doi:10.6041/j.issn.1000-1298.2024.05.017

基于改进 DeepLabV3 + 的荞麦苗期无人机遥感图像分割识别方法研究

武锦龙[1,2]　吴虹麒[1,3]　李　浩[1,3]　雷兴鹏[1,3]　宋海燕[1,3]

(1. 山西农业大学农业工程学院, 太谷 030801; 2. 山西农业大学信息科学与工程学院, 太谷 030801;
3. 旱作农业机械关键技术与装备山西省重点实验室, 太谷 030801)

摘要: 针对 DeepLabV3 + 语义分割模型计算复杂度高、内存消耗大、难以在计算力有限的移动平台上部署等问题,提出一种改进的轻量化 DeepLabV3 + 深度学习语义分割算法,用于实现无人机荞麦苗期图像的分割与识别。该算法采用 RepVGG(Re-parameterization visual geometry group)与 MobileViT(Mobile vision transformer)模块融合的方式建立主干网络实现特征提取;同时,在 RepVGG 网络结构中引入 SENet(Squeeze-and-excitation networks)注意力机制,通过利用通道间的相关性,捕获更多的全局语义信息,保证荞麦分割的性能。实验结果表明,与 FCN(Fully convolutional networks)、PSPNet(Pyramid scene parsing network)、DenseASPP(Dense atrous spatial pyramid pooling)、DeepLabV3、DeepLabV3 + 模型相比,本文提出的改进算法在较大程度上降低了模型参数规模,更适合在移动端部署,自建荞麦苗期分割数据集上的语义分割平均像素准确率(Mean pixel accuracy, mPA)和平均交并比(Mean intersection over union, mIoU)分别为 97.02% 和 91.45%,总体参数量、浮点运算次数(Floating-point operations, FLOPs)和推理速度分别为 9.01×10^6、8.215×10^{10}、37.83 f/s,综合表现最优。在全尺寸图像分割中,训练模型对不同飞行高度的荞麦苗期分割的 mPA 和 mIoU 均能满足要求,也具有较好的分割能力和推理速度,该算法可为后期荞麦补种、施肥养护和长势监测等提供重要技术支持,进而促进小杂粮产业智能化发展。

关键词: 荞麦苗期; 无人机遥感; 图像语义分割; DeepLabV3 +; 轻量化

中图分类号: TP753　　**文献标识码**: A　　**文章编号**: 1000-1298(2024)05-0186-10　　**OSID**:

Segmentation of Buckwheat by UAV Based on Improved Lightweight DeepLabV3 + at Seedling Stage

WU Jinlong[1,2]　WU Hongqi[1,3]　LI Hao[1,3]　LEI Xingpeng[1,3]　SONG Haiyan[1,3]

(1. *College of Agricultural Engineering*, *Shanxi Agricultural University*, *Taigu* 030801, *China*
2. *College of Information Science and Engineering*, *Shanxi Agricultural University*, *Taigu* 030801, *China*
3. *Dryland Farm Machinery Key Technology and Equipment Key Laboratory of Shanxi Province*, *Taigu* 030801, *China*)

Abstract: In view of the problems of high computational complexity, large memory consumption, and difficulty in deployment on mobile platforms with limited computing power in DeepLabV3 + segmentation model, an improved lightweight DeepLabV3 + algorithm was proposed to realize the segmentation and recognition of buckwheat by UAV at seedling stage. The algorithm adopted the fusion of re-parameterization visual geometry group (RepVGG) and mobile vision transformer (MobileViT) modules to establish the backbone network for feature extraction. At the same time, the squeeze-and-excitation networks (SENet) attention mechanism was introduced into the RepVGG network structure to capture more global semantic information by using the correlation between channels, and ensure the performance of buckwheat segmentation. Experimental results showed that compared with fully convolutional networks (FCN), pyramid scene parsing network (PSPNet), dense atrous spatial pyramid pooling (DenseASPP), DeepLabV3, and DeepLabV3 + models, the improved algorithm proposed greatly reduced the model parameters, making it more suitable for deployment on mobile terminals. The mean pixel accuracy

收稿日期: 2024 - 02 - 26　修回日期: 2024 - 03 - 18

基金项目: 国家重点研发计划项目(2021YFD1600602 - 09)和山西省基础研究计划项目(202203021212414、202303021222067)

作者简介: 武锦龙(1990—),男,博士生,讲师,主要从事遥感数据智能信息处理研究,E-mail: wujinlong8192@163.com

通信作者: 宋海燕(1977—),女,教授,博士,主要从事田间信息检测技术研究,E-mail: yybbao@163.com

(mPA) and mean intersection over union (mIoU) on the self-built buckwheat segmentation dataset were 97.02% and 91.45%, the overall parameters, floating-point operations (FLOPs) and inference speed were 9.01×10^6, 8.215×10^{10} and 37.83 f/s, respectively, with the best performance. In the full-size image segmentation, the mPA and mIoU for buckwheat segmentation can meet the requirements at different flight heights, which had good segmentation ability and inference speed. The algorithm can provide technical support for the later buckwheat seed replacement, fertilization maintenance, and growth monitoring, and promote the intelligent development of small and coarse grain industry.

Key words: buckwheat at seedling stage; UAV remote sensing; image segmentation; DeepLabV3 +; lightweight

0　引言

荞麦是我国重要的小杂粮作物,也是典型的“药食同源”作物,具有丰富的保健功效,能够降压、控制糖尿病、改善消化和胆固醇水平等[1-2]。荞麦的生长阶段主要分为幼苗期、生长期和成熟期 3 个阶段。秋播荞麦通常于 7 月种植,其幼苗期处于雨季高发季,这个阶段荞麦对外界环境变化较为敏感,田间杂草容易与作物争夺养分、水分和光照,因此,环境变化及杂草生长情况都会影响作物的出苗情况[3]。幼苗期是荞麦营养生长的关键阶段,及时地查苗、补苗以及监测荞麦长势能够为荞麦生产管理及产量形成提供保障[4]。苗期情况获取可为幼苗数量检测、长势均匀度评估、生长指标评估等提供技术支持。传统苗期情况评估主要以经验和人工采样为主,适用于较小种植规模地块,当种植规模较大时,人工方式则费时费力、时效性差、成本高等,难以满足精准农业快速、准确获取信息的需求[5]。无人机遥感作为一种高效化、现代化、智能化的技术,可以观测大规模种植地块的全尺度图像,成为农作物出苗监测、生长监测和产量预测的必备工具。将智慧农业技术应用到小杂粮产业中是目前发展趋势,能够降低农业工作者工作量的同时,促进小杂粮产业发展。

近年来,深度学习在计算机视觉领域已表现出优异的性能,部分研究者将其应用于农产品苗期图像分割中,并取得了较好的结果。DIAO 等[6]对 UNet 网络进行了改进,对玉米作物行分割的平均像素精度(Mean pixel accuracy, mPA)、平均精度和平均召回率分别为 90.18%、91.79% 和 90.18%,比传统图像分割方法有明显提高。LIU 等[7]提出了一种高效残差分解卷积神经网络(Multi-sacle effcient residual fcatorized convNet, MS-ERFNet)模型实现玉米作物行的图像分割,其平均交并比和像素准确率分别为 93.41% 和 97.54%,均高于其他模型。SUN 等[8]对 DeepLabV3 + 图像分割模型进行了改进,采用水稻倾斜通道特征金字塔(Rice lodging-channel-wise feature pyramid, RL-CFP)代替空间金字塔池化(Atrous spatial pyramid Pooling, ASPP),并在编解码阶段添加了通道注意力机制(Channel attention module, CAM),最终实现了水稻的倒伏分割,其分割 mIoU 和 mPA 分别达到 90.52% 和 94.37%。ZUO 等[9]对 UNet 网络进行了改进,编码网络采用密集网络复用和融合多层特征改进了信息传输方式,解码网络采用转置卷积进行上采样,该模型对于玉米苗木植株分割的平均像素精度为 90.54%。ZHANG 等[10]采用 2 种轻量化模型 LW-Segnet(Lightweight Segnet)和 LW-Unet(Lightweight Unet)实现了小麦苗期高精度分割,其分割的 F1 值和 mIoU 分别为 88.00% 和 83.42%。

目前,基于深度学习的作物苗期植株语义分割主要应用在玉米、水稻等大规模种植作物领域,在荞麦领域的研究相对较少。采用无人机遥感技术和人工智能技术结合的方式对荞麦进行田间管理,能够不断促进小杂粮产业的信息化、现代化发展,最终为小杂粮产业发展提供技术支持。由于移动平台的计算能力有限、尺寸较小、质量较轻,对分割模型的总体参数量、FLOPs 和推理时间的要求都较高[11-12],因此,建立荞麦苗期分割数据库,并开发一种能够搭载于移动端的轻量级深度学习语义分割框架实现荞麦苗期图像分割,为后期作物数量统计和作物行识别提供技术支持。综上所述,本文提出一种轻量化的荞麦苗期深度学习图像分割方法,以 DeepLabV3 + 模型为基础网络,采用 RepVGG Block 与 MobileViT Block 模块融合的方式替换原始主干网络降低模型参数量,引入 SENet 注意力机制保证通道间相关性的同时捕获更多的全局语义信息,并与其他模型进行对比实验,在不降低图像分割精度和速度的同时,完成轻量化模型的构建,以期为无人机实时荞麦苗期分割提供理论基础,最终助力小杂粮智慧农业发展。

1　材料与方法

1.1　研究区域概况

于 2022 年和 2023 年在山西省晋中市太谷区申

奉村山西农业大学荞麦研究实验田(37°26′2.4″~37°26′06″N,112°35′34.8″~112°35′42.0″E)开展实验。荞麦播种时间为 6—7 月,收获时间为 9—10 月,种植品种为甜荞“晋荞 1 号”。图 1 为荞麦研究区地理位置及实验田概况。

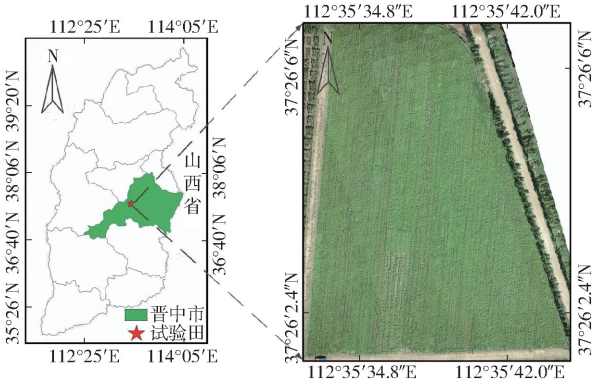

图 1 研究区地理位置及荞麦实验田

Fig. 1 Geographical location of study area and buckwheat experimental field

1.2 数据获取

航拍图像采集设备为大疆 PHANTOM 4RTK 型无人机(搭载 2×10^7 像素 1 英寸 CMOS 传感器),该无人机搭载 RTK 模块,兼容高精度 GNSS 移动站,具有厘米级定位能力。航拍参数设置:飞行高度设置为 10 m,飞行速度为 2.6 m/s,GSD 为 0.41 cm/像素,航向重叠率为 80%,旁向重叠率为 70%。在轨迹飞行模式下,绘制飞行轨迹线,无人机就会按照预设的轨迹线开始飞行。本文研究对象为荞麦的苗期数据,因此数据采集时间主要集中在荞麦出苗后 10 d 左右,具体采集时间为 2022 年 8 月 10 日和 2023 年 8 月 3 日。

将采集后的图像利用 PIX4Dmapper 软件生成正射影像图,图像格式为 TIFF,分辨率为 38 446 像素×14 183 像素。由于图像分辨率过高,无法采用深度语义分割模型进行训练预测,需对其进行分块处理。图像分块处理策略为将正射影像图按照从左到右,从上到下的顺序以 512 像素×512 像素的窗口进行分块,其滑动步长为 512 像素,所得子图之间无重叠部分,当滑动窗口到达正射影像边缘时,如子图尺寸不足 512 像素×512 像素,则以实际像素为准。图像分块策略能够在有效减少深度学习语义分割计算量的同时增加样本数据集,图像分块策略如图 2 所示。图像分块之后总共获得 2 128 幅图像,图像格式为 JPG。

1.3 数据增强

为了增强数据的多样性,采用水平翻转、垂直翻转、0°~360°随机翻转、颜色抖动等多种方式实现数据增强[13],数据增强效果如图 3 所示。通过数据增强共获得 8 456 幅图像,按照比例 7∶2∶1将数据集分为训练集(5 919 幅)、验证集(1 691 幅)和测试集(846 幅)。采用 Labelme 软件对荞麦图像进行标注,将荞麦植株部分标记为白色,将背景部分标记为黑色,标注结果文件类型为 JSON 文件,然后将其转换为 PNG 标注格式,其示例图如图 4 所示。



图 2 图像分块策略

Fig. 2 Image block strategy

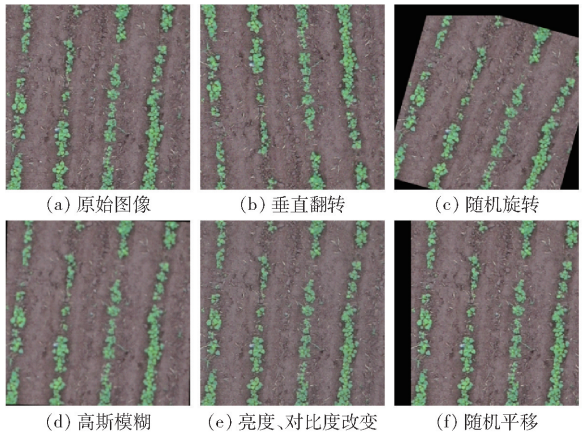

图 3 图像增强示例

Fig. 3 Examples of image enhancement

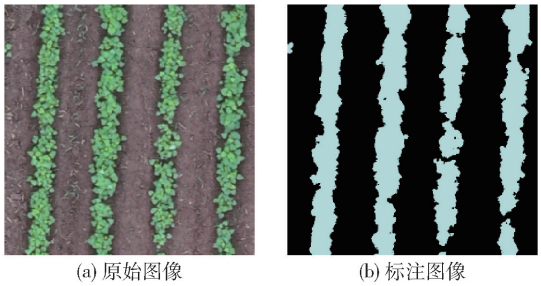

图 4 图像标注示例

Fig. 4 Examples of image annotation

1.4 改进的轻量化 DeepLabV3 + 无人机荞麦苗期分割模型

本研究以 DeepLabV3 + 模型为基础网络,提出一种改进的 DeepLabV3 + 方法实现对荞麦苗期图像的分割与识别。DeepLabV3 + 模型的工作原理是将图像的浅层特征和深层上采样特征进行融合,形成编码-解码工作机制,在关注图像边界信息的同时有

效提升了分割效果[14]；编码阶段主要包括特征提取主干网络和空间金字塔池化(Atrous spatial pyramid pooling，ASPP)模块，主干网络通常采用常见的主流框架，如 ResNet、Xception、MobileNet 和 ShuffleNet 等网络，通过主干网络可以完成特征学习优化和生成粗略特征图等功能，并采用 ASPP 结构采用不同大小的卷积获取高级抽象特征图，进行更深层次的特征采样[15]；解码阶段利用高维特征和低维特征，将像素位置信息与空间信息相结合，最后利用双线性插值上采样完成像素级分割[16]。DeepLabV3 + 模型将深度可分离卷积(Depth wise separable convolution，DSC)应用在了 ASPP 和编码模块中，在抑制模型参数的同时增加网络深度[17]。

DeepLabV3 + 模型有较好的分割效果，但由于其较为复杂的主干特征提取网络和深层次的网络结构，容易出现训练、部署困难等问题[18]，且难以实现移动平台的搭载。基于以上问题，本文设计一种改进的轻量级语义分割模型 DeepLabV3 + 用于荞麦苗期图像分割，采用 RepVGG 与 MobileViT 模块融合的方式进行特征提取代替原始主干网络，降低了模型参数量。为了更好地对特征通道之间的关系进行建模，在 RepVGG 网络模型中引入 SENet 注意力机制，合理利用通道间的相关性，捕获更多的全局语义信息。改进后 DeepLabV3 + 的模型网络结构如图 5 所示。改进后的轻量化 DeepLabV3 + 分割模型与无人机遥感荞麦数据集相结合，能够实现大规模种植时荞麦的精准识别，能够及时、准确地针对荞麦长势及产量做出预估，合理判断大田所处苗期对生育期水、肥、药的管理具有重要的指导意义，实现荞麦田间的智能化管理。

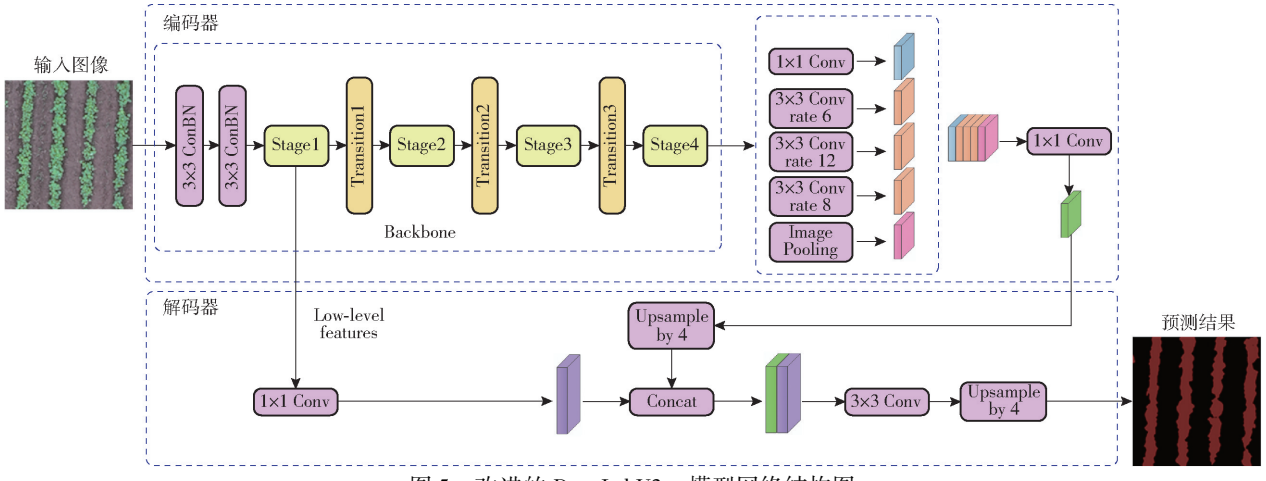

图 5　改进的 DeepLabV3 + 模型网络结构图

Fig. 5　Structure of improved DeepLabV3 + model

1.4.1　MobileViT 模块

MobileViT[19]是轻量级网络，传统的轻量级卷积神经网络(Light-wieght convlolutional neural networks，LWCNNs)主要利用空间归纳偏差能力使其在不同的视觉任务中需要学习的模型参数变少，但获取的特征为局部特征。为了能够获取全局特征，通常会通过添加 ViTs(Vision transformers)机制来实现，但该机制会增加模型参数，不满足轻量级的需求。而 MobileViT 能够将卷积神经网络(Convolutional neural networks，CNNs)和 ViTs 完美结合，形成一种全新的轻量化模型，其结构如图 6 所示。MobileViT 模块能够利用更少的参数实现对输入量中局部和全局信息的建模。

对于一个输入为 $H \times W \times C$ 的数据而言，首先通过一个卷积核为 3 × 3 的卷积层进行局部特征建模，然后通过一个卷积核大小为 1 × 1 的卷积层调整通道个数，接着为了提高其归纳偏置的全局表达能力，该模块在 Unfold 和 Fold 模块之间引入一个

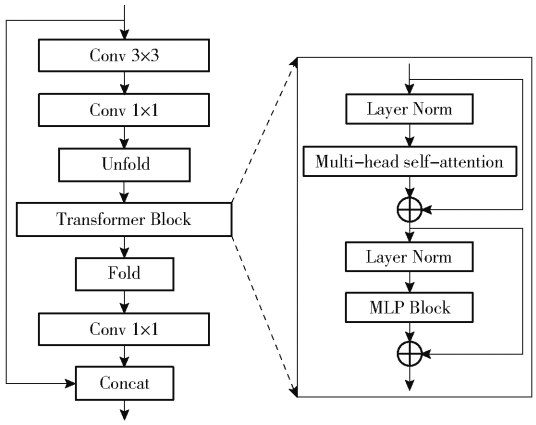

图 6　MobileViT Block 结构图

Fig. 6　Structure of MobileViT Block

Transformer Block，其中 Unfold 能够将所得到的特征映射到 N 个非重叠的 Patch，而 Fold 模块则可以将其恢复到原始特征维度，在完成全局特征建模后，利用卷积核大小为 1 × 1 的卷积层将通道调整为原始大小，并与原始输入特征图进行拼接，最终利用 3 ×

3 的卷积层实现特征融合。Transformer Block 结构由归一化层(Layer Norm)、多头自注意力机制(Multi - head self - attention)、多层感知机(Multilayer perceptron, MLP)模块构成。

1.4.2 SENet 注意力机制

在图像分类、图像分割以及目标检测等计算机视觉领域,注意力机制的应用较为广泛。在神经网络中加入注意力机制能够有效提高神经网络的识别性能。常见的注意力机制有 CBAM(Convolution block attention module)、SK(Selective kernel)和 SENet 注意力机制等。在荞麦苗期植株的语义分割过程中,特征提取非常重要。在传统的卷积神经网络架构中,通常采用卷积层和池化层来提取图像特征,但这种方法并没有明确地对特征通道之间的关系进行建模,导致一些通道对于特定任务的贡献相对较小。SENet 机制通过引入压缩(Squeeze)和激励(Excitation)操作来建立通道之间的关系,使得模型能够自适应地学习到每个通道的重要性,并且根据任务的需要加权调整特征图中的通道贡献,网络可以更加关注重要的特征通道,提高特征的判别能力,从而提升模型的性能[20]。RepVGG[21]是一种类 VGG 网络结构,它使用结构重参数化将训练时的多分支拓扑与推理时的纯体系结构分离,在图像分类和语义分割中表现出良好的速度精度折衷。但 RepVGG、Inception 都属于串联型 CNN 网络模型,输入数据的全局信息很难传达到模型尾部。这一部分全局信息的丢失,难免会影响模型性能。

因此,本文在 RepVGG 网络模型中引入 SENet 注意力机制来保证全局信息的传递,SE - RepVGG 结构如图 7 所示。假设输入特征图为 X,尺寸为 $H \times W \times C$,H、W 和 C 分别为图像高度、宽度和通道数。首先对输入特征进行 Squeeze 操作,其主要思路是进行全局平均池化(Global average pooling),将输入维度由 $H \times W \times C$ 变成 $1 \times 1 \times C$。然后进行 Excitation 操作,使用全连接层和非线性激活函数学习每个通道的权重,以捕捉通道之间的关系。具体采用 2 个全连接层(Fully connected, FC)实现,第 1 个全连接层将通道数 C 压缩为原来的 $1/r$ 来降低计算量,后接一个 ReLU 激活函数;第 2 个全连接层则将通道数恢复为 C,后接一个 Sigmod 激活函数,将获得的权重进行归一化处理。通过 2 个全连接层的特征提取能够将权重加权到每个通道的特征上,保证合理利用通道间的相关性,同时也极大地减少了参数量和计算量。最后,将所有加权的特征图重新组合起来,得到最终的输出特征图 $\tilde{X}$。

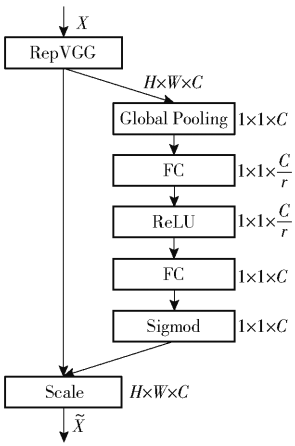

图 7 SE - RepVGG 结构

Fig. 7 Structure of SE - RepVGG

1.4.3 主干网络

本研究提出一种轻量化主干网络模型,网络模型以及详细参数如表 1 所示,主干网络由 2 个卷积模块、4 个 Stage 模块和 3 个 Transition 模块构成,其中 Stage 模块为轻量化主干网络的重要组成部分,主要采用 SE - RepVGG 与 MobileViT 结构融合的方式实现局部、全局特征提取。首先,采用 2 个步长为 2 的 3 × 3 卷积核 ConvBN 模块对输入数据进行初步特征提取,其特征图尺寸分别变为原始输入图像的 1/2 和 1/4。然后,通过 Stage 模块和 Transition 模块交替进行特征提取,其中 Stage 模块采用的卷积核步长为 1,使得输入前后的特征图宽和高没有发生改变;Transition 模块主要结构为步长为 1 的 3 × 3 卷积和步长为 2 的 3 × 3 卷积级联,分别连接 BatchNorm 和 ReLU,该模块处理后特征图尺寸缩小为输入的 1/2。

表 1 主干网络结构

Tab. 1 Structure of backbone network

模块	卷积核尺寸,步长	输入	输出
ConvBN	3 × 3,2	512 × 512 × 3	256 × 256 × 16
ConvBN	3 × 3,2	256 × 256 × 16	128 × 128 × 16
Stage1	1 × 1,1	128 × 128 × 16	128 × 128 × 32
Transition1	3 × 3,2	128 × 128 × 32	64 × 64 × 64
Stage2	1 × 1,1	64 × 64 × 64	64 × 64 × 64
Transition2	3 × 3,2	64 × 64 × 64	32 × 32 × 128
Stage3	1 × 1,1	32 × 32 × 128	32 × 32 × 128
Transition3	3 × 3,2	32 × 32 × 128	16 × 16 × 256
Stage4	1 × 1,1	16 × 16 × 256	16 × 16 × 256

1.4.4 损失函数

在图像分割任务中,每次学习迭代过程中通常利用损失函数来反向传播指导网络参数的优化,最终完成样本中像素的分类。损失函数通常用来评估模型的预测值和真实值(标注值)之间的不一致程度,损失函数值越小,则认为模型的分类性能越高、

鲁棒性越高。苗期荞麦植株和非荞麦植株无人机图像分割问题属于二分类问题。图像分割领域，常用的损失函数有交叉熵损失函数（Cross entropy loss，CE_Loss）、Dice_Loss 损失函数以及 Focal 损失函数（Focal_Loss）[22]。

本文将以上损失函数进行组合优化，对比分析不同损失函数组合在无人机荞麦苗期图像上的分割结果，寻找最优损失函数。

1.5　模型运行环境和评估参数

1.5.1　模型运行环境及参数

实验所用操作系统为 Windows 10，CPU 型号为 Intel（R） Core（TM） i7 - 12700F @ 2.10 GHz，GPU 型号为 NVIDIA GeForce RTX 3080，运行内存为 32 GB，1 TB 机械硬盘，编程语言为 Python 3.9，使用的深度学习框架为 Pytorch 1.13.0，GPU 加速库为 CUDA11.7、CUDNN8.4.1。输入图像分辨率为 512 像素 ×512 像素，训练优化器类型为 Adam，权重衰退指数为 0.000 1，初始化学习率为 0.000 5，Batch size 为 6，训练批次为 100。

1.5.2　模型评估参数

本文从分割性能和部署性能 2 方面对模型进行评估。分割性能指标通常包括平均像素准确率（Mean pixel accuracy，mPA）、mIoU、平均精确度（Mean precision，MP）和平均召回率（Mean recall，MR）。

部署性能指标包括模型内存占用量、总体参数量、浮点运算次数和推理速度。其中模型内存占用量为模型训练结束的实际存储空间（MB）；总体参数量为模型训练过程中需要训练参数的总个数，用于描述模型的空间复杂度；浮点运算次数为一种常用的衡量深度学习网络模型速度指标，用于描述模型的时间复杂度；推理速度为模型在每秒内可以进行图像分割操作的图像数量，单位为每秒传输帧数（f/s），通常认为每秒传输帧数越高，处理速度越快。

2　实验结果与分析

2.1　注意力机制选择结果

注意力机制主要添加在主干网络 Stage 模块的 RepVGG Block 结构中，本文主干网络共有 4 个 Stage，即 Stage1、Stage2、Stage3 和 Stage4。本文对比分析 SENet 添加位置对荞麦图像分割结果的影响，实验结果如表 2 所示。

从表 2 可知，当仅在 Stage4 位置添加 SENet 注意力机制时，其 mIoU、MP、MR 和 mPA 4 个评价参数分别为 89.71%、93.87%、95.12%、96.38%；同时在 Stage3 和 Stage4 阶段添加 SENet 时，与仅在 Stage4 位置添加 SENet 相比，4 个指标分别提高 1.74、0.87、1.13、0.64 个百分点，效果提高显著；当同时在 Stage2、Stage3 和 Stage4 添加 SENet 时，模型在增加网络参数的同时，效果不仅没有明显增加，反而出现了下降现象，仅 MP 提高 1.32 个百分点，而 mIoU、MR 和 mPA 指标则分别降低 1.46、2.90、0.66 个百分点，实验结果较差。通过对比分析，SENet 添加位置为 Stage3 和 Stage4 实验结果最佳。因此，本文在以上最佳位置添加不同注意力机制进行模型对比，以便得到最优模型。SENet 和 SK 注意力机制的实验结果表现相当。在网络模型最佳位置添加 SENet 时，其 mIoU 和平均召回率两个评价参数达到最优，而 SK 注意力机制则在 mPA 上表现最优，但二者模型大小差距较为明显，SENet 最佳位置模型内存占用量仅为 35.68 MB，而 SK 注意力机制最佳位置模型内存占用量高达 62.74 MB，差距较为明显。因此，通过注意力机制添加位置及不同注意力机制的对比实验可知，采用 SENet 注意力机制的分割模型参数更少，更适合在移动端进行实验部署。

表 2　注意力机制选择结果

Tab.2　Results of attention mechanism selection

注意力机制	网络结构位置				mIoU/%	MP/%	MR/%	mPA/%	模型内存占用量/MB
	Stage1	Stage2	Stage3	Stage4					
SENet				√	89.71	93.87	95.12	96.38	35.29
SENet			√	√	91.45	94.74	96.25	97.02	35.68
SENet		√	√	√	89.99	96.06	93.35	96.36	35.77
SENet	√	√	√	√	90.97	95.87	94.55	96.93	35.79
CBAM			√	√	89.99	95.68	93.66	96.60	35.69
SK			√	√	91.39	95.21	96.09	97.05	62.74

注：“√”表示添加此模块。

2.2　损失函数对比实验结果

损失函数的选择能够激发模型学习优化，在模型收敛速度方面具有至关重要的作用。因此，本文设计了 CE_Loss、Focal_Loss、CE_Loss + Dice_Loss、Focal_Loss + Dice_Loss 4 组损失函数组合，分析其对

模型训练性能的影响。4 组损失函数在训练集上的性能对比和损失值曲线如表 3 和图 8 所示。从表 3 和图 8 可知,单独采用 SE_Loss 损失函数和 Focal_Loss 损失函数时,其 mIoU、MP、MR 和 mPA 表现不佳,但采用 Focal_Loss 损失函数在迭代前期时,收敛速度较快;采用 CE_Loss + Dice_Loss 损失函数组合时,与单独使用 CE_Loss 损失函数相比,平均精确率有一定程度下降,其他参数均明显提升;采用 Focal_Loss + Dice_Loss 整体表现最佳,在迭代前期损失值较小且收敛速度最快,在迭代次数为 35 轮次时,损失值达到基本平稳,其 mIoU、MP、MR 和 mPA 分别为 91.45%、94.74%、96.25% 和 97.02%,各项指标表现最优。

表 3 不同损失函数组合对比结果

Tab. 3 Comparison of different loss functions %

模型	mIoU	MP	MR	mPA
CE_Loss	90.55	93.92	96.10	96.66
Focal_Loss	88.91	93.00	95.14	96.03
CE_Loss + Dice_Loss	90.64	94.26	95.81	96.72
Focal_Loss + Dice_Loss	91.45	94.74	96.25	97.02

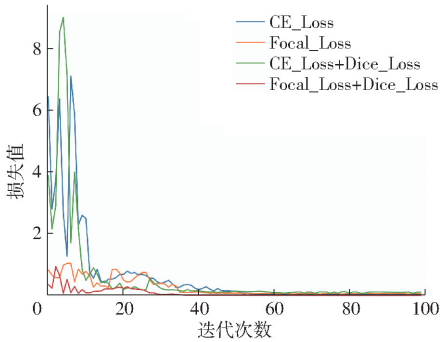

图 8 不同损失函数在训练集上的损失值曲线对比

Fig. 8 Training loss results of different loss functions

2.3 无人机荞麦苗期图像分割结果

在无人机荞麦苗期分割实验中,本研究采用网络默认初始参数进行模型预训练,并且在评估的基础上不断调整网络参数,根据训练结果最终确定最优训练参数。为了能够准确评估不同分割模型的分割性能,本研究将实验结果与 FCN[23]、PSPNet[24]、DenseASPP[25]、DeepLabV3、DeepLabV3 + 模型进行了对比分析,同时,对每个模型训练参数进行了统一。

不同模型分割结果如图 9 所示,分割结果图像中红色表示荞麦植株部分,黑色表示背景部分。图 9 中样本图像 1 出苗情况较为完整,样本图像 2 存在轻微缺苗情况,样本图像 3 缺苗现象较为严重,样本图像 4 田间存在杂草干扰情况。表 4 为不同分割模型在测试集中的性能结果。PSPNet、

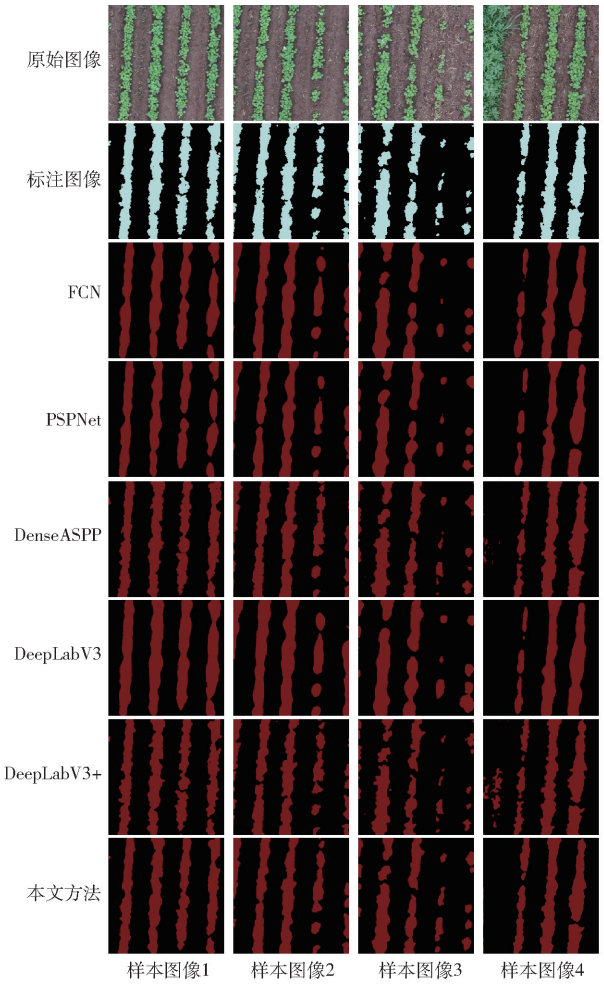

图 9 不同图像分割模型对比结果

Fig. 9 Comparison results of different image segmentation models

DeepLabV3、FCN 3 个网络结构的主干网络采用 Resnet 结构实现,模型内存占用量分别达到 190.36、163.18、137.07 MB,总体参数量和浮点运算次数也相对较大,其中总体参数量分别为 4.658×10^7、3.963×10^7 和 3.295×10^7,而浮点运算次数则分别为 1.776×10^{11}、$1.641\,2\times10^{11}$ 和 $1.388\,6\times10^{11}$,推理速度均较慢,分别为 9.85、10.66、10.10 f/s。DeepLabV3 模型的 mPA 和 mIoU 值分别为 93.71% 和 83.84%,PSPNet 模型的 mPA 和 mIoU 值分别为 92.00%、84.44%,FCN 模型的 mPA 和 mIoU 值分别为 92.00%、85.83%,3 种常见模型在自建无人机荞麦数据集上表现一般,模型总体参数量和浮点运算次数都较大,推理速度较慢。DenseASPP 分割模型由于其特殊的网络结构,主干网络采用 densenet121 实现,其 mPA 和 mIoU 值分别为 95.00%、89.36%,模型相对较小,为 40.58 MB,总体参数量和浮点运算次数较少,分别为 9.17×10^6 和 8.309×10^{10},相对于前 3 种模型在模型总体参数量和浮点运算次数等方面有较明显的提升,实现了模型轻量化,但其推理

表 4　不同分割模型在测试集中的性能结果

Tab. 4　Performance of different segmentation models

模型	mIoU/%	mPA/%	模型内存占用量/MB	总体参数量	FLOPs	推理速度/(f·s^{-1})
FCN	85.83	92.00	137.07	3.295×10^{7}	$1.388\,6\times10^{11}$	10.10
PSPNet	84.44	92.00	190.36	4.658×10^{7}	$1.776\,0\times10^{11}$	9.85
DenseASPP	89.36	95.00	40.58	9.170×10^{6}	$8.309\,0\times10^{10}$	7.92
DeepLabV3	83.84	93.71	163.18	3.963×10^{7}	$1.641\,2\times10^{11}$	10.66
DeepLabV3 +	91.33	96.96	22.97	5.810×10^{6}	$7.342\,0\times10^{10}$	34.71
本文模型	91.45	97.02	35.68	9.010×10^{6}	$8.215\,0\times10^{10}$	37.83

速度较慢,仅为 7.92 f/s。DeepLabV3 + 模型的主干网络采用 MobileNetV2 轻量化网络,其 mPA 和 mIoU 值分别为 96.96%、91.33%,其模型内存占用量为 22.97 MB,总体参数量、浮点运算次数和推理速度均有明显提升,表现良好,但是在实际数据分析中发现,由于编码阶段采用轻量化主干网络,其特征提取细节不足,存在将杂草误认为是荞麦植株的情况。本文提出的改进 DeepLabV3 + 方法的 mPA 和 mIoU 值分别 97.02%、91.45%,模型内存占用量为 35.68 MB,总体参数量和浮点运算次数分别为 9.01×10^{6} 和 8.215×10^{10},推理速度为 37.83 f/s,单幅图像推理时间为 26.46 ms,在模型准确率、轻量化和推理速度等方面综合表现最优。

2.4　荞麦地块全尺寸分割结果

本文最终的目标是能够在移动端实现荞麦图像的实时分割,因此,需要实现全尺寸荞麦苗期图像的分割与识别以增强模型的鲁棒性。由于无人机单次采集图像分辨率为 5 472 像素 × 3 648 像素,图像分辨率较高,需采用 1.2 节中图像分块方法对原始图像进行滑动分块,滑动步长为 512 像素,滑动分块窗口尺寸为 512 像素 × 512 像素,分块数量为 88 块(其中 512 像素 × 512 像素图像 70 块,352 像素 × 512 像素图像 7 块,512 像素 × 64 像素图像 10 块,352 像素 × 64 像素图像 1 块),子图之间无重叠部分;然后对切割得到的图像进行预测,将预测后的结果进行拼接,最终得到全尺寸荞麦苗期分割图像。不同的飞行高度能够获取实验田不同尺度的信息,飞行高度较低时,能够提供荞麦实验田苗期的更多细节,而飞行高度较高时,则能够为荞麦实验田提供更多的广视角的宏观信息。采用无人机进行荞麦苗期状态的评估,主要就是评估实验田的整体情况,因此需要验证本文训练模型针对不同飞行高度的荞麦苗期分割情况,具体实验结果如图 10 所示,针对不同高度性能对比如表 5 所示。

不同飞行高度采用的模型为同一训练模型,其模型内存占用量、总体参数量和浮点运算次数均一致,因此,本文选取 mIoU、mPA 和推理速度 3 个参数

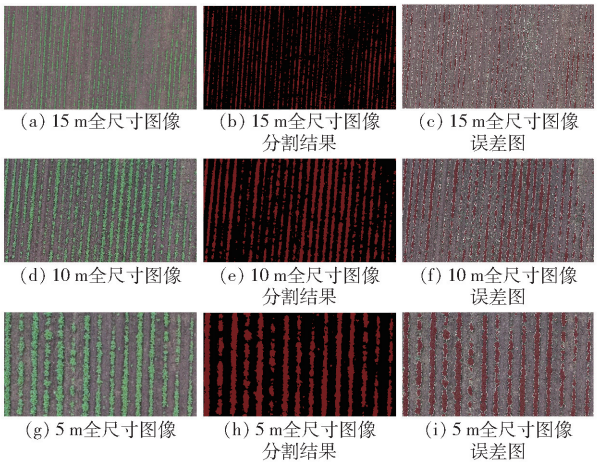

图 10　改进 DeepLabV3 + 模型对不同飞行高度全尺寸图像分割结果

Fig. 10　Improved DeepLabV3 + full-size image segmentation results in different flight altitudes

表 5　改进 DeepLabV3 + 针对不同飞行高度的性能对比

Tab. 5　Performance comparison of improved DeepLabV3 + for different flight altitudes

飞行高度/m	mIoU/%	mPA/%	推理速度/(f·s^{-1})
5	83.15	91.42	38.04
10	91.45	97.02	37.91
15	90.21	96.85	35.57

来实现算法运行效率的对比。图 10a 和图 10b 为飞行高度为 15 m 时的全尺寸图像及其分割结果,图 10c 为对应的误差结果图,该图由分割结果叠加到原图上所得,能够更加直观地看出分割结果。模型在飞行高度 15 m 的 mPA 和 mIoU 分别为 96.85% 和 90.21%,由于飞行高度升高,图像中所含的荞麦苗期信息较多,其推理速度相对较慢,为 35.57 f/s;图 10d 和图 10e 为飞行高度为 10 m 时的全尺寸图像及其分割结果,其推理速度为 37.91 f/s;图 10g 和图 10h 分别为飞行高度为 5 m 时的图像和分割结果,由于模型训练采用的训练集均为 10 m 图像数据,其 mPA 和 mIoU 下降较为明显,分别下降 5.6、8.3 个百分点,但也能明显识别荞麦苗位置,而由于飞行高度较低,图像中所包含的苗期信息较少,推理速度相对较快,达到 38.04 f/s。通过实验对比验

证,本文模型能够实现不同高度下全尺寸荞麦苗期图像的分割,识别结果较好,验证了模型的鲁棒性,同时也能实现无人机广视角图像的整体苗期分割。

3 结论

(1)随着荞麦种植面积的不断提高,利用科学手段有效检测荞麦生长过程中的多项生理指标及生长形态,对荞麦的生长发育具有重要的意义。在荞麦苗期利用无人机技术进行准确的作物植株提取,可对田间苗期的出苗率、重播率、漏播率以及均匀度进行有效监测,据此评估农田的播种质量并为产量估测提供参考,为小杂粮产业智能农业建设提供理论基础。

(2)以荞麦苗期图像分割为例,针对原始 DeepLabV3 + 语义分割模型参数量较大、难以在移动端部署等问题进行改进,在编码阶段采用 RepVGG Block 与 MobileViT Block 模块融合的方式作为主干网络实现局部特征和全局特征的提取,同时为了提高通道间的特征提取效率,加入 SENet 注意力机制,构建了一种全新的轻量化语义分割模型。与 FCN、PSPNet、DenseASPP、DeepLabV3、DeepLabV3 + 模型相比,本文改进模型的 mPA 和 mIoU 值分别为 97.02%、91.45%,总体参数量、浮点运算次数和推理速度分别为 9.01×10^6、8.215×10^{10}、37.83 f/s,综合表现最优。

(3)改进后的算法在模型参数量上以及分割准确率上效果都较好,但仍然存在不足之处。由于注意机制的引入在一定程度上增加了模型复杂性,因此在模型复杂性和参数量方面还需进一步研究。

参 考 文 献

[1] 陈善雄,伍胜,于显平,等. 基于卷积神经网络结合图像处理技术的荞麦病害识别[J]. 农业工程学报, 2021, 37(3): 155 - 163.
CHEN Shanxiong, WU Sheng, YU Xianping, et al. Buckwheat disease recognition using convolution neural network combined with image processing[J]. Transactions of the CSAE, 2021, 37(3): 155 - 163. (in Chinese)

[2] 吕少中,杜文亮,陈震,等. 基于机器视觉的荞麦剥壳性能参数在线检测方法[J]. 农业机械学报, 2019, 50(10): 35 - 43.
LÜ Shaozhong, DU Wenliang, CHEN Zhen, et al. On-line measuring method of buckwheat hulling efficiency parameters based on machine vision[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(10): 35 - 43. (in Chinese)

[3] XU S, ZHANG Y, DONG W, et al. Early identification and localization algorithm for weak seedlings based on phenotype detection and machine learning[J]. Agriculture, 2023, 13(1): 212.

[4] ZOU K, CHEN X, WANG Y, et al. A modified U - Net with a specific data argumentation method for semantic segmentation of weed images in the field[J]. Computers and Electronics in Agriculture, 2021, 187: 106242.

[5] KONG S, LI J, ZHAI Y, et al. Real-time detection of crops with dense planting using deep learning at seedling stage[J]. Agronomy, 2023, 13(6): 1503.

[6] DIAO Z, GUO P, ZHANG B, et al. Maize crop row recognition algorithm based on improved UNet network[J]. Computers and Electronics in Agriculture, 2023, 210: 107940.

[7] LIU X, QI J, ZHANG W, et al. Recognition method of maize crop rows at the seedling stage based on MS - ERFNet model[J]. Computers and Electronics in Agriculture, 2023, 211: 107964.

[8] SUN J, ZHOU J, HE Y, et al. RL - DeepLabv3 +: a lightweight rice lodging semantic segmentation model for unmanned rice harvester[J]. Computers and Electronics in Agriculture, 2023, 209: 107823.

[9] ZUO X, LIN H, WANG D, et al. A method of crop seedling plant segmentation on edge information fusion model[J]. IEEE Access, 2022, 10: 95281 - 95293.

[10] ZHANG P, SUN X, ZHANG D, et al. Lightweight deep learning models for high-precision rice seedling segmentation from UAV-based multispectral images[J]. Plant Phenomics, 2023, 5: 0123.

[11] WAN H, ZENG X, FAN Z, et al. U2ESPNet - a lightweight and high-accuracy convolutional neural network for real-time semantic segmentation of visible branches[J]. Computers and Electronics in Agriculture, 2023, 204: 107542.

[12] ZHANG Y, LV C. Tinysegformer: a lightweight visual segmentation model for real-time agricultural pest detection[J]. Computers and Electronics in Agriculture, 2024, 218: 108740.

[13] ZOU K, LIAO Q, ZHANG F, et al. A segmentation network for smart weed management in wheat fields[J]. Computers and Electronics in Agriculture, 2022, 202: 107303.

[14] WANG C, DU P, WU H, et al. A cucumber leaf disease severity classification method based on the fusion of DeepLabV3 + and U - Net[J]. Computers and Electronics in Agriculture, 2021, 189: 106373.

[15] PENG H, ZHONG J, LIU H, et al. ResDense - focal - DeepLabV3 + enabled litchi branch semantic segmentation for robotic harvesting[J]. Computers and Electronics in Agriculture, 2023, 206: 107691.

[16] HE X, XUAN S, WANG K, et al. DeepLabV3 + image segmentation algorithm fusing cumulative distribution function and channel attention mechanism[J]. Journal of Computer Applications, 2023, 43(3): 936.

[17] 曹英丽,赵雨薇,杨璐璐,等. 基于改进 DeepLabV3 + 的水稻田间杂草识别方法[J]. 农业机械学报, 2023, 54(12): 242 - 252.
CAO Yingli, ZHAO Yuwei, YANG Lulu, et al. Weed identification method in rice field based on improved DeepLabV3 + [J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(12): 242 - 252. (in Chinese)
[18] 顾文娟,魏金,阴艳超,等. 基于改进 DeepLabv3 + 的番茄图像多类别分割方法[J]. 农业机械学报,2023,54(12): 261 - 271.
GU Wenjuan, WEI Jin, YIN Yanchao, et al. Multi class segmentation method for tomato images based on improved DeepLabv3 + [J]. Transactions of The Chinese Society for Agricultural Machinery, 2023, 54(12): 242 - 252. (in Chinese)
[19] PEI Y, LIU Y, LING N. MobileViT - GAN: a generative model for low bitrate image coding[C]//2023 IEEE International Conference on Visual Communications and Image Processing (VCIP), 2023.
[20] FANG M, LIANG X, FU F, et al. Attention mechanism based semi-supervised multi-gain image fusion[J]. Symmetry, 2020, 12(3): 451.
[21] WANG Z, DONG L, ZHANG M, et al. Remu - Net: multi-branch net framework for 3D brain tumor segmentation[J]. Journal of Circuits, Systems and Computers, 2022,31(12):2250215.
[22] 毛昊,李新利,王孝伟,等. 基于多类别 Focal Loss 损失函数的变电站场景图像语义分割研究[J]. 华北电力大学学报(自然科学版), 2022, 49(5):84 - 92.
MAO Hao, LI Xinli, WANG Xiaowei, et al. Research on semantic segmentation of transformer substation image based on multi-category Focal Loss function[J]. Journal of North China Electric Power University (Natural Science Edition), 2022, 49(5):84 - 92. (in Chinese)
[23] ZHAO C. A novel semantic segmentation method based on improved FCN for unmanned driving[C]//2023 IEEE 6th International Conference on Automation, Electronics and Electrical Engineering (AUTEEE), 2023.
[24] 武花,张新长,孙颖,等. 融合多特征改进型 PSPNet 模型应用于复杂场景下的建筑物提取[J]. 测绘通报, 2021(6):21 - 27.
WU Hua, ZHANG Xinchang, SUN Ying, et al. Building extraction in complex scenes based on the fusion of multi-feature improved PSPNet model[J]. Bulletin of Surveying and Mapping, 2021(6): 21 - 27. (in Chinese)
[25] WU K, ZHANG S, XIE Z. Monocular depth prediction with residual DenseASPP network[J]. IEEE Access, 2020(99): 1.

(上接第 185 页)

[32] 王楠. 乡村空间重构背景下的聚落演变与调控研究——以河北省为例[D]. 北京:中国农业大学,2022.
WANG Nan. Research on the evolution and regulation of settlements under the background of rural spatial reconstructing: a case study of Hebei Province[D]. Beijing:China Agricultural University,2022. (in Chinese)
[33] 海贝贝,李小建,许家伟. 巩义市农村居民点空间格局演变及其影响因素[J]. 地理研究,2013,32(12):2257 - 2269.
HAI Beibei,LI Xiaojian,XU Jiawei. Spatio-temporal evolution of rural settlements in Gongyi[J]. Geographical Research, 2013,32(12):2257 - 2269. (in Chinese)
[34] 徐州,林孝松,余情,等. 巫山县农村居民点分布与地形因子关系[J]. 水土保持研究,2018,25(4):338 - 343.
XU Zhou,LIN Xiaosong,YU Qing,et al. Distribution of rural settlements in relation to landform factors in Wushan County[J]. Research of Soil and Water Conservation,2018,25(4):338 - 343. (in Chinese)
[35] 姜广辉,张凤荣,秦静,等. 北京山区农村居民点分布变化及其与环境的关系[J]. 农业工程学报,2006,22(11):85 - 92.
JIANG Guanghui,ZHANG Fengrong,QIN Jing,et al. Relationship between distribution changes of rural residential land and environment in mountainous areas of Beijing[J]. Transactions of the CSAE,2006,22(11):85 - 92. (in Chinese)
[36] 钟国敏,张学雷,段金龙,等. 郑州市土壤多样性和土地利用多样性研究及关联分析[J]. 土壤通报,2013,44(3):513 - 521.
ZHONG Guomin,ZHANG Xuelei,DUAN Jinlong,et al. Evaluation and correlative analysis of pedodiversity and land use diversity in Zhengzhou[J]. Chinese Journal of Soil Science,2013,44(3):513 - 521. (in Chinese)
[37] 徐咏文,段萍,罗志华. 浅析中国土壤分类的发生与现状[J]. 安徽农业科学,2005,33(10):2003 - 2004.
[38] 李冬梅,王冬艳,李红. 城市等级对农村居民点演变的影响[C]//2016 中国新时期土地资源科学与新常态创新发展战略研讨会暨中国自然资源学会土地资源研究专业委员会 30 周年纪念会论文集,2016:255 - 261.
[39] 李晓江,尹强,张娟,等.《中国城镇化道路、模式与政策》研究报告综述[J]. 城市规划学刊,2014(2):1 - 14.
LI Xiaojiang,YIN Qiang,ZHANG Juan,et al. The path, model and policy of urbanization in China[J]. Urban Planning Forum,2014(2):1 - 14. (in Chinese)
[40] 王楠,郝晋珉,张金懿. 河北省乡村聚落的历史起源与发展研究[J]. 中国农业大学学报,2023,28(12):214 - 223.
WANG Nan,HAO Jinmin,ZHANG Jinyi. Research on the historical origin and development of rural settlement in Hebei Province[J]. Journal of China Agricultural University,2023,28(12):214 - 223. (in Chinese)
[41] 王楠,郝晋珉,星野敏,等. 黄淮海平原农村产业结构与用地结构关系研究——以曲周县为例[J]. 北京大学学报(自然科学版),2023,59(5):801 - 812.
WANG Nan,HAO Jinmin, HOSHINO Satoshi,et al. Study on the relationship between industrial structure and land use structure of rural areas in Huang - Huai - Hai Plain: a case study of Quzhou County[J]. Acta Scientiarum Naturalium Universitatis Pekinensis,2023,59(5):801 - 812. (in Chinese)