

文章编号:1007-2780(2024)08-1001-13

基于轻量化的 DeepLabV3+ 遥感图像 地物分割方法

马 静^{1,2}, 郭中华^{1,2*}, 马志强¹, 马小艳^{1,2}, 李迦龙^{1,2}

(1. 宁夏大学 电子与电气工程学院, 宁夏 银川 750021;
2. 宁夏大学 沙漠信息智能感知重点实验室, 宁夏 银川 750021)

摘要: 针对 DeepLabV3+ 在遥感图像地物分割中出现的细节信息丢失、类别不均衡等问题引起的误差, 提出一种基于轻量化网络的 DeepLabV3+ 遥感图像地物分割方法。首先, 使用 MobileNetV2 替换原始基准网络中的骨干网络, 提高训练效率并减少模型的复杂度。其次, 增大 ASPP 结构中空洞卷积的膨胀率, 并在 ASPP 最后一层使用最大池化, 有效地捕获不同尺度的上下文信息, 同时在 ASPP 每个分支中引入 SE 注意力机制, 并在提取浅层特征之后引入 ECA 注意力机制, 提高模型对不同类别和细节的感知能力。最后, 使用加权的 Dice-Focal 联合损失函数进行优化, 处理类别不均衡的问题。将改进的模型分别在 CCF 和 华为昇腾杯竞赛数据集上进行验证, 实验结果表明, 本文所提出的方法相较于原始 DeepLabV3+ 模型在两种测试集上的各个指标均有不同程度的提高。其中, mIoU 达到了 73.47%、63.43%, 分别提高了 3.24% 和 15.11%; 准确率达到了 88.28%、86.47%, 分别提高了 1.47% 和 7.83%; F1 指数达到了 84.29% 和 77.04%, 分别提高了 3.86% 和 13.46%。改进后的 DeepLabV3+ 模型可以更好地解决细节信息丢失和类别不均衡的问题, 提高遥感图像地物分割的性能和准确性。

关键词: MobileNetV2; 空洞卷积; 注意力机制; 损失函数

中图分类号: TP391 文献标识码: A doi: 10.37188/CJLCD.2023-0293

Remote sensing image land feature segmentation method based on lightweight DeepLabV3+

MA Jing^{1,2}, GUO Zhonghua^{1,2*}, MA Zhiqiang¹, MA Xiaoyan^{1,2}, LI Jialong^{1,2}

(1. School of Physics and Electronic-Electrical Engineering, Ningxia University,
Yinchuan 750021, China;
2. Ningxia Key Lab on Information Sensing & Intelligent Desert, Ningxia University,
Yinchuan 750021, China)

Abstract: A lightweight network based DeepLabV3+ remote sensing image land feature segmentation method is proposed to address the errors caused by the loss of detail information and imbalanced categories in remote sensing image segmentation. Firstly, MobileNetV2 is adopted to replace the backbone network in original baseline network to improve training efficiency and reduce model complexity. Secondly, the

收稿日期: 2023-09-06; 修订日期: 2023-10-11.

基金项目: 国家自然科学基金 (No.62365016); 中央支持地方专项资金 (No.2023FRD05034)

Supported by National Natural Science Foundation of China (No.62365016); Central Support for Local Special Funds (No.2023FRD05034)

*通信联系人, E-mail: guozhh@nxu.edu.cn

dilation rate of atrous convolutions within ASPP structure is increased and max-pooling in final ASPP layer is incorporated to effectively capture context information at different scales. At the same time, SE attention mechanism is introduced into each branch of ASPP, and ECA attention mechanism is introduced after extracting shallow features to improve the model's perception ability for different categories and details. Finally, the weighted Dice-Local joint loss function is used for optimization to address class imbalance issues. The improved model is validated on both the CCF and Huawei Ascend Cup competition datasets. Experimental results show that the proposed method outperforms original DeepLabV3+ model on both test sets, with various metrics showing different degrees of improvement. Among them, mIoU reaches 73.47% and 63.43%, representing improvements of 3.24% and 15.11%, respectively. The accuracy reaches 88.28% and 86.47%, showing enhancements of 1.47% and 7.83%, respectively. The F1 index reaches 84.29% and 77.04%, increasing by 3.86% and 13.46%, respectively. The improved DeepLabV3+ model can better solve the problems of loss of detail information and class imbalance, which improves the performance and accuracy of remote sensing image feature segmentation.

Key words: MobileNetV2; dilated convolution; geometric figure; attention mechanism; loss function

1 引 言

遥感图像地物分割^[1]是遥感图像处理和分析中的一个重要研究领域,其应用范围涵盖了农业、城市规划、环境监测等多个领域^[2-3]。遥感图像地物分割的目标是将遥感图像中不同的地物分隔开,以便进一步处理和分析遥感图像。然而,由于遥感图像数据的复杂性和高维性以及地物分割的多样性和复杂性^[4],遥感图像地物分割一直是一个具有挑战性的问题。早期的遥感图像地物分割主要采用基于阈值^[5]和基于区域生长^[6]等传统图像处理方法,这些方法在处理简单场景时效果较好,但在处理复杂多变的遥感图像时缺乏灵活性,难以捕获到地物的细节信息,导致分割效果有限。

近年来,深度学习技术迅猛发展,为遥感图像地物高精度分割的研究和应用带来了新的希望。深度学习模型可以通过学习大量的训练数据自动提取遥感图像中的地物特性,实现对地物的准确分割^[7]。最早应用于遥感图像地物分割的深度学习模型是基于神经卷积网络(Convolutional Neural Network, CNN)的方法。Long等人^[8]提出的全卷积网络(Fully Convolutional Network, FCN)将传统的全连接层替换为卷积层,可以处理任意大小的输入图像^[9],并生成与输入图像相同大小的像素级别的预测结果,更好地处理遥感图像中的复杂地物。但卷积操作会导致

图像的分辨率下降,忽略了像素之间的联系性^[10]。Ronneberger等人^[11]提出了Unet网络,采用包含编码器和解码器的U形结构,引入上采样和跳跃连接^[12]的概念,并在解码器部分使用反卷积层来进行上采样,以提高分割结果的精度。由于采用了跳跃连接,该模型中包含了大量的参数和计算量^[13],导致模型庞大。Vijay等人^[14]提出了SegNet网络,采用与Unet相似的编码器-解码器结构,利用最大池化层进行特征提取并引入交叉熵损失函数^[15]和权重衰减^[16]等优化策略,提高了网络训练的效果。但该模型在分割小目标时会出现边界模糊的问题。PSPNet是由Zhao等人^[17]提出的一种卷积神经网络,采用金字塔池化模块^[18]来捕捉不同大小的上下文信息,但对于分割目标的形状和大小变异较大的情况,会出现分割错误或漏检现象。Chen等人^[19]在ECCV会议上提出了DeepLabV3+网络,采用空洞卷积^[20]和多尺度上下文模块^[21]来提高遥感图像地物分割的准确度,但其存在细节信息丢失、计算复杂性和类别不均衡等问题。沈言善等人^[22]提出了DL-Unet网络,引入扩张卷积与交叉熵损失函数对遥感图像不同类别的地物进行有效分割。周羿等人^[23]在DeepLabV3+中引入了金字塔拆分注意力模块(Pyramid Split Attention, PSA)和密集空间金字塔池化模块(Dense Atrous Spatial Pyramid Pooling, DenseASPP)来捕获关键信息。

上述学者提出的模型在遥感图像地物分割中存在细节信息丢失和类别不均衡等问题。本文针对这些问题引起的误差,对DeepLabV3+原始模型进行以下改进:(1)替换骨干网络。将原始基准网络中的Xception骨干网络替换为轻量级的骨干网络MobileNetV2,以便快速进行特征提取并处理遥感图像大规模数据集。(2)改进ASPP模块。增大ASPP结构中空洞卷积的膨胀率并在ASPP中的最后一层引用最大池化,同时每个分支后引入SE(Squeeze and Excitation)注意力机制,扩大感受野,提高深层特征信息的利用率。(3)在解码区引入ECA(Efficient Channel Attention)注意力机制,提高浅层特征信息的利用率。(4)使用加权的Dice-Focal联合损失函数进行优化,处理类别不均衡问题。基于以上策略对模型进行改进,并在CCF和华为昇腾

杯高分辨遥感影像数据集上验证改进模型分割的性能。

2 DeepLabV3+网络模型

2.1 DeepLabV3+模型

DeepLabV3+的基础模型如图1所示,采用了Xception^[24]骨干网络的编码器和解码器结构处理原始图像。首先,骨干网络提取不同尺度的特征图(1/16、1/8、1/4和1/2)。然后,使用深度可分离卷积对得到的特征图进行融合。在经过特征金字塔网络处理后,这些特征图输出到ASPP结构中。ASPP层使用不同采样率的空间卷积捕获不同尺度下的空间信息。最后,ASPP层进行全局平均池化和1×1卷积操作。解码器将ASPP输出4倍上采样,与骨干网络特征图融合,经过3×3卷积和4倍上采样输出最终结果。

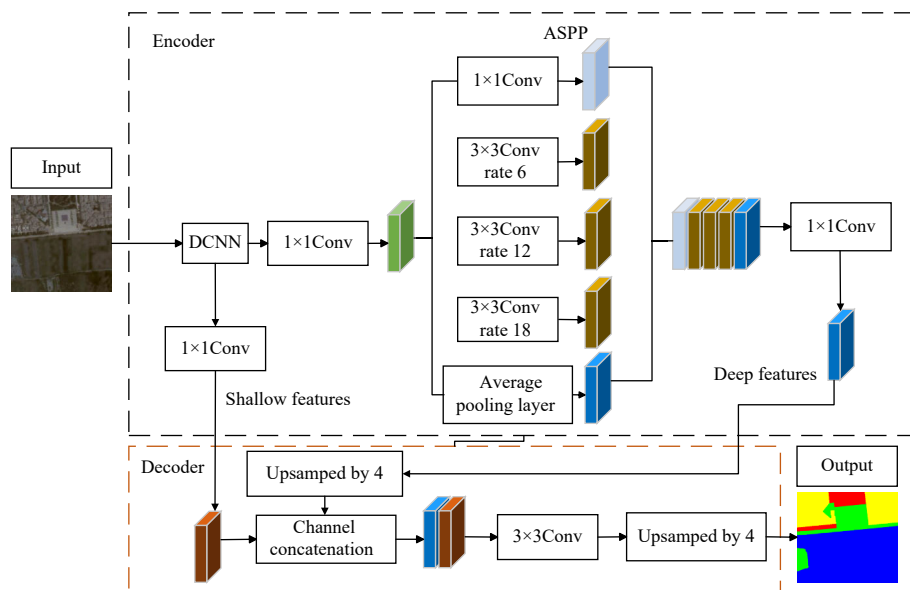


图1 DeepLabV3+模型结构图

Fig. 1 Structure diagram of DeepLabV3+ model

传统卷积操作使用固定的卷积核尺寸在每个位置上对输入特征图进行局部感受野的计算。空洞卷积引入了一个称为膨胀率(Dilation Rate)的参数,用于在卷积核中插入额外的间隔。假设以一个变量 l 来衡量空洞卷积的扩张系数,则加入空洞之后的实际卷积核尺寸与原始卷积核尺寸之间的关系为: $K_l = K + (k - 1)(l - 1)$, $l = 1, 2, 4$,如

图2所示。原始卷积核大小为 3×3 ,图2(a)中膨胀率为1,卷积核中的元素之间没有间隔,可视为普通卷积;图2(b)中膨胀率为2,对应的感受野变为 7×7 ;图2(c)中膨胀率为4,对应的感受野变为 15×15 。空洞卷积可以在不增加参数和计算量的情况下增加卷积层的感受野,从而增强模型对图像中不同尺度上下文信息的感知能力。

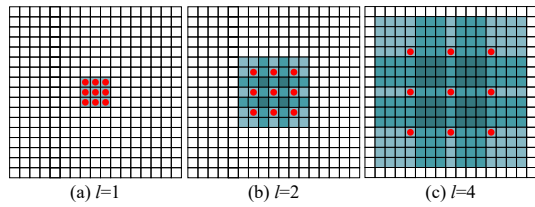


图 2 不同膨胀率的感受野对比图

Fig. 2 Comparison chart of receptive fields with different expansion rates

2.2 深度可分离卷积

Xception 是一种深度卷积神经网络模型,它是由 Google 在 2017 年提出的,旨在提高 Inception 模型^[25]的性能。Xception 的设计思想是将 Inception 模型中的卷积层和深度可分离卷积层^[26]完全分离,从而提高模型的表达能力和计算效率。具体来说,在传统的卷积神经网络中,卷积层既负责通道之间的特征混合,也负责空间特征的提取,这样的设计在一定程度上限制了模型对特征的表达能,并且在计算效率上也存在瓶颈。为了解决这个问题,Xception 提出了深度可分离卷积,如图 3 所示,其将通道之间的特征混合与空间特征的提取完全分离。这种操作方式相比于传统的卷积操作,可以有效提高模型的表达能,并且大幅减少了计算量和参数数量,提高了模型的计算效率。

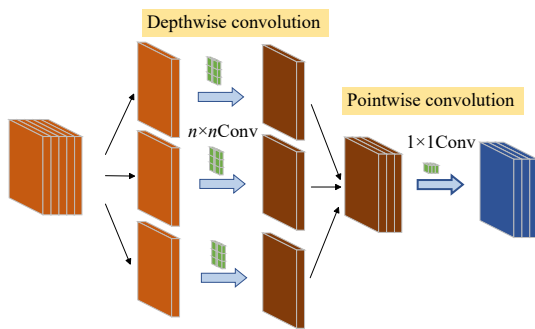


图 3 深度可分离卷积结构图

Fig. 3 Deeply separable convolutional structure diagram

3 实验数据及改进方法

3.1 实验数据与数据处理

本文实验使用 2020 年发布的 CCF 遥感影像地块分割和 华为昇腾杯 AI+ 遥感影像竞赛数

据集。CCF 数据集为中国南方某地区的高分辨率遥感影像,每个像素的分辨率为 2 m,成像波段包括 R、G、B 波段。数据覆盖 7 类地物类型,包括建筑、耕地、林地、水域、道路、草地和其他。为了使各类地物占比平衡,将林地、草地归为一类,将道路归为其他类。将数据裁剪为 256×256 大小,通过缩放、旋转等方法进行数据增强,挑选 10 916 张图片作为样本。华为昇腾杯 AI+ 遥感影像数据为 0.1~4 m 分辨率的高分一、二、六号,高景二号,北京二号以及部分航空等数据源的可见光、多光谱载荷图像,成像波段包括 R、G、B 波段,数据覆盖包含 9 类地貌:水体、交通运输、建筑、耕地、草地、林地、裸地和其他。为了使各类地物占比平衡,将耕地、草地、林地归为植被类。将数据裁剪为 512×512 大小,经过数据增强,挑选 9 736 张图片作为样本。将两类数据均以 9:1 的比例划分为训练集和测试集,信息详见表 1。

表 1 数据集详情

Tab. 1 Dataset details

数据集	分辨率/m	波段	标注类别
CCF2020	2	RGB	其他、耕地、林地、水域、建筑
华为昇腾杯	0.1~4	RGB	其他、水体、交通运输、建筑、植被、裸土

3.2 轻量化网络 MobileNetV2

鉴于 Xception 网络计算复杂度较高、训练和推理时间较长等缺陷,本文选择将 DeepLabV3+ 的骨干网络 Xception 替换为 MobileNetV2。MobileNetV2 是由 Google 团队提出的一种轻量级卷积神经网络^[27]架构,具有较低的计算复杂度和内存占用,同时具备更快的训练和推理速度,网络结构如表 2 所示。

MobileNetV2 主要采用反向线性残差结构,其包括反向残差结构(Inverted Residuals)和线性瓶颈层(Linear Bottlenecks)两个关键性技术,如图 4 所示。这种设计使得 MobileNetV2 能够在保持高性能的同时,具有较小的模型尺寸和较低的计算复杂度。

3.3 注意力机制

本文在编码区的 ASPP 结构分支中引入 SE

表 2 MobileNetV2 网络结构

Tab. 2 MobileNetV2 network structure

Input	Operator	t	c	n	s
$2\ 242 \times 3$	conv2d	—	32	1	2
$1\ 122 \times 32$	bottleneck	1	16	1	1
$1\ 122 \times 16$	bottleneck	6	24	2	2
562×24	bottleneck	6	32	3	2
282×32	bottleneck	6	64	4	2
142×64	bottleneck	6	96	3	1
142×96	bottleneck	6	160	3	2
72×160	bottleneck	6	320	1	1
72×320	conv2d 1×1	—	1 280	1	1
$72 \times 1\ 280$	avgpool 7×7	—	—	1	—
$1 \times 1 \times 1\ 280$	conv2d 1×1	—	k	—	—

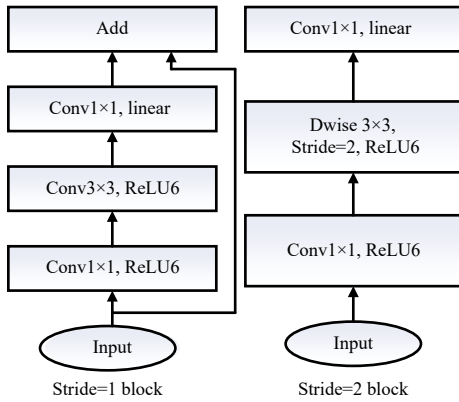


图 4 MobileNetV2 残差块结构

Fig. 4 Residual block structure of MobileNetV2

注意力机制,通过 SEBlock 特征进行通道注意力加权,帮助网络自适应地选择和强调最具有信息量的通道特征。在骨干网络提取浅层特征之后,添加 ECA 注意力机制,通过学习通道之间的相关性来提升特征表示的能力,帮助网络更好地捕捉特征之间的上下文信息,从而增强模型的感受野和语义理解能力。

3.3.1 SE 注意力机制

SE 模块的设计思想是将全局的空间特征转化为通道特征并学习通道权重,通过压缩(Squeeze)和激励(Excitation)阶段实现。结构如图 5 所示,给定一个特征图 X ,经过 F_{tr} 操作生成特征图 U ,将其进行全局平均池化,生成一个 $1 \times 1 \times C$ 的向量,每个通道表示为式(1):

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j). \quad (1)$$

然后通过两个全连接层 W_1, W_2 对得到的 z 向量进行式(2)所示的处理,得到我们想要的通道权重值 s :

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)). \quad (2)$$

最后将生成的特征向量 $s(1 \times 1 \times C)$ 与特征图 $U(H \times W \times C)$ 对应通道相乘,得到特征图 \tilde{X} ,如式(3)所示:

$$\tilde{X} = F_{scale}(u_c, s_c) = s_c u_c. \quad (3)$$

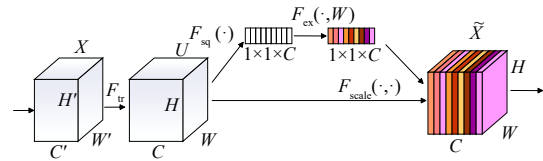


图 5 SE 注意力模块结构图

Fig. 5 Structure diagram of SE attention module

3.3.2 ECA 注意力机制

传统的注意力机制使用全局平均池化或全连接层来计算通道之间的权重,而 ECA 注意力机制通过引入一维卷积操作来实现高效的通道注意力计算。ECA 注意力模块结构如图 6 所示,首先,对特征图进行一维卷积操作,以实现每个通道的特征图进行局部的交互和整合;其次,在卷积操作后,应用一个 sigmoid 函数,将卷积结果映射到 $[0, 1]$ 的范围内,得到每个通道的注意力权重;最后,将注意力权重与输入特征图相乘,得到加权后的特征图。

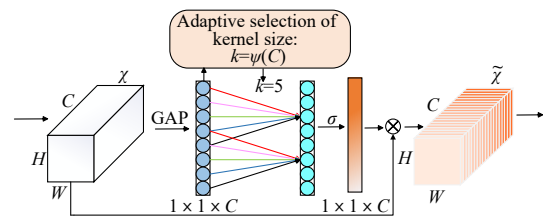


图 6 ECA 注意力模块结构图

Fig. 6 Structure diagram of ECA attention module

3.4 ASPP 的改进

为了捕获更大范围的上下文信息,将原始 ASPP 扩张系数分别为 1、6、12、18 的空洞卷积扩大为 2、12、24、36,扩大卷积核在输入特征图上的采样间隔,从而增加每个位置感受野的范围,提

供更全局的上下文信息。

在提取深层特征时,为了加强对特征信息的充分关注,使模型能够自适应地关注重要特征,抑制不重要特征,并更好地利用特征之间的相关性,在ASPP结构的分支中引入SE注意力机制。每个SEBlock内部都包含了全局池化、全连接层、激活函数等操作,用于学习通道权重。通过通道注意力机制对卷积之后的特征进行加权处理,实现对重要特征的关注并抑制对不重要特征的响应,提高模型的表达能力、泛化能力和性能。

原始ASPP模型中的全局平均池化被用于计算全局特征向量。由于对每个特征通道进行平均操作,将每个通道的特征取平均得到一个标

量,因而会忽略特征图中的位置信息,导致一些细节信息的丢失。为了强调显著特征,保留空间信息,在ASPP的最后一层使用全局最大池化,能够保留特征图中每个通道的最强响应位置,提供一定的空间位置信息。对输入图像进行微小平移或缩放时,输出特征向量保持不变,具备一定的空间不变性。改进的ASPP结构如图7所示。

3.5 优化损失函数

为了更好地衡量模型输出与真实标签之间的差异,本文采用Dice损失函数和Focal损失函数在训练过程中进行联合优化。Dice loss更多地关注难以分类的样本,通过计算预测结果和真实标签的交集与并集之比来衡量相似度,并且对边界细节敏感,提高了分割结果的边界细节准确性。Focal loss通过调整样本的权重,使模型更加关注难以分类的样本,提高对少数类别的学习能力,同时引入了可调参数gamma,用于调节难易样本的关注程度。

3.6 改进的网络结构

改进的DeeplabV3+结构如图8所示,输入图像经过MobileNetV2骨干网络进行提取特征,深层特征从MobileNetV2最后一层获取,并进入改进的ASPP结构,与经过ECA通道注意力特征提取的浅层特征连接,进行卷积、上采样等步骤,输出最终的特征图。

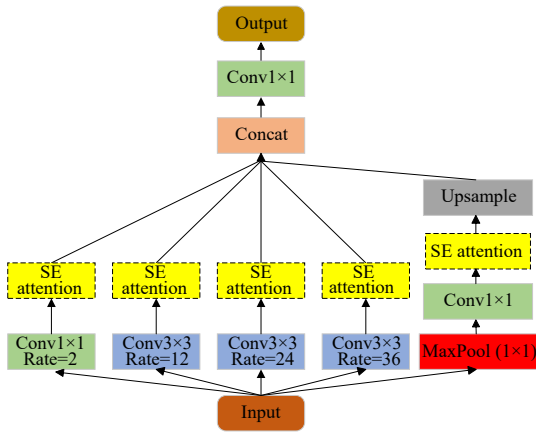


图7 改进的ASPP结构图

Fig. 7 Improved ASPP structure diagram

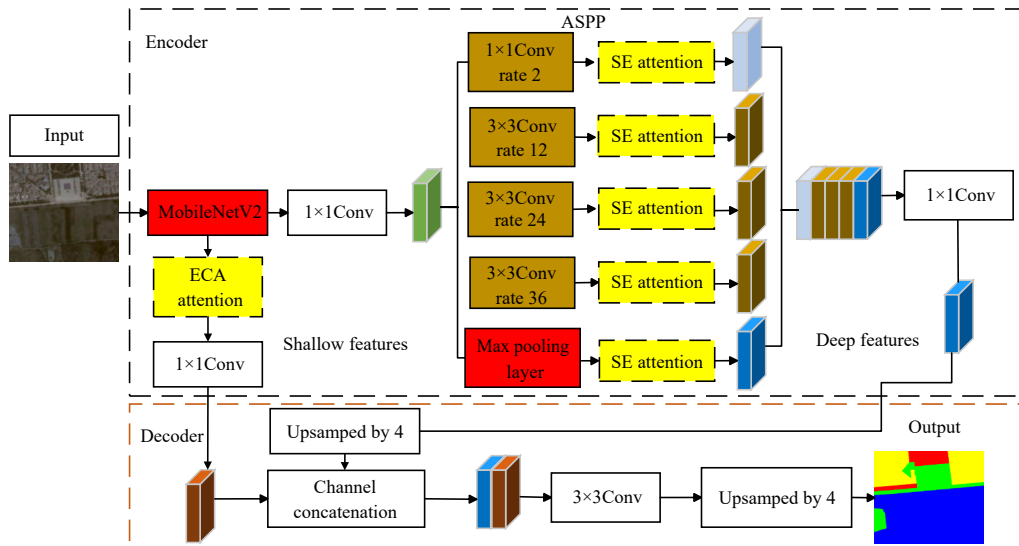


图8 改进的DeepLabV3+模型结构图

Fig. 8 Structure diagram of improved DeepLabV3+ model

4 实验结果与分析

4.1 实验平台及相关参数设置

本文实验平台基于 Windows, CPU 为 Intel (R) Xeon(R) Platinum 8255C, 显卡为 NVIDIA RTX2080, 使用 1.11.0 版本的 Pytorch 框架, Cuda 版本为 11.3。训练参数如表 3 所示。

表 3 训练参数

Tab. 3 Training parameters

Keys	Values
Epoch	110
Batch_size	8
Init_lr	0.007
Min_lr	7e-5
optimizer_type	sgd
momentum	0.9
lr_decay_type	cos

4.2 评价指标

在遥感图像地物分割领域, 预测结果会被分为 4 类: 真正 (True Positive, TP), 表示被正确分类为目标类别的像素数量; 真负 (True Negative, TN), 表示被正确分类为非目标类别的像素数量; 假正 (False Positive, FP), 表示被错误分类为目标类别的像素数量; 假负 (False Negative, FN), 表示被错误分类为非目标类别的像素数量。为了评估改进模型的可行性, 本文采用交并比 (Intersection over Union, IoU)、平均交并比 (Mean Intersection over Union, mIoU)、像素精

度 (Pixel Accuracy, PA)、平均像素精度 (mean Pixel Accuracy, mPA)、精确率 (Precision, P)、准确率 (Accuracy, Acc)、召回率 (Recall, R)、F1 指数、参数量以及训练速率 (每秒处理的迭代次数) 评价指标进行验证。评价指标计算公式如式 (4)~(11) 所示:

$$IoU = \frac{TP}{TP + FN + FP}, \quad (4)$$

$$mIoU = \frac{1}{N} \sum_{i=1}^N IoU, \quad (5)$$

$$PA = \frac{\sum_{i=1}^N n_{ij}}{\sum_{i=1}^N \sum_{j=1}^N n_{ij}}, \quad (6)$$

$$mPA = \frac{1}{N} \sum_{i=1}^N PA, \quad (7)$$

$$P = \frac{TP}{TP + FP}, \quad (8)$$

$$R = \frac{TP}{TP + FN}, \quad (9)$$

$$F1 = 2 \cdot \frac{P \cdot R}{P + R}, \quad (10)$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}, \quad (11)$$

其中: n_{ij} 是混淆矩阵第 i 行第 j 列的值, N 为类别总数。

4.3 实验结果

DeepLabV3+ 网络模型的特征可视化结果如图 9 所示。图像经过网络提取到的深层特征图经过 ASPP 结构的最大池化和 SE 注意力机制处理后, 与经过 ECA 注意力机制处理的浅层特征图融合, 最终输出预测图像的特征图。

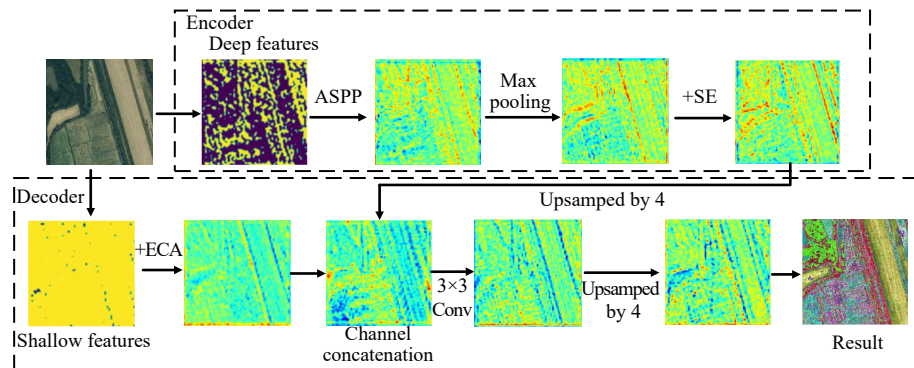


图 9 DeepLabV3+模型的特征图可视化

Fig. 9 Feature map visualization diagram of DeepLabV3+ model

为了直观地展示不同层特征图的变化,从网络中提取了第4层浅层特征图、第5层经过ECA机制处理的特征图、第18层深层特征图、第24层经过最大池化处理的特征图、第28层SE机制处理的特征图、第30层融合特征图以及通过卷积和上采样得到的特征图(经过伪彩色处理)。观察可见,引入最大池化和SE机制后,耕地、林地的

像素点更加突出。在引入ECA的浅层特征图中,林地的像素点更加集中,呈现出更明显的特征。

4.4 消融实验

为验证不同因素对遥感图像地物分割结果的影响,本文设置了两类消融对比实验在两种不同的数据集上进行验证:(1)以 MobileNetV2 为骨干网络,使用 Dice 和 Focal 联合损失函数,在

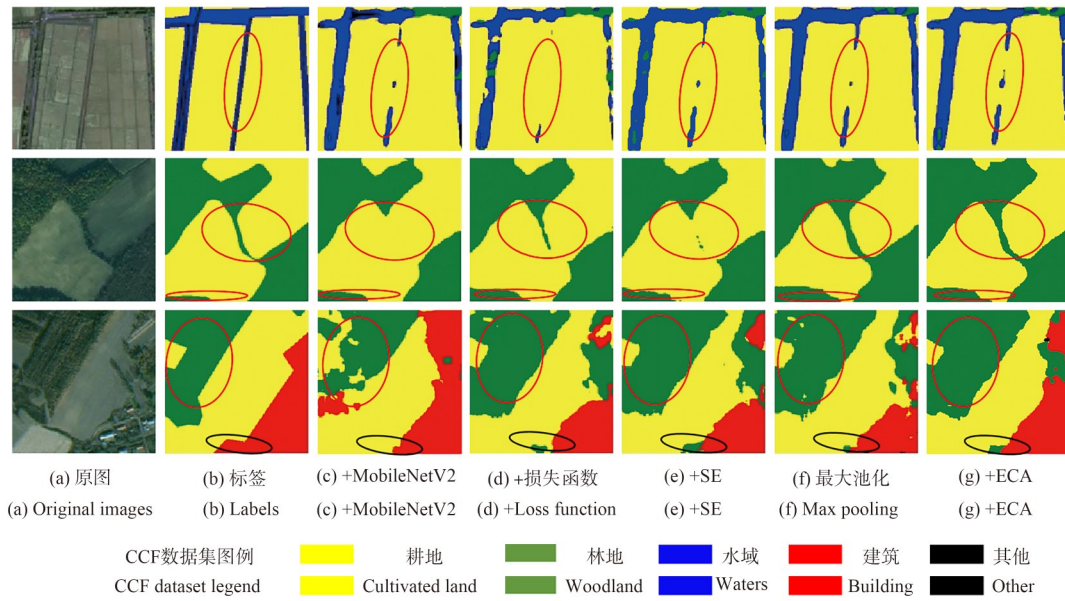


图 10 DeepLabV3+消融实验对比图(CCF数据集图例)

Fig. 10 Comparison chart of DeepLabV3+ ablation experiments (Legend of CCF Dataset)

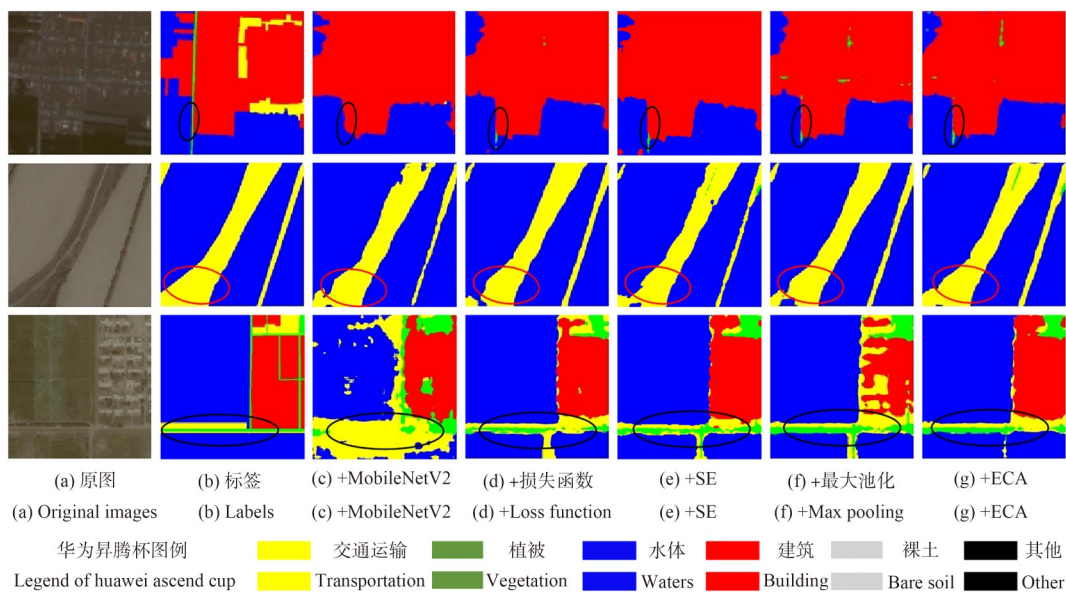


图 11 DeepLabV3+消融实验对比图(华为昇腾杯数据集图例)

Fig. 11 Comparison chart of DeepLabV3+ ablation experiments (Legend of Huawei Ascend Cup Dataset)

ASPP 分支结构中引入 CA、CBAM、ECA、SE 4 种注意力机制,不同模型的参数量对比如表 4 所示,具体评价指标如表 5 所示。(2)沿用原始的 DeepLabV3+的思路,分别对原始 Xception 骨干网络和使用 MobileNetV2 骨干网路和添加 Dice 和 Focal 联合优化损失函数、在 ASPP 结构中引入 SE 注意力机制、引入最大池化层、在浅层特征之后引入 ECA 注意力机制这 5 类改进的方法进行消融实验,经过 110 轮迭代训练后,模型收敛,最终测试图如图 10 和图 11 所示,具体评价指标如表 4、表 5 和表 6 所示。

表 4 不同模型的参数量对比

Tab. 4 Comparison of the number of parameters of different models

方法	参数量/个
原始 Xception	54 709 445
+MobileNetV2	5 814 037
+损失函数	5 874 037
+SE	5 854 997
引入最大池化	5 854 997
+ECA	5 855 000

在 CCF 数据集上进行不同注意力机制的对比实验。通过表 5 可以看出,所有方法相对于基线模型都有所改进,mIoU 和 PA 都有提升。在像素精度上,+CBAMNet 和 +SENet 方法在其他类别和建筑类别相对较高,而 +ECANet 方法在耕地类别相对较高。在交并比方面,+SENet 方法在耕地和水域类别表现较好,而 +ECANet 方法在林地类别表现较好。总体来说,+SENet 模型实现的效果最好,mIoU、PA、F1 指数达到了 73.39%、82.42%、84.21%,相较于原始基准网络分别提高了 0.42%、0.6%、0.34%,每秒处理的迭代次数达到了 27.45 it/s。

由表 4、表 6 和表 7 的实验结果可以看出,本文所改进的方法均能不同程度地提高 DeepLabV3+模型在遥感图像地物分割中的性能。

(1)将原始骨干网络换为 MobileNetV2 明显减少了模型参数量,在运行速度加快的同时各指标也有一定程度的提高。在 CCF 和 华为昇腾杯数据集上,mIoU 分别提高了 2.64% 和 8.91%,准确率分别提高了 1.57% 和 5.47%,F1 指数分别提高了 3.28% 和 7.98%,运行速率分别提高了 12.38 it/s 和 2.68 it/s。

表 5 不同注意力模块的 CCF 数据集实验对比

Tab. 5 Experimental comparison of different attention modules on CCF dataset

方法	IoU/%						PA/%						F1/%	速率/ (it·s ⁻¹)
	其他	耕地	林地	水域	建筑	mIoU	其他	耕地	林地	水域	建筑	mPA		
基线	58.76	83.41	71.39	82.39	68.89	72.97	67.05	95.23	78.91	87.03	80.88	81.82	83.87	26.74
+CANet	87.08	83.43	72.77	83.19	68.31	73.09	66.49	94.50	79.62	90.12	81.43	82.43	83.89	25.88
+CBAMNet	59.78	84.04	71.79	83.11	67.52	73.25	70.42	94.94	79.63	88.09	75.01	81.62	84.05	25.74
+ECANet	59.08	84.18	72.01	82.81	68.25	73.27	68.77	95.31	78.87	88.15	80.35	82.29	84.12	26.45
+SENet	60.57	83.79	71.76	83.34	67.51	73.39	71.74	94.50	80.70	86.91	76.91	82.42	84.21	27.75

表 6 DeeplabV3+模型改进算法的 CCF 数据集消融实验对比

Tab. 6 Comparison of ablation experiments of improvement algorithm for DeeplabV3+ model on CCF dataset

方法	IoU/%						Precision/%						F1/%	速率/ (it·s ⁻¹)
	其他	耕地	林地	水域	建筑	mIoU	其他	耕地	林地	水域	建筑	Acc		
原始 Xception	50.53	82.34	72.79	81.60	63.88	70.23	59.30	91.11	88.53	91.49	71.07	86.36	80.43	13.42
+MobileNetV2	59.00	83.77	71.57	83.25	66.75	72.87	76.37	88.13	88.51	94.50	79.68	87.93	83.71	25.80
+损失函数	58.76	83.41	71.39	82.39	68.89	72.97	82.61	87.04	88.22	93.92	82.29	87.83	83.87	26.74
+SE	60.57	83.79	71.76	83.34	67.51	73.39	79.56	88.08	86.62	95.29	84.89	88.10	84.11	27.75
引入最大池化	60.60	84.15	72.06	82.78	67.45	73.41	82.68	87.98	88.20	93.78	79.11	88.26	84.23	27.78
+ECA	58.62	84.00	71.68	83.43	69.58	73.47	78.37	88.25	87.61	94.02	83.23	88.28	84.29	28.12

表 7 DeeplabV3+模型改进算法的昇腾杯数据集消融实验对比

Tab. 7 Comparison of ablation experiment of improvement algorithm for DeeplabV3+ model on Huawei Ascend Cup dataset

方法	IoU/%							Precision/%							F1/%	速率/ (it·s ⁻¹)
	其他	水体	交通 运输	建筑	植被	裸土	mIoU	其他	水体	交通 运输	建筑	植被	裸土	Acc		
原始Xception	37.70	64.46	77.00	52.27	23.95	34.53	48.32	46.26	80.26	90.00	65.33	33.63	46.18	78.64	63.58	2.99
+MobilNetV2	64.06	66.09	82.51	54.63	32.95	43.16	57.23	85.71	82.55	88.89	71.14	45.46	61.47	84.11	71.56	5.67
+损失函数	66.43	71.73	84.17	62.37	43.59	49.77	63.01	84.17	84.37	91.12	72.64	60.20	63.74	86.48	76.46	8.20
+SE	68.04	71.67	84.13	61.48	41.09	52.68	63.18	86.56	84.68	91.21	70.62	55.28	66.10	84.44	76.62	8.02
引入最大池化	68.71	71.39	83.94	62.64	41.58	50.80	63.28	85.82	84.14	90.77	74.70	57.49	66.71	86.42	76.88	8.07
+ECA	69.29	71.63	84.08	62.40	42.73	50.51	63.43	88.11	84.82	91.10	73.86	55.64	63.33	86.47	77.04	8.05

(2)在替换骨干网络的基础上,结合 Dice-Focal 损失函数,引入最大池化和 SE 机制的 ASPP 模型相对于原始模型表现更好。在 CCF 和华为昇腾杯数据集上 mIoU 分别提高了 3.18% 和 14.96%,准确率分别提高了 1.9% 和 7.78%,F1 指数分别提高了 3.8% 和 13.3%。

(3)在上述方法改进的基础上,在浅层特征之后引入了 ECA 注意力机制,相较于原始 DeepLabV3+ 模型,在 CCF 和华为昇腾杯数据集上 mIoU 分别达到了 73.47% 和 63.43%,提高了 3.24% 和 15.11%;准确率分别达到了 88.28% 和 86.47%,提高了 1.92% 和 7.83%;F1 指数分别达到了 84.29% 和 77.04%,提高了 3.86% 和 13.46%;本文方法相较于原始方法的运行速率在 CCF 和华为升腾杯数据集上分别提高了 14.7 it/s 和 5.06 it/s。

为了直观感受改进方法在数据集上的表现,随机选取了每个数据集上 3 张图片的仿真结

果。从图 10 和图 11 的仿真结果可以看出,引入 SE 机制和最大池化模型的分割呈现一定的误判,在植被和交通运输上分割的范围较大,将不属于其类别的地物也分割到其中。而引入了 ECA 机制后误判范围明显减小。图 10(g) 和图 11(g) 的本文最终方法在各个地物上的分割效果都更接近于原标签图像,对于目标边缘的分割更加清晰。

本文方法在两个数据集上均取得最高的 mIoU,进一步验证了本文方法的鲁棒性。

4.5 主流分割算法实验

为了验证本文算法的有效性,将本文算法与 U-Net、PSPNet、HRNetV2、DL-Unet、原 DeepLabV3+ 以及 MS-DeepLabV3+ 进行了对比。

通过表 8 实验结果可以看出,本文改进的算法在 CCF 数据集上获得了最高性能,在各个地物上都有较好的分割结果,mIoU 达到了 73.47%,准确度达到了 88.08%,F1 指数达到了 84.29%。

表 8 不同模型在 CCF 数据集的实验对比

Tab. 8 Experimental comparison of different model on CCF dataset

方法	骨干 网络	IoU/%							Precision/%					F1/%
		其他	耕地	林地	水域	建筑	mIoU	其他	耕地	林地	水域	建筑	Accuracy	
UNet	VGG16	46.81	80.98	68.79	81.22	61.64	67.89	77.75	83.71	88.79	95.34	78.32	85.73	81.96
PSPNet	ResNet50	45.31	80.42	69.95	78.17	61.41	67.05	75.72	84.79	87.85	88.98	78.14	85.25	80.58
	MobileNetV2	54.80	83.64	72.53	78.92	65.56	71.11	71.35	89.78	89.61	87.18	73.52	87.23	82.67
HRNetV2	—	44.05	78.85	65.65	77.69	61.99	65.65	85.20	82.33	87.29	89.52	72.05	84.16	79.72
DL-Unet	—	55.36	82.22	70.22	81.41	60.67	69.97	75.28	86.46	91.43	92.29	67.67	86.67	82.30
DeepLabV3+	Xception	50.53	82.34	72.79	81.60	63.88	70.23	76.37	88.13	88.51	94.50	79.68	87.93	82.32
MS-DeepLabV3+	ResNet50	56.07	82.82	71.21	82.73	65.06	71.58	74.37	86.88	90.91	93.72	75.84	87.24	83.20
本文方法	MobileNetV2	58.62	84.00	71.68	83.43	69.58	73.47	78.37	88.25	87.61	94.02	83.23	88.08	84.29

5 结 论

本文采用轻量化网络 MobileNetV2 为 DeepLabV3+ 的骨干网络,利用空洞卷积更好地捕获不同尺度的上下文信息。改进的 ASPP 结构进一步增加了感受野,同时融合 SE、ECA 注意力机制增强了重要特征的代表能力。最后采用加权的 Dice-Focal 联合优化损失函数,在 CCF 和 华为昇腾杯竞赛的数据集上进行训练和测试,得到以下结论:

(1) 将原始基准网络中的 Xception 骨干网络替换为轻量化骨干网络 MobileNetV2,可以快速进行特征提取并处理遥感图像大规模数据集。

(2) 在训练和验证的过程中引用加权的 Dice-Focal 联合优化损失函数,提高了训练效率并减少了模型的复杂度。

(3) 对 ASPP 结构进行了增大空洞卷积膨胀率并在 ASPP 结构中使用最大池化,同时引入 SE 注意力机制,可以更有效地提取深层特征信息。

在 CCF 和 华为昇腾杯数据集上的 mIoU 分别提高了 3.18% 和 14.96%, 准确率分别提高了 1.92% 和 7.83%。

(4) 在网络中引入 ECA 注意力机制,可以更全面地利用浅层特征。最终改进的 DeepLabV3+ 模型获得了最佳性能,在 CCF 和 华为昇腾杯数据集上 mIoU 分别达到了 73.47% 和 63.43%, 准确率分别达到了 88.28% 和 86.47%, F1 指数分别达到了 84.29% 和 77.04%。相较于原始 DeepLabV3+ 模型, mIoU 提高了 3.24% 和 15.11%, 准确率提高了 1.47% 和 7.83%, F1 指数提高了 3.86% 和 13.46%。

(5) 本文改进的算法相比于其他模型取得了最高性能。

本文研究方法证明了基于轻量化的 DeepLabV3+ 模型在遥感图像地物分割中有较高的鲁棒性和科学性,但对于小目标分割和相似目标分割还有待改进,后续可加入增强边缘模型 (EEM),改善地物分割的精度和细节保留能力。

参 考 文 献:

- [1] ZHANG Z W, LU W, CAO J S, *et al.* MKANet: An efficient network with sobel boundary loss for land-cover classification of satellite remote sensing imagery [J]. *Remote Sensing*, 2022, 14(18): 4514.
- [2] 程瑞. 基于深度学习的 GF-1 影像下城市公共空间信息提取[D]. 南京:南京邮电大学, 2021.
CHENG R. The extraction of urban public space information from GF-1 image based on deep learning [D]. Nanjing: Nanjing University of Posts and Telecommunications, 2021. (in Chinese)
- [3] 龚轩, 郭中华, 陈旺. 基于 CA-TransUNet 的遥感图像道路分割[J]. *计算机与现代化*, 2023(7): 112-118.
GONG X, GUO Z H, CHEN W. Remote sensing image road segmentation based on CA-TransUNet [J]. *Computer and Modernization*, 2023(7): 112-118.
- [4] 金宇伟. 基于深度学习的高分辨率遥感影像建筑物识别及变化检测[D]. 成都:电子科技大学, 2021.
JIN Y W. Building recognition and change detection from high-resolution remotely sensed imagery using deep learning [D]. Chengdu: University of Electronic Science and Technology of China, 2021. (in Chinese).
- [5] FAN R, BOCUS M J, ZHU Y L, *et al.* Road crack detection using deep convolutional neural network and adaptive thresholding [C]//2019 IEEE Intelligent Vehicles Symposium (IV). Paris: IEEE, 2019: 474-479.
- [6] CUI W H, GUAN Z Q, ZHANG Z Y. An improved region growing algorithm for image segmentation [C]. 2008 International Conference on Computer Science and Software Engineering. Wuhan, China: IEEE, 2008: 93-96.
- [7] 李新娜. 遥感图像林地资源信息提取方法研究[D]. 兰州:兰州交通大学, 2022.
LI X N. Research on extraction method of forest land resources information from remote sensing images [D]. Lanzhou: Lanzhou Jiaotong University, 2022. (in Chinese)
- [8] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Boston: IEEE, 2015: 3431-3440.
- [9] 张丽佳. 采用改进的 DeepLabV3+ 算法的多尺度公路灾害图像分割研究[D]. 银川:宁夏大学, 2022.
ZHANG L J. Multi-scale highway disaster image segmentation using improved DeepLabV3+ algorithm [D]. Yinchuan: Ningxia University, 2022. (in Chinese)

- [10] 张天文. 基于深度学习的 SAR 图像舰船检测及识别技术研究[D]. 成都: 电子科技大学, 2022.
ZHANG T W. Research on deep learning-based SAR ship detection and recognition technology [D]. Chengdu: University of Electronic Science and Technology of China, 2022. (in Chinese)
- [11] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation [C]. 18th International Conference Medical Image Computing and Computer-Assisted Intervention. Munich: Springer, 2015: 234-241.
- [12] HE K M, ZHANG X Y, REN S Q, *et al.* Deep residual learning for image recognition [C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016: 770-778.
- [13] SAIDA T, RASHID M, NEMOTO Y, *et al.* CNN-based segmentation frameworks for structural component and earthquake damage determinations using UAV images [J]. *Earthquake Engineering and Engineering Vibration*, 2023, 22(2): 359-369.
- [14] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(12): 2481-2495.
- [15] HO Y, WOOKEY S. The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling [J]. *IEEE Access*, 2020, 8: 4806-4813.
- [16] LI G Z, JIAN X C, WEN Z C, *et al.* Algorithm of overfitting avoidance in CNN based on maximum pooled and weight decay [J]. *Applied Mathematics and Nonlinear Sciences*, 2022, 7(2): 965-974.
- [17] ZHAO H S, SHI J P, QI X J, *et al.* Pyramid scene parsing network [C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017: 6230-6239.
- [18] HE K M, ZHANG X Y, REN S Q, *et al.* Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1904-1916.
- [19] CHEN L C, ZHU Y, PAPANDREOU G, *et al.* Encoder-decoder with atrous separable convolution for semantic image segmentation [C]// *Proceedings of the 15th European Conference on Computer Vision (ECCV)*. Munich: Springer, 2018: 833-851.
- [20] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions [C]//*Proceedings of the 4th International Conference on Learning Representations*. San Juan: ICLR, 2016: 823-834.
- [21] 石甜甜, 郭中华, 闫翔, 等. 基于多尺度融合注意力改进 UNet 的遥感图像水体分割[J]. 液晶与显示, 2023, 38(3): 397-408.
SHI T T, GUO Z H, YAN X, *et al.* Water body segmentation in remote sensing images based on multi-scale fusion attention module improved UNet [J]. *Chinese Journal of Liquid Crystals and Displays*, 2023, 38(3): 397-408. (in Chinese)
- [22] 沈言善, 王阿川. 基于深度学习的遥感图像地物分割方法[J]. 液晶与显示, 2021, 36(5): 733-740.
SHEN Y S, WANG A C. Remote sensing image feature segmentation method based on deep learning [J]. *Chinese Journal of Liquid Crystals and Displays*, 2021, 36(5): 733-740. (in Chinese)
- [23] 周羿, 刘德儿. 融合注意力机制及 DenseASPP 改进的 DeeplabV3+ 遥感图像分割方法[J]. 遥感信息, 2023, 38(3): 85-92.
ZHOU Y, LIU D R. A semantic segmentation method for remote sensing image based on fusion attention mechanism and DenseASPP improved DeeplabV3+ [J]. *Remote Sensing Information*, 2023, 38(3): 85-92. (in Chinese)
- [24] SHAHEED K, MAO A H, QURESHI I, *et al.* DS-CNN: A pre-trained Xception model based on depth-wise separable convolutional neural network for finger vein recognition [J]. *Expert Systems with Applications*, 2022, 191: 116288.
- [25] SZEGEDY C, LIU W, JIA Y Q, *et al.* Going deeper with convolutions [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Boston: IEEE, 2015: 1-9.
- [26] HOWARD A G, ZHU M L, CHEN B, *et al.* MobileNets: Efficient convolutional neural networks for mobile vision applications [J/OL]. *arXiv*, 2017: 1704. 04861.

- [27] KHAN M A, PARK H, CHAE J. A lightweight convolutional neural network (CNN) architecture for traffic sign recognition in urban road networks [J]. *Electronics*, 2023, 12(8): 1802.

作者简介:



马 静,女,硕士研究生,2021年于宁夏大学获得学士学位,主要从事遥感图像处理和深度学习算法方面的研究。E-mail:2280355382@qq.com



郭中华,男,博士,教授,2010年于西北工业大学获得博士学位,主要从事遥感图像处理、高光谱图像技术、机器视觉等方面的研究。E-mail:guozhh@nxu.edu.cn