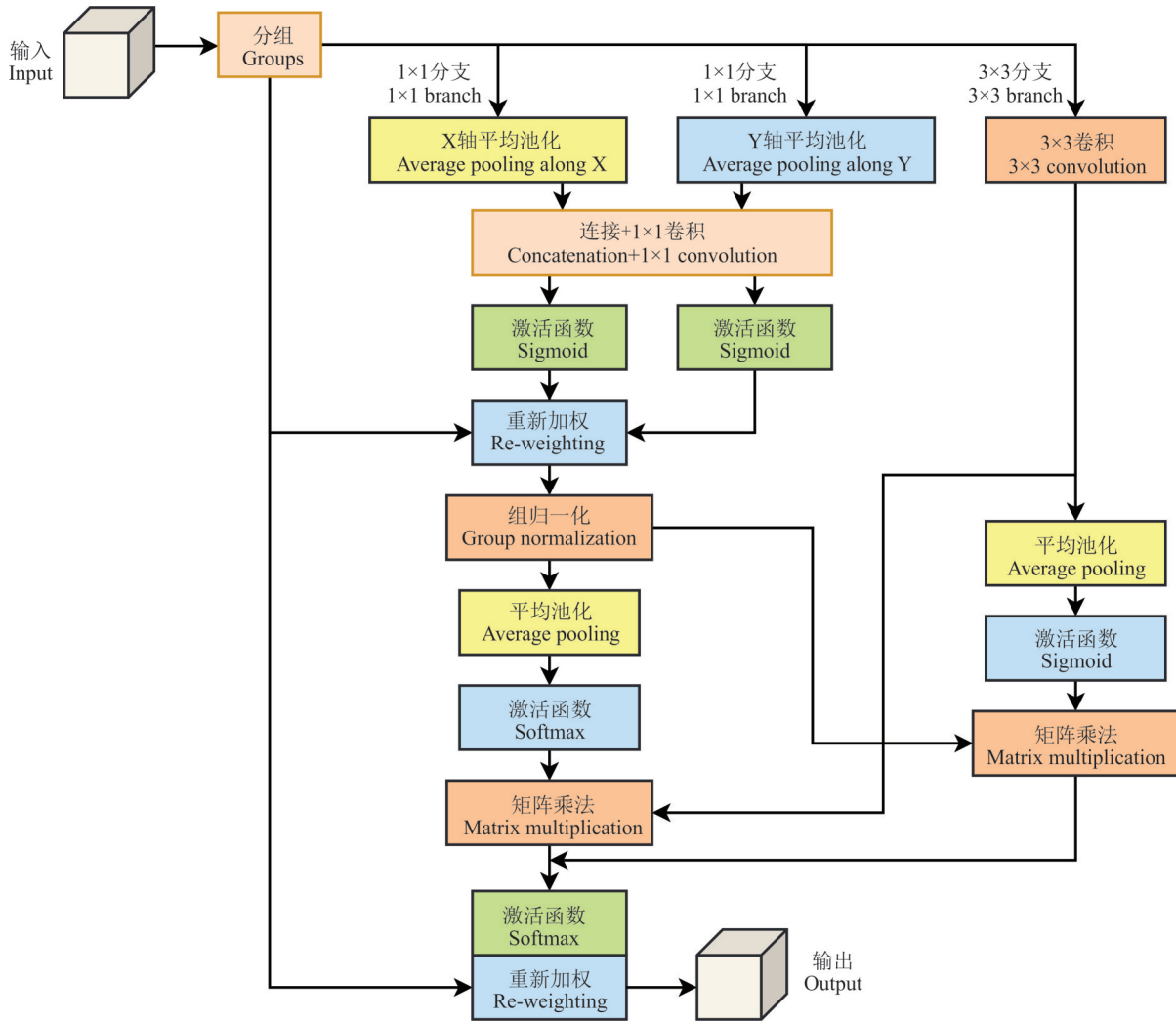


图6 EMA 模块结构
Fig. 6 EMA module structure

择了一系列评价指标。对于YOLOv5模型,本研究选取精确度(precision, P)、召回率(recall, R)、平均准确率(mean average precision, mAP)以及每秒帧数(frames per second, FPS)这4个关键指标。针对DeeplabV3+模型,本研究选取交并比(intersection over union, IoU)、像素准确率(pixel accuracy, PA)、参数量(params)和浮点运算次数(giga floating-point operations per second, GFLOPs)作为评价标准。计算公式分别如下:

$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP + FN}$$

$$mAP = \frac{\int_0^1 P(R) dR}{M}$$

$$FPS = \frac{N}{t}$$

$$IoU = \frac{TP}{TP + FP + FN}$$

$$PA = \frac{TP + TN}{TP + TN + FP + FN}$$

式中: TP为整体样本中正确预测为高原植被的样本数量; TN表示整体样本中正确预测为非高原植被的样本数

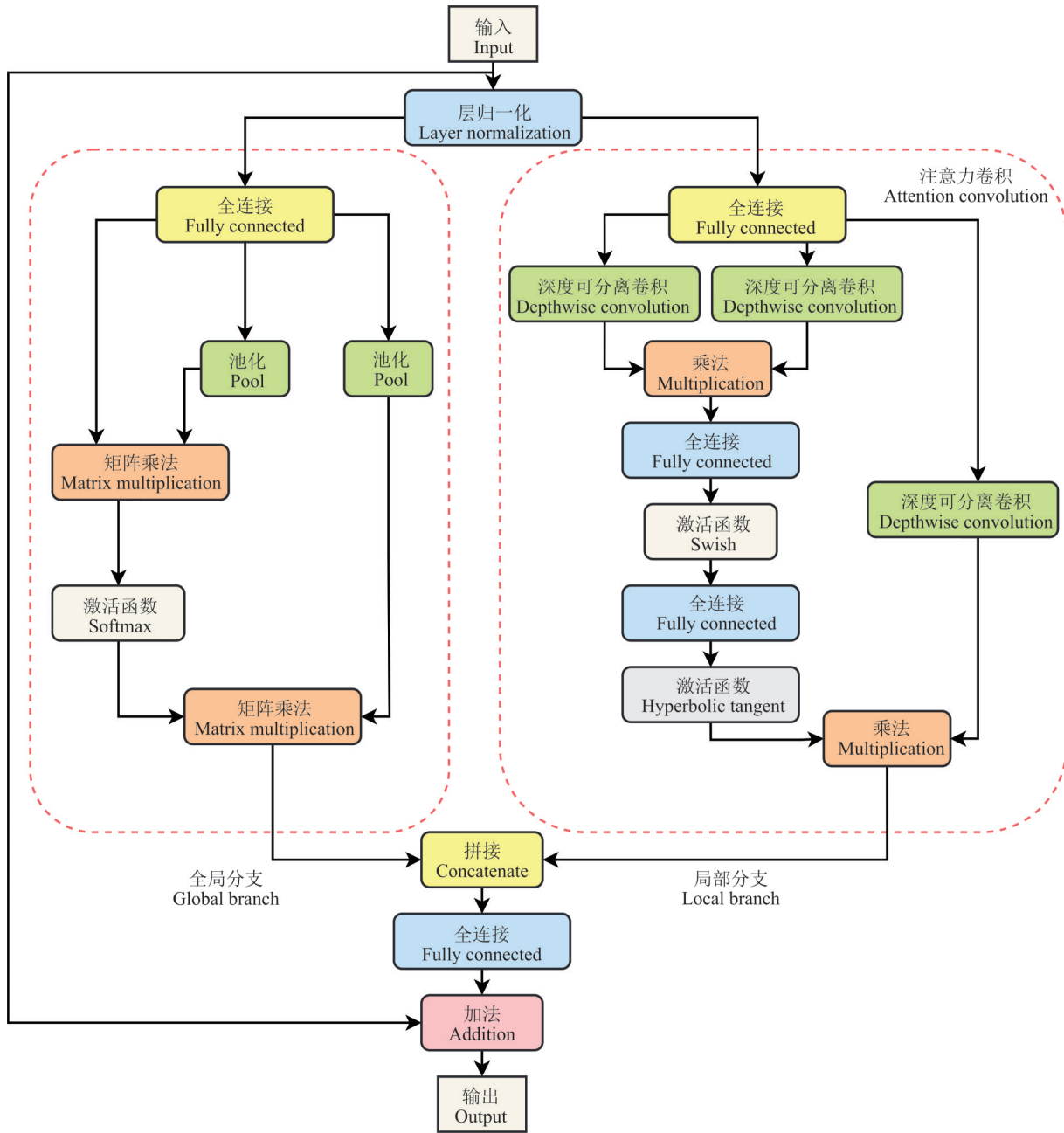


图 7 CloAttention 结构
Fig. 7 CloAttention structure

量; FP 为整体样本中错误预测为高原植被的样本数量; FN 为未检测出高原植被的样本数量; M 为目标类别数量; N 为检测样本个数; t 为检测总时间。

2.2 实验环境

本研究实验环境为 64 位 Ubuntu 18.04.6 LTS 操作系统下搭建的 Pytorch 1.12.1 深度学习框架, Python 3.11.5 以及 CUDA 11.4 编程软件。计算机 CPU 为 Intel Xeon CPU E5-2603 v4 @ 1.70GHz, GPU 为 NVIDIA GeForce GTX 1080 Ti, 显存为 11 GB。

2.3 基于 YOLOv5 的植被识别定位实验

为了验证 YOLOv5 系列模型在植被识别定位方面的性能, 本研究对其 4 种不同规模的变体——YOLOv5s、YOLOv5m、YOLOv5l 和 YOLOv5x 进行了比较实验。训练采用统一的参数设置: 迭代周期 (epochs) 定为 100, 批处理大小 (batch-size) 设为 16。4 种模型的训练评估结果 (表 1) 表明, YOLOv5s 在检测效率每秒帧数 (FPS) 上的

性能优于其他3种模型,且具有高准确率,P指标达到97.9%,R指标达到97.3%,mAP达到98.9%,证明了本研究选取YOLOv5s作为植被识别定位模型的适用性。

2.4 基于改进DeeplabV3+的植被提取实验

为验证改进的DeeplabV3+模型在植被提取方面的性能,本研究使用经典分割网络PSPnet、HRnet、Unet和DeeplabV3+共4种网络作为实验对象进行对比,训练采用统一的参数设置:迭代周期(epochs)定为100,批处理大小(batch-size)设为8。5种模型的训练评估结果(表2)表明,改进后的DeeplabV3+模型在IoU和PA指标上获得大幅提升,同时Params和GFLOPs大幅下降,明显减少了参数量和计算复杂度。与PSPnet相比,IoU和PA指标分别提升了10.20%和8.85%;与HRnet相比,提升了8.13%和8.93%;与Unet相比,提升了5.85%和6.54%;与原始DeeplabV3+相比,提升了7.01%和4.60%。因此,本研究提出的基于改进DeeplabV3+的植被提取算法可以更好地适用于青藏高原无人机航拍图像的植被语义特征的分割提取。图8展示了5种网络模型的植被提取效果。

2.5 稳定性评估

为了验证本研究提出的结合YOLOv5和改进DeeplabV3+模型的双阶段植被提取算法的可靠性和稳定性,本研究对算法在青藏高原不同用地类型下的表现进行了全面评估。实验中选择了5种具有代表性的用地类型:农业用地、工业用地、生活用地以及自然地貌中的河流湿地和裸岩石区,以全面评估模型的提取性能。为确保评估的有效性,每种类型均选取50张在光照条件、季节变化和拍摄角度等方面具有代表性的图像。在性能评估指标方面,选用交并比、像素准确率、精确度和召回率来衡量算法的提取精度。所有选定的图像均经过必要的预处理,包括尺寸调整和归一化等,以适配模型的输入要求。实验中,图像数据首先通过YOLOv5模型进行植被区域的初步定位,随后利用改进的DeeplabV3+模型进行详细的语义分割,从而对定位的区域进行精细化提取。

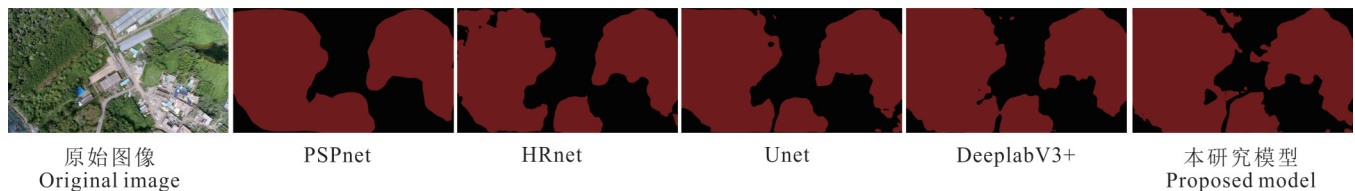


图8 不同网络模型植被提取效果

Fig. 8 Vegetation extraction effect of different network models

5种类型的稳定性评估结果(表3)表明,本研究提出的双阶段植被提取算法在不同用地类型下均展示出高度的可靠性和稳定性。裸岩石区的IoU值最高,达到93.12%,表明模型在辨别具有复杂背景的裸岩石区植被时,能够提供最准确的分割结果。河流湿地和农业用地也表现出较高的IoU值,分别为90.40%和88.96%,证实了模型在处理水体邻近区域及农田的植被识别上的有效性。所有用地类型的PA值均高于95%,表明模型在整体像

表1 YOLOv5对比实验评估结果

Table 1 Evaluation results of YOLOv5 comparative experiments

模型 Model	精确率 Precision (%)	召回率 Recall (%)	平均准确率 Mean precision (%)	每秒帧数 Frames per second
YOLOv5x	98.3	98.0	99.2	52
YOLOv5l	98.5	97.3	99.1	82
YOLOv5m	97.9	96.4	98.8	119
YOLOv5s	97.9	97.3	98.9	204

表2 植被提取对比实验评估结果

Table 2 Vegetation extraction comparative experimental evaluation results

模型 Model	交并比 Intersection over union (%)	像素准确 率 Pixel accuracy (%)	参数量 Params	浮点运算次数 Giga floating- point operations per second
PSPnet	80.20	87.47	46.707	118.427
HRnet	82.27	87.39	29.538	79.915
Unet	84.55	89.78	43.933	184.100
DeeplabV3+	83.39	91.72	54.709	166.841
改进 DeeplabV3+ Improved Deep- labV3+	90.40	96.32	6.565	54.630

素级别的分类准确性很高。裸岩石区在 P 值上达到了最高值,为 97.64%,说明模型在该类型地区准确标定植被像素点的能力极强。R 值在各类用地中均表现良好,最低值为 95.05%,表明模型几乎能够检测到所有相关的植被区域,且误报较少。

模型在各种环境条件下均能保持较高的识别精度和准确性,这为其应用于青藏高原复杂地形下的植被提取提供了有力的验证。此外,模型在裸岩石区和河流湿地等地形下表现出的卓越性能,证明了其在处理复杂背景条件下的高效性。图 9 展示了本模型在 5 种用地类型下的植被提取效果。

表 3 不同用地类型评估结果

Table 3 Assessment results across different land use types (%)

用地类型 Land use type	交并比 Intersection over union	像素准确率 Pixel accu- racy	精确率 Preci- sion	召回率 Recall
农业用地 Agricultural land	88.96	96.48	90.54	96.48
工业用地 Industrial land	87.74	95.57	91.80	95.57
生活用地 Residential land	88.12	96.04	90.69	96.04
河流湿地 Riverine wetlands	90.40	95.13	94.64	95.13
裸岩石区 Bare rock area	93.12	95.05	97.64	95.05



图 9 5 种用地类型下的植被提取效果

Fig. 9 Vegetation extraction outcomes across five land use types

3 讨论

3.1 模型精度分析

本研究提出的结合 YOLOv5 和改进 DeeplabV3+ 的双阶段植被提取算法,显著提高了植被提取精度并优化了检测效率。该成效主要得益于优化的算法流程和模型结构。双阶段植被提取算法融合了两种模型的优势:YOLOv5 模型准确识别和定位出植被区域,有效减轻了高分辨率图像处理的计算负担,同时提升了算法对细小植被目标的识别效率。改进后的 DeeplabV3+ 模型则通过精细的语义分割,确保植被信息的高精度提取。此外,对 DeeplabV3+ 模型的改进优化增强了算法在处理青藏高原复杂多样地形时的适用性。通过引入 MobileNetV2 作为主干网络,显著降低了模型的计算需求;优化 ASPP 空洞卷积率组合为 4、12、36、48,提升了模型大范围分割提取的精准度,同时保持了对小范围目标的捕捉能力;引入的 EMA 和 CloAttention 注意力机制显著提升了模型对植被细节的捕捉能力,尤其是在处理局部纹理和形状的微观特征时。因此,算法在多个关键性能指标上均实现了显著提升,在青藏高原各类环境下展现了出色的植被提取效果。

3.2 影响因素分析

数据集标注的准确性是影响模型整体精度的主要因素。青藏高原地区的植被类型复杂多样,且在地形上分布极为不均匀。由于青藏高原的独特生态环境和气候条件,当地植被展现出与其他地区不同的生长状态和空间分布特征,形成了大量植被覆盖度稀疏且不连续的区域。这些特点导致植被边界难以准确界定,使得人工标注过程中不可避免地引入了一定的主观性和误差。此外,工业用地和生活用地等区域由于包含众多人造结构和复杂地面覆盖物,导致其与植被区域的对比度较低,增加了标注的难度。这些标注误差直接影响了模型训练时使用的监督信息质量,进而影响模型在实际应用中的提取精度。

在不同用地类型的表现评估中,自然地貌的提取精度最高,而工业用地、生活用地等类型下的提取精度相对较低。这主要是因为自然地貌区域受人为干扰较小,光照条件更为均匀,从而有助于算法更有效地执行植被识别和提取任务。相反,生活和工业用地除了受到数据集标注的影响外,还因建筑物和其他人造设施的存在,不仅增加了阴影区域,也使得植被与非植被的界限变得模糊,从而增加了植被提取的复杂性和误差率。

3.3 不足与展望

尽管本研究提出的双阶段植被提取算法在青藏高原地区展示了显著的性能,但仍存在一些局限性和未来的改进空间。目前采用的数据集虽覆盖了多样的地理特征,但未能充分包含青藏高原在极端气候条件下的环境特征,这影响了图像数据处理的鲁棒性。因此,未来工作将扩展数据收集范围,整合具有更广泛环境和季节变化的数据集,以增强算法的适应性和鲁棒性。

此外,数据集标注的准确性是影响本算法植被提取精度的关键因素。人为标注过程中不可避免的误差会导致训练模型的监督信息出现偏差。未来,为提高数据集的标注质量,本研究将探索具有更高精度的标注技术以提升标注信息的一致性与精确度。这些措施将显著提升模型训练的效率 and 植被提取任务的总体精度,为青藏高原植被覆盖度的准确监测提供技术支持,帮助研究人员更好的理解和评估植被分布的生态和环境价值。

4 结论

本研究提出基于YOLOv5和改进DeeplabV3+模型的青藏高原植被提取算法,克服了传统植被检测方法在精度、成本和时效性方面的缺陷。算法弥补了运用深度学习技术对青藏高原植被进行精准提取研究的不足,通过综合运用多种模型的优势,实现了对青藏高原植被的高精度提取。采用无人机航拍数据作为研究对象,以规避多源遥感数据在植被识别方面的局限性。算法引入YOLOv5进行植被区域的初步定位,再通过改进的DeeplabV3+进行精细的语义分割。针对青藏高原复杂的植被信息和YOLOv5的识别定位结果,本研究对DeeplabV3+模型进行了4方面的关键改进:替换主干网络为轻量级的MobileNetV2,优化ASPP模块的空洞卷积率组合,以及引入EMA和CloAttention注意力机制。实验结果显示,本算法在青藏高原无人机航拍数据集上的交并比和像素准确率分别达到了90.40%和96.32%,在参数量和计算复杂度上均实现了显著优化,超越了传统语义分割算法,既实现了网络的轻量化又提升了对青藏高原航拍图植被的提取精度。在不同用地类型下,本算法均展示了出色的提取精度,证明了其高度可靠性和稳定性。本研究不仅可以为青藏高原植被覆盖度检测提供精确的科学依据,也可以为快速评估偏远或难以接触地区的生态状况提供重要的技术支持。

参考文献 References:

- [1] Dong S. Revitalizing the grassland on the Qinghai-Tibetan Plateau. *Grassland Research*, 2023, 2(3): 241-250.
- [2] Sang J W, Song C Y, Jia N X, *et al.* Vegetation survey and mapping on the Qinghai-Tibet Plateau. *Biodiversity Science*, 2023, 31(3): 56-71.
桑佳文, 宋创业, 贾宁霞, 等. 青藏高原植被调查与制图评估. *生物多样性*, 2023, 31(3): 56-71.
- [3] Zhang W B, Fu S H, Liu B Y. Error assessment of visual estimation plant coverage. *Journal of Beijing Normal University (Natural Science)*, 2001, 37(3): 402-408.
章文波, 符素华, 刘宝元. 目估法测量植被覆盖度的精度分析. *北京师范大学学报(自然科学版)*, 2001, 37(3): 402-408.
- [4] Zhang W B, Liu B Y, Wu J D. Monitoring of plant coverage of plots by visual estimation and overhead photograph. *Bulletin of Soil and Water Conservation*, 2001, 21(6): 60-63.
章文波, 刘宝元, 吴敬东. 小区植被覆盖度动态快速测量方法研究. *水土保持通报*, 2001, 21(6): 60-63.
- [5] Yang Q, Pu H M, Zhao X C, *et al.* Comparison of different plant cover investigation methods for three artificial grasslands. *Chinese Journal of Applied & Environmental Biology*, 2021, 27(1): 220-227.
杨琴, 蒲红梅, 赵学春, 等. 3种人工草地不同植被覆盖度实地测量方法比较. *应用与环境生物学报*, 2021, 27(1): 220-227.
- [6] Canfield R H. Application of the line intercept method in sampling range vegetation. *Journal of Forestry*, 1941, 39(4): 388-394.

- [7] Zhang Y X, Li X B, Chen Y H. Overview of field and multi-scale remote sensing measurement approaches to grassland vegetation coverage. *Advances in Earth Science*, 2003, 18(1): 85–93.
张云霞, 李晓兵, 陈云浩. 草地植被盖度的多尺度遥感与实地测量方法综述. *地球科学进展*, 2003, 18(1): 85–93.
- [8] Huang P, Pu J W, Zhao Q Q, *et al.* Research progress and development trend of remote sensing information extraction methods of vegetation. *Remote Sensing for Natural Resources*, 2022, 34(2): 10–19.
黄佩, 普军伟, 赵巧巧, 等. 植被遥感信息提取方法研究进展及发展趋势. *自然资源遥感*, 2022, 34(2): 10–19.
- [9] Shen M X, He R Y, Cong J H, *et al.* Study on extraction of vegetation information of ETM+ by using PCA method and Brovey transform. *Transactions of the Chinese Society for Agricultural Machinery*, 2007, 38(9): 87–89.
沈明霞, 何瑞银, 丛静华, 等. 基于主成分分析与 Brovey 变换的 ETM+ 影像植被信息提取. *农业机械学报*, 2007, 38(9): 87–89.
- [10] Tang P Q, Wu W B, Yao Y M, *et al.* New method for extracting multiple cropping index of North China Plain based on wavelet transform. *Transactions of the Chinese Society of Agricultural Engineering*, 2011, 27(7): 220–225.
唐鹏钦, 吴文斌, 姚艳敏, 等. 基于小波变换的华北平原耕地复种指数提取. *农业工程学报*, 2011, 27(7): 220–225.
- [11] Zhang X Y, Jing Y S, Li W G. Optimal scale screening of paddy rice in remote sensing imagery based on high pass filter fusion. *Chinese Journal of Agrometeorology*, 2018, 39(5): 344–353.
张晓忆, 景元书, 李卫国. 基于高通滤波算法的水稻遥感影像适宜尺度筛选. *中国农业气象*, 2018, 39(5): 344–353.
- [12] Dai P Q, Ding L X, Liu L J, *et al.* Tree species identification based on FCN using the visible images obtained from an unmanned aerial vehicle. *Laser & Optoelectronics Progress*, 2020, 57(10): 101001.
戴鹏钦, 丁丽霞, 刘丽娟, 等. 基于 FCN 的无人机可见光影像树种分类. *激光与光电子学进展*, 2020, 57(10): 101001.
- [13] Redmon J, Divvala S, Grishick R, *et al.* You Only Look Once: unified, real time object detection//Institute of Electrical and Electronic Engineers. *Conference on Computer Vision and Pattern Recognition*. Las Vegas: Institute of Electrical and Electronic Engineers, 2016: 779–788.
- [14] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 2017, 60(6): 84–90.
- [15] Chen L C, Papandreou G, Kokkinos I, *et al.* Semantic image segmentation with deep convolutional nets and fully connected CRFs. *Computer Science*, 2014(4): 357–361.
- [16] Chen L C, Papandreou G, Kokkinos I, *et al.* Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 40(4): 834–848.
- [17] Chen L C, Papandreou G, Schroff F, *et al.* Rethinking atrous convolution for semantic image segmentation. 2017, DOI: 10.48550/arXiv.1706.05587.
- [18] Chen L C, Zhu Y, Papandreou G, *et al.* Encoder-Decoder with atrous separable convolution for semantic image segmentation//European Conference on Computer Vision. Germany: Springer, 2018: 801–818.
- [19] Szegedy C, Liu W, Jia Y, *et al.* Going deeper with convolutions//Institute of Electrical and Electronic Engineers. 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: Institute of Electrical and Electronic Engineers, 2015: 1–9.
- [20] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation//Medical Image Computing and Computer Assisted Intervention Society. *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015; 18th International Conference*. Munich, Germany: Springer, 2015: 234–241.
- [21] He K, Zhang X, Ren S, *et al.* Deep residual learning for image recognition//Institute of Electrical and Electronic Engineers. 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: Institute of Electrical and Electronic Engineers, 2016: 770–778.
- [22] Zhao H, Shi J, Qi X, *et al.* Pyramid scene parsing network//Institute of Electrical and Electronic Engineers. 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: Institute of Electrical and Electronic Engineers, 2017: 2881–2890.
- [23] Sun K, Xiao B, Liu D, *et al.* Deep high-resolution representation learning for human pose estimation//Institute of Electrical and Electronic Engineers. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. California: Institute of Electrical and Electronic Engineers, 2019: 5693–5703.
- [24] Zhou X X, Wu Y L, Li M Y, *et al.* Automatic vegetation extraction method based on feature separation mechanism with deep learning. *Journal of Geo-information Science*, 2021, 23(9): 1675–1689.
周欣昕, 吴艳兰, 李梦雅, 等. 基于特征分离机制的深度学习植被自动提取方法. *地球信息科学学报*, 2021, 23(9): 1675–

1689.

- [25] Zhang Y, Wang H, Liu J, *et al.* A lightweight winter wheat planting area extraction model based on improved DeepLabv3+ and CBAM. *Remote Sensing*, 2023, 15(17): 4156.
- [26] da Silva Mendes P A, Coimbra A P, de Almeida A T. Vegetation classification using DeepLabv3+ and YOLOv5//Institute of Electrical and Electronic Engineers. ICRA 2022 Workshop in Innovation in Forestry Robotics: Research and Industry Adoption. USA: Institute of Electrical and Electronic Engineers, 2022.
- [27] Hu Y N, An R, Ai Z T, *et al.* Researches on grass species fine identification based on UAV hyperspectral images in Three—River Source region. *Remote Sensing Technology and Application*, 2021, 36(4): 926—935.
胡宜娜, 安如, 艾泽天, 等. 基于无人机高光谱影像的三江源草种精细识别研究. *遥感技术与应用*, 2021, 36(4): 926—935.
- [28] Zhang Y P, Wu X T, Li X L, *et al.* Identification of degraded grassland in Qinghai area of Yellow River Source based on high-resolution images. *Acta Agriculturae Boreali-occidentalis Sinica*, 2023, 32(2): 198—211.
张宇鹏, 吴笑天, 李希来, 等. 基于高分影像的黄河源青海片区退化草地识别. *西北农业学报*, 2023, 32(2): 198—211.
- [29] Wen T, Liu X N, Ji T, *et al.* Studying on plant classification and recognition method for Three—River Source alpine grassland plant based on vegetation index. *Acta Agrestia Sinica*, 2022, 30(7): 1811—1818.
文铜, 柳小妮, 纪童, 等. 基于植被指数的三江源高寒草地植物分类与识别方法研究. *草地学报*, 2022, 30(7): 1811—1818.
- [30] Lv C H, Liu Y Q. UAV-derived raster data of the Tibetan Plateau in 2020. National Tibetan Plateau/Third Pole Environment Data Center. <https://doi.org/10.11888/Geogra.tpdc.271124>. <https://cstr.cn/18406.11.Geogra.tpdc.271124>.
吕昌河, 刘亚群. 青藏高原无人机航拍栅格数据(2020). 国家青藏高原数据中心. <https://doi.org/10.11888/Geogra.tpdc.271124>. <https://cstr.cn/18406.11.Geogra.tpdc.271124>.
- [31] Lv C H, Zhang Z M. UAV-derived raster data of the Tibetan Plateau (2021). National Tibetan Plateau/Third Pole Environment Data Center. <https://doi.org/10.11888/Terre.tpdc.271903>. <https://cstr.cn/18406.11.Terre.tpdc.271903>.
吕昌河, 张泽民. 青藏高原无人机航拍栅格数据(2021). 国家青藏高原数据中心. <https://doi.org/10.11888/Terre.tpdc.271903>. <https://cstr.cn/18406.11.Terre.tpdc.271903>.
- [32] Ouyang D, He S, Zhang G, *et al.* Efficient multi-scale attention module with cross-spatial learning// ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Greece: IEEE, 2023: 1—5.
- [33] Fan Q H, Huang H B, Guan J Y, *et al.* Rethinking local perception in lightweight vision transformer. *ArXiv*, 2023, abs/2303.17803.