

基于 DeeplabV3+ 网络的高分遥感影像分类

黄冬青^{1,2,3}, 徐伟铭^{1,2,3*}, 许文迪^{1,2,3}, 何小英^{1,2,3}, 潘凯祥^{1,2,3}

¹福州大学数字中国研究院 (福建), 福建 福州 350108;

²福州大学空间数据挖掘与信息共享教育部重点实验室, 福建 福州 350002;

³福州大学地理空间信息技术国家地方联合工程研究中心, 福建 福州 350002

摘要 针对卷积神经网络在遥感影像分类时遇到的模型参数量过大和分类精度低等问题, 在 DeeplabV3+ 网络的基础上, 将编码器中的深层特征提取器替换为轻量化网络 MobilenetV2 和 Xception_65, 将解码器结构改为逐层特征融合实现解码区上采样的细化, 引入通道注意力模块加强编解码器之间的信息关联, 引入多尺度监督实现感受野自适应。构建 4 种具有不同编解码结构的网络, 在 CCF 数据集上对网络进行验证测试。实验结果表明, 编码器采用 Xception_65, 解码器同时引入逐层连接、通道注意力模块和多尺度监督的 MS-XDeeplabV3+ 网络在减少模型参数量、加快模型训练速度的同时能更细化地物的边缘信息, 提高对道路、水体等线状地物和草地的分类精度, 像素总体精度和 Kappa 系数分别达 0.9122 和 0.8646, 在遥感影像分类中效果最佳。

关键词 遥感影像分类; 卷积神经网络; 编解码结构; 逐层特征融合; 通道注意力模块; 多尺度监督

中图分类号 TP751

文献标志码 A

DOI: 10.3788/LOP222553

High-Resolution Remote Sensing Image Classification Based on DeeplabV3+ Network

Huang Dongqing^{1,2,3}, Xu Weiming^{1,2,3*}, Xu Wendi^{1,2,3}, He Xiaoying^{1,2,3}, Pan Kaixiang^{1,2,3}

¹The Academy of Digital China, Fuzhou University, Fuzhou 350108, Fujian, China;

²Key Laboratory of Spatial Data Mining & Information Sharing, Ministry of Education, Fuzhou University, Fuzhou 350002, Fujian, China;

³National Engineering Research Centre of Geospatial Information Technology, Fuzhou University, Fuzhou 350002, Fujian, China

Abstract This paper addresses the challenges of high model complexity and low classification accuracy in remote sensing image classification using convolutional neural networks. To overcome these challenges, a modified DeeplabV3+ network is proposed, which replaces the deep feature extractor in the encoder with lightweight networks MobilenetV2 and Xception_65. The decoder structure is also modified to feature fusion layer by layer in order to refine the up-sampling process in the decoding region. In addition, a channel attention module is introduced to strengthen the information association between codecs, and multiscale supervision is used to adapt the receptive field. Four networks with different encoding and decoding structures are constructed and verified on the CCF dataset. The experimental results show that the MS-XDeeplabV3+ network, which uses Xception_65 in the encoder and layer by layer connection, channel attention module, and multiscale supervision in the decoder, has reduced number of model parameters, faster training speed, refined edge information for ground objects, and improved classification accuracy for grassland and linear ground objects such as roads and water bodies. The overall pixel accuracy and Kappa coefficient of the MS-XDeeplabV3+ network reach 0.9122 and 0.8646, respectively, which show the best performance among all networks in remote sensing image classification.

Key words remote sensing image classification; convolutional neural network; encode and decode structure; layer by layer feature fusion; channel attention module; multiscale supervision

收稿日期: 2022-09-15; 修回日期: 2022-10-15; 录用日期: 2022-11-24; 网络首发日期: 2023-01-04

通信作者: *xwming2@126.com

1 引言

遥感影像分类^[1]指根据地物的纹理、光谱等特点,对不同的目标地物进行识别,从而对影像中每个像素赋予地物类别的过程。对遥感影像进行分类,将其转换为可被计算机理解和处理的地物分类信息,为城市规划、环境监测和土地覆盖变化检测等高层应用提供支撑。

随着计算机硬件和深度学习的快速发展,卷积神经网络(CNN)由于强大的特征表达能力,被广泛应用到遥感影像分类领域。张乾坤等^[2]在 CNN 中加入地块数据,完成对地块的分类,实验结果表明所设计的模型的分精度明显优于传统机器算法。但 CNN 分类对象为图像块,当存在混合图像块时,会将其分类为唯一类,影响分类精度与效果。Shelhamer 等^[3]提出了全卷积神经网络(FCN),该网络去除 CNN 的全连接层并添加上采样还原图像的尺寸,实现像素级分类。谢锦莹等^[4]将面向对象的多尺度分割方法与 FCN 相结合,有效消除了 FCN 分类结果存在的椒盐现象,提高了对滨海湿地植被的分类精度。但由于 FCN 存在下采样和池化,特征图中丢失了很多细节信息,对地物细节分割效果欠佳。在 FCN 的基础上,Ronneberger 等^[5]提出了基于编解码结构的 U-Net,编码区对图像进行深度特征提取,解码区对不同尺度的特征图进行融合。李传林等^[6]提出了一种改进型 U-Net 网络,通过引入综合注意力机制和残差模块,该网络增强对建筑物的特征学习,提高卷积结构的鲁棒性。但 U-Net 结构过于单一,简单的编解码结构导致在网络训练过程中收敛速度较慢。Badrinarayanan 等^[7]提出了 SegNet 网络,在解码过程中利用编码器池化层的下标索引恢复图像分辨率,加速了网络的推理。张哲晗等^[8]通过减少 SegNet 网络的池化层和增加卷积层,拓宽网络深度,丰富图像特征,网络对地物分类的平均精度达到

86%。但 SegNet 上采样仅对索引位置的特征进行赋值,其余位置的信息则发生丢失。谷歌公司提出了 Deeplab 系列模型^[9-12],其中 DeeplabV3+ 模型是遥感影像分类常用模型。DeeplabV3+ 模型在空洞空间金字塔池化(ASPP)和空洞卷积的基础上构建了编解码结构,对浅层特征与深层特征进行融合,提高了对地物的分割精度。许泽宇等^[13]提出了 E-Deeplab 网络模型,通过增加上采样层,缩小上采样倍数,提高编解码器之间的紧密性,使用自适应权重损失函数,自动调节地物损失权重,模型精度明显优于其他相关模型。但 DeeplabV3+ 网络显存开销大且计算速度慢,上采样过程中对特征图的多尺度连接不够充分,对细小地物分割仍不够精细。

针对基于卷积神经网络的高分遥感影像分类模型存在的参数量大、训练时间长和对细小地物分割不精细等问题,本文基于 DeeplabV3+ 网络,构建 4 种具有不同编解码结构的网络。为减少模型参数量和训练时间,在 DeeplabV3+ 网络的编码区引入 MobilenetV2 和 Xception_65 两种轻量级网络,构建 MDeeplabV3+ 和 XDeeplabV3+ 网络。在此基础上,为提高模型的分精度,在解码区引入逐层特征融合、通道注意力模块和多尺度监督,构建 MS-MDeeplabV3+ 和 MS-XDeeplabV3+ 网络,最后在 CCF 数据集上对网络进行验证测试。

2 网络结构设计

为能减少基于深度学习的高分遥感影像分类模型的参数量的同时提高分类精度,基于 DeeplabV3+ 网络,引入轻量化网络、逐层特征融合、通道注意力机制和多尺度监督等模块来构建 MDeeplabV3+、XDeeplabV3+、MS-MDeeplabV3+ 和 MS-XDeeplabV3+ 4 种具有不同编解码结构的模型,模型结构特点如表 1 所示。

表 1 4 种网络结构的异同点

Table 1 Similarities and differences between the four network structures

Model structure	Encoder		Decoder			
	MobilenetV2	Xception_65	Skip a layer fusion	Layer-by-layer fusion	Channel module	Multi-scale supervision
MDeeplabV3+	✓		✓			
XDeeplabV3+		✓	✓			
MS-MDeeplabV3+	✓			✓	✓	✓
MS-XDeeplabV3+		✓		✓	✓	✓

2.1 MDeeplabV3+

DeeplabV3+ 网络是编解码结构,可分为编码区和解码区,如图 1 所示。编码区的深度卷积神经网络(DCNN)对原始图像进行特征提取,ASPP 对 DCNN 提取的特征图进行多尺度的空洞卷积采样,之后将多尺度的高级语义特征图组合后进行 1×1 卷积,得到深

度特征图。对于解码区,从 DCNN 中提取一张浅层特征图,对其与 4 倍上采样之后的深度特征图进行融合,对深度特征图丢失的位置信息进行补充,最后进行 4 倍上采样得到语义分割结果图。

当前地物分类大都使用卷积神经网络进行特征提取,如 AlexNet、VGGNet 和 ResNet 等优秀的基础网

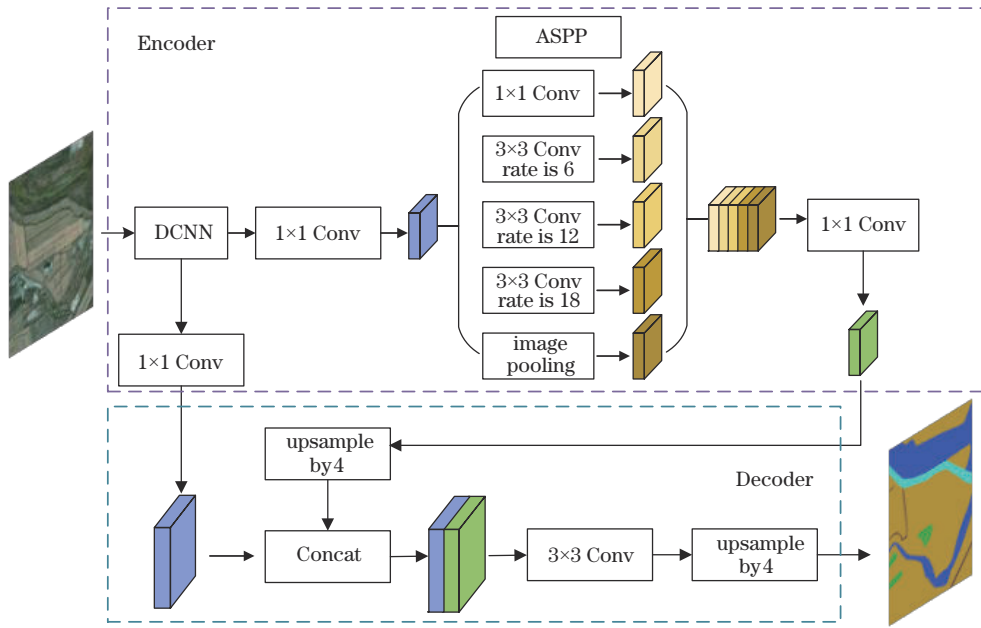


图 1 DeeplabV3+网络结构

Fig. 1 Network architecture of DeeplabV3+

络,但这些网络计算量大,需要较大的存储空间和算力资源。为能在保持模型精度的同时减小模型参数数量和复杂度,在 DCNN 中采用 MobilenetV2 网络,构建 MDeeplabV3+。Mobilenet^[14]是为了终端设备计算机视觉应用设计的,特点是参数量可配置、耗存小、运行高效。MobilenetV2 主要引入了倒残差结构和深度可分离卷积,网络结构如图 2 所示,详细配置如表 2 所示,其中 Sep Conv2d 表示深度可分离卷积, t 表示升维倍数, c 表示输出通道数, n 表示倒残差块数量, s 表示步

表 2 MobilenetV2 的详细配置

Table 2 Detailed configuration of the MobilenetV2

Input size	Operation	t	c	n	s
$256^2 \times 3$	Conv2d		32	1	2
$128^2 \times 32$	Bottleneck	1	16	1	1
$128^2 \times 16$	Bottleneck	6	24	2	2
$64^2 \times 24$	Bottleneck	6	32	3	2
$32^2 \times 32$	Bottleneck	6	64	4	2
$16^2 \times 64$	Bottleneck	6	96	3	1
$16^2 \times 96$	Bottleneck	6	160	3	1
$16^2 \times 160$	Bottleneck	6	320	1	1

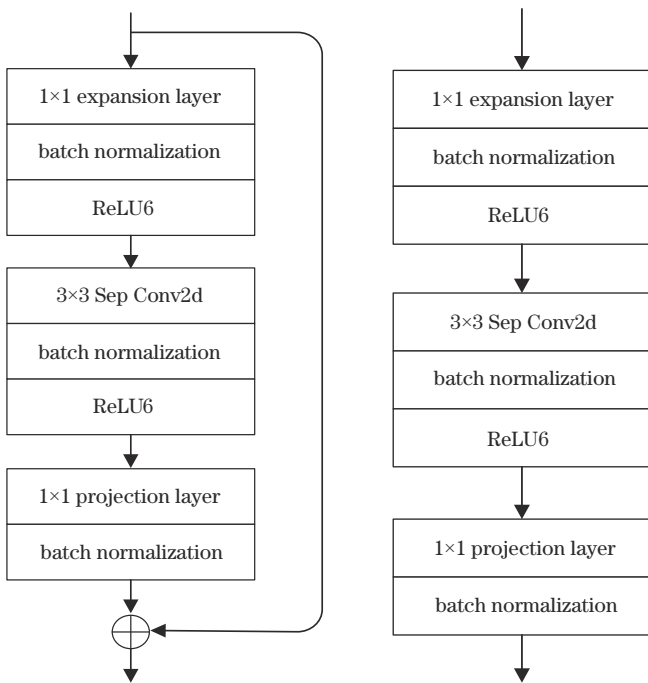


图 2 倒残差结构

Fig. 2 Bottleneck residual block

长。倒残差结构是对输入特征进行先升维后降维,从而提取更多特征,并使用 ReLU6 函数代替 ReLU 函数的结构。深度可分离卷积是将一个卷积层分为处理空间相关的逐通道卷积和处理通道相关的逐点卷积的过程。假设输入特征图大小为 $C_1 \times H_1 \times W_1$, 输出特征图大小为 $C_2 \times H_2 \times W_2$, 其中 $C_2 > C_1$, 卷积核大小为 $e \times e$ 。对于普通的一个卷积层,参数量总共有 $(e \times e \times C_1 + 1) \times C_2$ 。深度可分离卷积则分为两步,先逐通道卷积后逐点卷积。逐通道卷积不改变通道数 C , 只改变特征图大小 H 和 W , 卷积核大小为 $e \times e$, 输出特征图大小为 $C_1 \times H_2 \times W_2$, 参数量共有 $(e \times e + 1) \times C_1$ 。逐点卷积不改变特征图大小 H 和 W , 改变通道数 C , 在通道方向进行 1×1 卷积, 输出特征图大小为 $C_2 \times H_2 \times W_2$, 参数量为 $(1 \times 1 \times C_1 + 1) \times C_2$, 与普通卷积相比,参数大量减少,加快了模型训练速度。

2.2 XDeeplabV3+

遥感影像中地物类型与空间位置关系密切,且地物的纹理特性比较复杂,所以需要有一个具有较强的特

征信息提取能力的网络。MobilenetV2 网络层数较少且结构简单,提取的特征有限,因此为拓宽网络深度,将 Xception_65 网络作为特征提取器,构建 XDeeplabV3+ 网络。Xception_65^[15] 是谷歌公司提出的一种用于提取图像深层特征的轻量化网络,将普通卷积替换为深度

可分离卷积,并引入残差块。残差块在扩展网络深度的同时,有效地解决了网络过深造成的梯度消失和爆炸等问题。深度可分离卷积用于网络压缩和提速,减少网络参数量,网络结构如图 3 所示,其中 Sep Conv 表示深度可分离卷积。

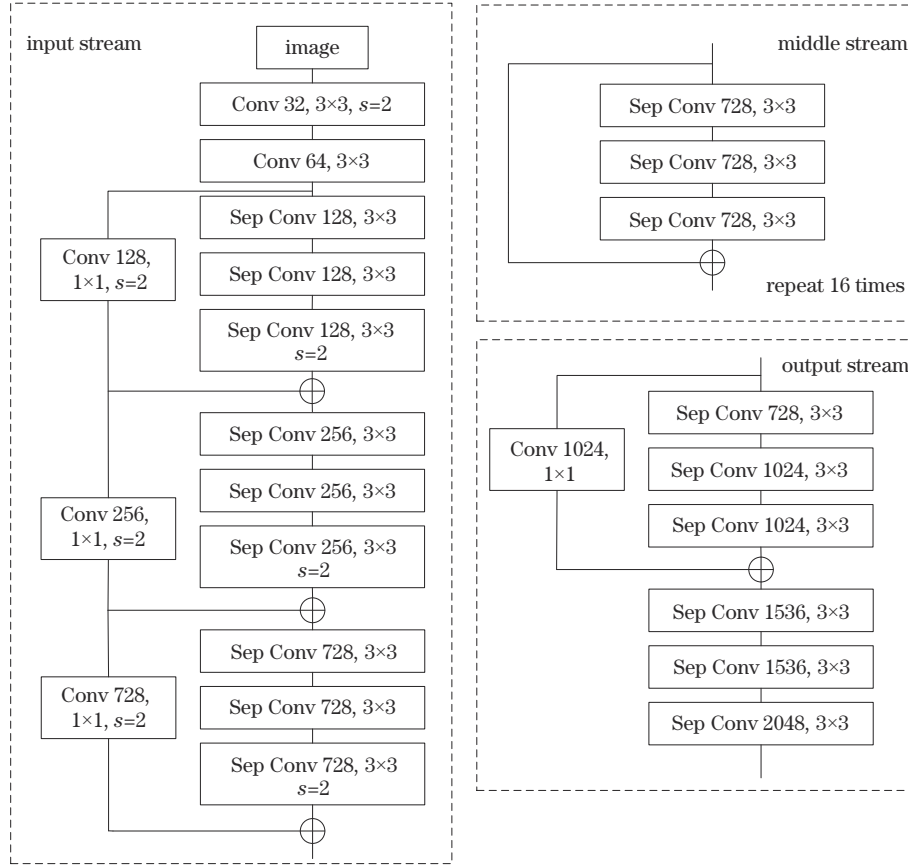


图 3 Xception_65 网络结构
Fig. 3 Network architecture of Xception_65

2.3 MS-MDeeplabV3+与 MS-XDeeplabV3+

针对模型在处理遥感影像分类时遇到的边缘分类不准确、细小分类精度低等问题,在 MDeeplabV3+ 和 XDeeplabV3+ 的基础上将解码区设计为逐层特征融合,引入通道注意力机制和多尺度监督等模块,构建了 MS-MDeeplabV3+ 和 MS-XDeeplabV3+ 网络,模型结构如图 4 所示。

1) 逐层特征融合

MDeeplabV3+ 和 XDeeplabV3+ 网络在解码部分仅利用 2 个 4 倍上采样和 1 个特征融合层实现解码操作,对特征图的多尺度连接不够充分,提取的特征包含较少的空间位置信息,导致地物边界不准确。因此,为了能在不增加计算量的前提下,提高模型的分类精度,综合考虑了解码区的功能要求,引入 U-Net 网络的逐层融合方法,网络结构如图 4 所示。输入图像经过特征提取网络和 ASPP 模块后得到的深度特征图在解码区,经过 4 个 2 倍上采样和 3 个特征融合模块,实现更细化的上采样操作,提高分类精度。

2) 通道注意力机制

通过逐层特征融合,解码区上采样过程中融入了更多具有位置信息的浅层特征,造成网络信息冗余,不能突出重要特征信息。因此,利用通道注意力(DCA)模型来抑制其他特征,提取关键特征,从而实现对目标区域的自动选择和权重分配。DCA 模块的结构如图 5 所示。对输入的特征图分别使用全局最大池化和全局平均池化进行压缩,然后采用大小为 5 的 1D 卷积获取局部跨通道交互信息,将交互信息相加后,再通过激活函数 Sigmoid 得到各通道权重,最后对各通道权重与输入特征图进行卷积运算。传统通道注意力模块通常采用 2 个全连接层捕捉非线性的跨通道交互信息,但降维计算量较大,且不利于捕获各通道之间的依赖关系。DCA 模块采用 1D 卷积,提出了一种不降维的局部跨特征通道交互策略,有效避免了降维对通道注意力学习效果的影响。

DCA 关注的重点是“哪个通道”,在通道维度对输入的特征图进行权重分配,提高模型判别和分析重要

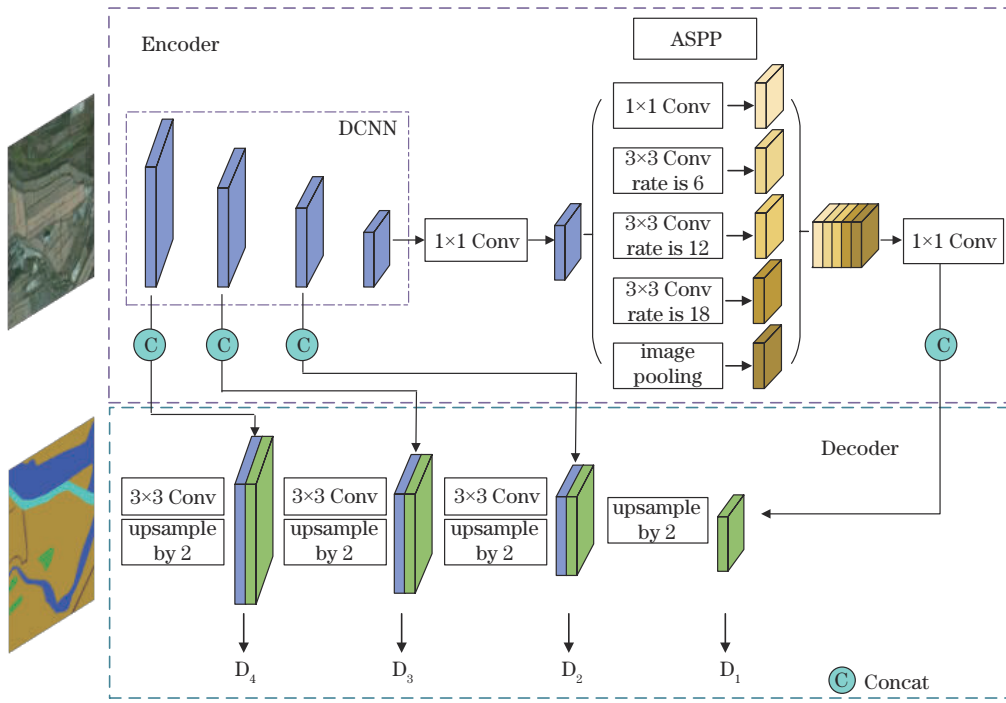


图 4 MS-XDeeplabV3+ 和 MS-MDeeplabV3+ 的网络结构

Fig. 4 Network architecture of MS-XDeeplabV3+ and MS-MDeeplabV3+

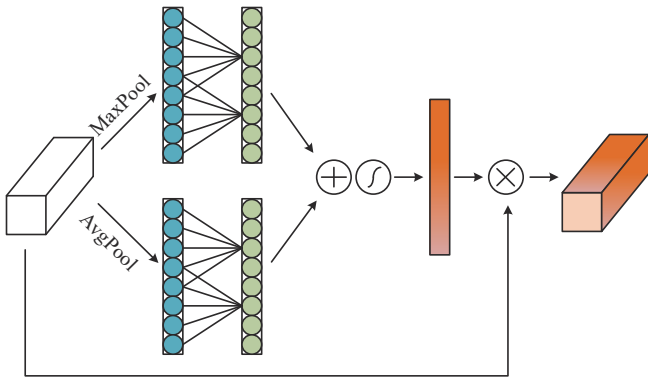


图 5 DCA 结构示意图

Fig. 5 Schematic of DCA structure

特征的能力,以极轻量方式从异构特征图中学习有用的特征信息,有效解决了跨通道维度的异构问题,提高了区别不同地物的能力。

3) 多尺度监督

对于监督学习,常用的方法是计算最后阶段的特征图与真实地表的损失。但是遥感影像上不同类别地物的大小差异较大,网络需要在不同尺度下对地物目标的特征进行有效提取。为提取不同尺度下的特征信息,引入多尺度监督(multiscale supervision)^[16-17],网络结构如图 4 所示,在网络的解码过程中,对 4 个不同尺寸的特征图进行侧输出,获取不同感受野下的预测图,并进行相应的损失计算。损失函数采用交叉熵损失函数和 Dice 损失函数。交叉熵损失函数是使用最广泛的损失函数,用于计算预测图和真实地表在像素级上的差异,Dice 损失函数用于计算预测图和真实地表在

整体上的差异。交叉熵损失函数的公式为

$$l_{CE} = -\frac{1}{N} \sum_{n=1}^N y_i \log p_i, \quad (1)$$

式中: N 为输入图像的像素总数; p_i 表示样本属于第 i 类的概率; y_i 表示当样本属于第 i 类别时, $y_i = 1$,否则 $y_i = 0$ 。

Dice 损失函数可以用混淆矩阵计算,公式为

$$l_{Dice} = 1 - \frac{2N_{TP}}{N_{FP} + 2 \times N_{TP} + N_{FN}}, \quad (2)$$

式中:TP 表示分类器正确分类的正样本像素个数;FP 表示分类器错分为正样本的像素个数;FN 表示分类器错分为负样本的像素个数。

总损失的计算公式为

$$l = l_{CE} + l_{Dice}. \quad (3)$$

网络通过 4 个不同分辨率的侧输出进行多尺度监督。随着分辨率的增加,特征图细节更加丰富,需要对 4 个损失值进行加权求和,因此最终的损失函数定义为

$$L = \sum_{i=1}^4 w_i l_i, \quad (4)$$

式中: l_i 和 w_i 是第 i' 个阶段的整体损失和权重输出。

3 分析与讨论

3.1 实验数据

实验数据来源于 2020 年“中国计算机学会”(CCF)卫星影像的 AI 分类与识别大赛。该数据集包含 140000 张尺寸为 256×256 的影像,并将地物分为

建筑、耕地、水体、道路、林地、草地和其他 7 类,各类地物数量分布极不均衡,尤其是建筑、道路和草地 3 类,具体情况如表 3 所示,数据集部分样本如图 6 所示。若采用地物类别数量不均衡的数据集进行训练,会导致对不同类别的特征学习能力有较大差异,很难得到

较好的分类模型。因此,本文对数据进行筛选,再根据遥感影像的旋转不变性,使用图像翻转、水平镜像和垂直镜像等方式进行数据增强,平衡各类别地物的分布,最后,将训练集、验证集和测试集按照 7:2:1 的比例进行划分。

表 3 CCF 数据集各类别占比
Table 3 Proportion of each category in the CCF dataset

Parameter	Building	Arable	Forest	Water	Road	Grass	Other
Label	0	1	2	3	4	5	6
Percentage /%	2.79	50.87	17.87	17.74	0.35	1.96	7.38

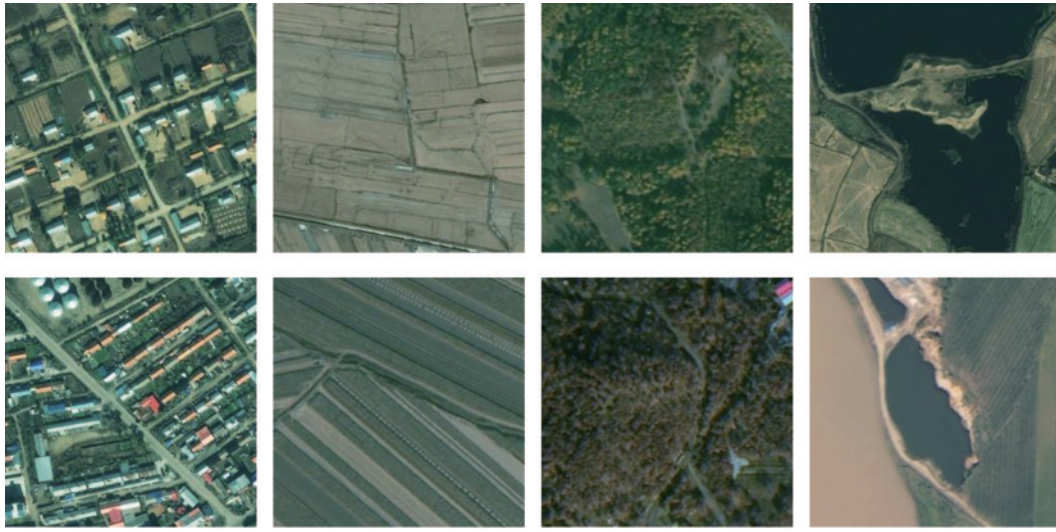


图 6 CCF 数据集部分遥感地物图像
Fig. 6 Partial samples of CCF dataset

3.2 实验环境与训练策略

实验环境为 Windows10 操作系统,采用 Python 编程语言,DeeplabV3+模型是基于 PyTorch 深度学习框架实现的,GPU 为 NVIDIA GeForce RTX 3070,内存为 8 GB。

在训练模型阶段,4 种网络模型使用相同的超参数进行训练。采用自适应学习率策略,初始学习率设置为 1×10^{-4} ,若验证集精度在连续的 10 个 epoch 不上升时,学习率乘上系数 0.3,批次大小设置为 4,采用交叉熵和 Dice 作为损失函数,采用 Adam 作为优化器,训练轮数设置为 60,保证充分拟合。

3.3 评价指标

选取总体精度(OA)、交并比(IoU)和 Kappa 系数作为精度评价指标。总体精度是正确分类的像素个数与总像素个数的比值,用于评价模型的总体分类精度,对不平衡数据中数量少的数据不敏感。交并比是各类别预测集合中的正确分类集合和真实集合的交集与并集的比值,用于评价各类别的分类精度。Kappa 系数用于一致性检验,也可以用于衡量分类精度。OA、IoU 和 Kappa 系数的计算公式分别为

$$P_{OA} = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{TN} + N_{FP} + N_{FN}}, \quad (5)$$

$$R_{IoU} = \frac{N_{TP}}{N_{TP} + N_{FP} + N_{FN}}, \quad (6)$$

$$C_{Kappa} = \frac{P_o - P_e}{1 - P_e}, \quad (7)$$

式中:TN 表示分类器正确分类的负样本像素个数; P_o 表示分类器正确分类的像素个数之和除以总像素个数的结果; P_e 表示各类别预测像素个数与真实像素个数乘积的和除以总像素个数的平方的结果。

3.4 结果分析

3.4.1 4 种模型对比分析

为对比 4 种具有不同编解码结构的网络对分类精度的影响,在 CCF 数据集中选取 6 组影像,对比分析模型得到的遥感影像分类结果,如图 7 所示。结果表明:MDeeplabV3+难以准确分割草地和细小地物,尤其是线状水体和道路;XDeeplabV3+在细小地物识别分类方面有所改善,但仍会出现不连续、错分等问题;在 MDeeplabV3+的基础上,通过改造解码器,引入通道注意力模块和多尺度监督,MS-MDeeplabV3+网络

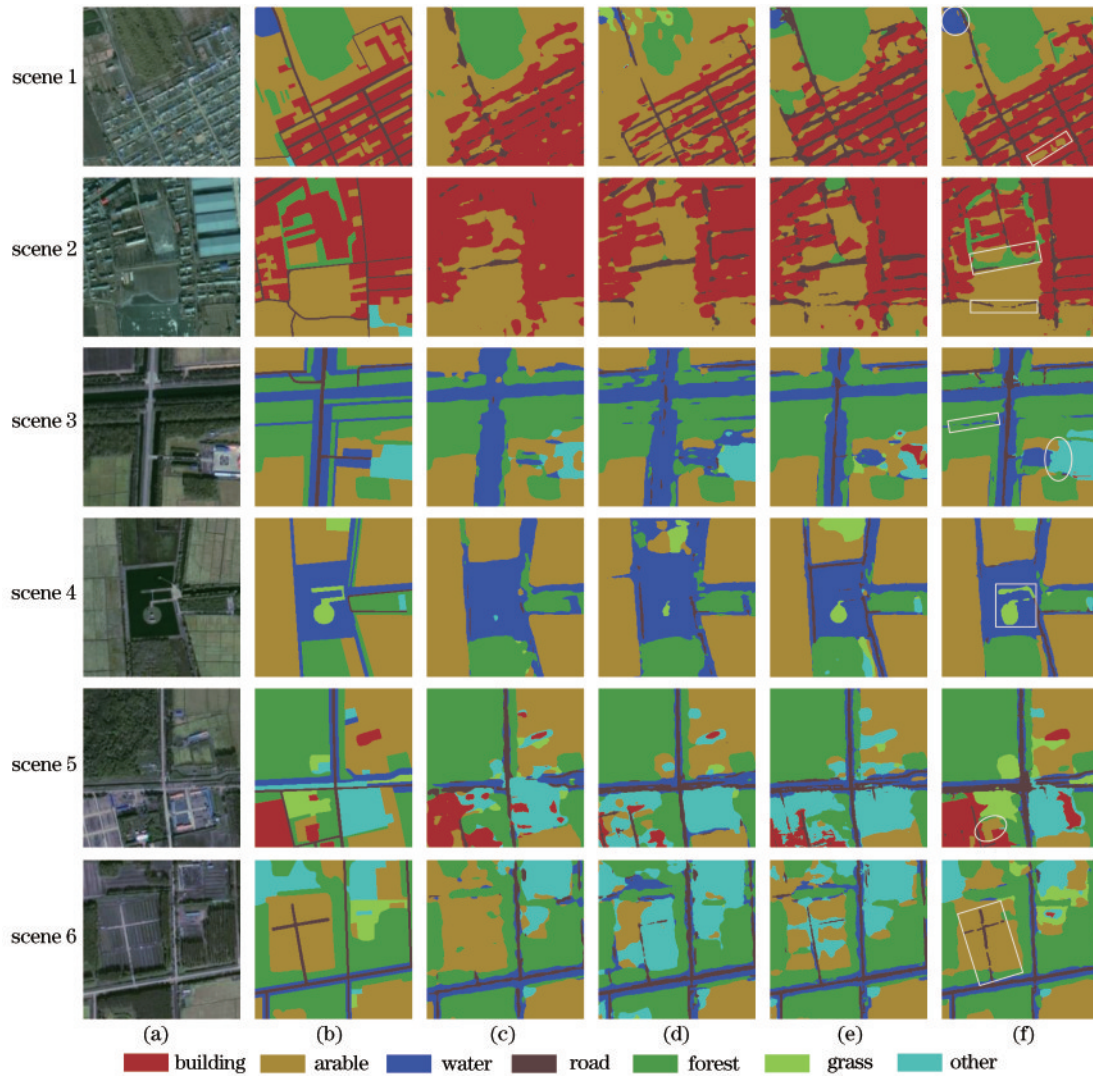


图 7 4 种模型的分类结果对比。(a)原始图像;(b)标签;(c) MDeeplabV3+;(d) XDeeplabV3+;(e) MS-MDeeplabV3+;(f) MS-XDeeplabV3+

Fig. 7 Comparison of the classification results of the four models. (a) Original image; (b) label; (c) MDeeplabV3+; (d) XDeeplabV3+; (e) MS-MDeeplabV3+; (f) MS-XDeeplabV3+

对草地的分类识别精度有所提高,但边缘分割不准确,对连续性线状地物如道路、水体等的分类精度有所提高,但仍会出现错漏和不连续等问题;融合了 XDeeplabV3+ 和 MS-MDeeplabV3+ 的优势构建的 MS-XDeeplabV3+ 网络明显提高了地物边缘分割精度,有效提高了整体和局部分类精度。在对模型的编码器特征提取部分的改进过程中, MDeeplabV3+ 的网络性能明显低于 XDeeplabV3+, XDeeplabV3+ 关

注到更细节的特征信息;在解码器部分,通道注意力模块突出了重要信息,多尺度监督对解码器中不同尺度特征图计算损失,提高了对线状地物的分类精度。

用 IoU、OA 和 Kappa 系数指标统计了 4 种模型对应分类结果的精度,如表 4 所示。MDeeplabV3+ 通过引入轻量化网络 MobilenetV2,网络训练速度加快且参数量极少,但对道路和草地的分类精度很低,导致模型总体精度不高;在 MDeeplabV3+ 的基础上,将

表 4 各模型分类结果的定量评价

Table 4 Quantitative evaluation of the classification results of each model

Method	IoU							mIoU	OA	Kappa
	Building	Arable	Forest	Water	Road	Grass	Other			
MDeeplabV3+	0.6882	0.8047	0.7828	0.8249	0.2015	0.2286	0.6233	0.5934	0.8027	0.7735
XDeeplabV3+	0.7144	0.8240	0.8267	0.8706	0.2554	0.2034	0.6480	0.6204	0.8348	0.7992
MS-MDeeplabV3+	0.7217	0.8234	0.8186	0.8625	0.3431	0.2298	0.6442	0.6348	0.8502	0.8297
MS-XDeeplabV3+	0.7650	0.8836	0.8437	0.9168	0.4774	0.3309	0.6615	0.6970	0.9122	0.8646

DCNN 中的 MobilenetV2 换成 Xception_65, 结果表明, XDeeplabV3+ 网络的 Kappa 系数和 OA 达 0.7992 和 0.8348, 分别比 MDeeplabV3+ 高出 0.0257 和 0.0321。在各类别方面, XDeeplabV3+ 对道路的 IoU 与 MDeeplabV3+ 相比提高最多, 提高 0.0539, 对草地的 IoU 下降 0.0252, 表明 XDeeplabV3+ 对草地的提取不如 MDeeplabV3+。

在 MDeeplabV3+ 基础上, 引入逐层特征融合、通道注意力机制和多尺度监督构建的 MS-MDeeplabV3+ 模型的 OA 和 Kappa 系数达 0.8502 和 0.8297, 相比 MDeeplabV3+, 分别高出 0.0475 和 0.0562。就各类别而言, 相比 MDeeplabV3+, MS-MDeeplabV3+ 精度提高程度最小的是草地, 最大的是道路, 草地的 IoU 提高 0.0012, 道路提高 0.1416, 说明对 MDeeplabV3+ 模型的解码区进行改进后对分类精度的提高优于对编码区的改进。

在 XDeeplabV3+ 的基础上构建 MS-XDeeplabV3+, 由表 4 可知, MS-XDeeplabV3+ 的 Kappa 系数和 OA 是 0.8646 和 0.9122, 比 MDeeplabV3+ 网络高出 0.1095 和 0.0911, 比 MS-MDeeplabV3+ 网络高出 0.0620 和 0.0349, 比 XDeeplabV3+ 网络高出 0.0774 和 0.0654。在 7 种地物的类别精度方面, MS-XDeeplabV3+ 对道路的 IoU 相比其他三种网络有明显提高, 达到 0.4774, 各类别平均 IoU 分别高出 MDeeplabV3+、MS-MDeeplabV3+ 和 XDeeplabV3+ 网络 0.1036、0.0622 和 0.0766, 表明特征提取网络采用 Xception 的同时对解码区进行改进, 能显著提高遥感影像分类精度。

总体来说, 同时考虑逐层特征融合、注意力模块和多尺度监督的 MS-XDeeplabV3+ 分类效果最好, OA 是 0.9122, 但对道路和草地类的 IoU 最高不超过 0.48 和 0.34, 表明该模型对道路和草地的分类精度还有待提高。

3.4.2 4 种模型与其他模型比较

为了验证所提模型的优势, 对其与传统的遥感影像分类方法 (FCN、U-Net、SegNet、DeeplabV3+) 以及基于 DeeplabV3+ 网络改进的遥感影像分类方法 (E-Deeplab^[13]、文献[18]中的方法、文献[19]中的方法) 进行对比, 结果如表 5 所示。由表 5 可知, 在参数量和训练时间方面, MDeeplabV3+、XDeeplabV3+、MS-MDeeplabV3+ 和 MS-XDeeplabV3+ 网络明显优于其他网络。在精度方面, MDeeplabV3+ 由于网络本身参数量的限制, 精度不高, mIoU 相对 E-Deeplab、文献[18]中的方法和文献[19]中的方法分别低 0.0901、0.0687 和 0.0603。在 MDeeplabV3+ 基础上构建的 MS-MDeeplabV3+ 的精度相对 MDeeplabV3+ 提高 0.0414。XDeeplabV3+ 网络的 mIoU 为 0.6204, 分别比 E-Deeplab、文献[18]中的方法和文献[19]中的方法低 0.0631、0.0417 和 0.0333。在 XDeeplabV3+ 基础上构建的 MS-XDeeplabV3+ 网络精度最高, 其 mIoU 为

0.6970, 相比 E-Deeplab、文献[18]中的方法和文献[19]中的方法分别高出 0.0135、0.0349 和 0.0433。实验结果表明, 本文构建的模型的优势在于减少模型参数量、加快模型训练速度的同时, 提高了分类精度。

表 5 不同网络模型的训练结果对比

Table 5 Comparison of different network models' training results

Method	Parameter size /MB	Time / min	mIoU
FCN	364.7	37.8	0.5586
U-Net	305.3	33.6	0.5739
SegNet	307.6	34.5	0.5748
DeeplabV3+	312.2	39.1	0.6199
E-Deeplab	387.3	47.8	0.6835
Algorithm in Ref. [18]	246.4	37.2	0.6621
Algorithm in Ref. [19]	332.6	42.4	0.6537
MDeeplabV3+	52.7	13.3	0.5934
XDeeplabV3+	148.5	18.7	0.6204
MS-MDeeplabV3+	55.3	17.6	0.6348
MS-XDeeplabV3+	151.1	23.9	0.6970

3.4.3 多尺度监督权重对分类精度的影响

MS-MDeeplabV3+ 和 MS-XDeeplabV3+ 网络通过 4 个侧输出进行多尺度监督, 如图 4 所示。对每侧输出设置不同的权重参数, 以评估多尺度监督对遥感影像分类精度的影响, 采用 mIoU、像素精度和 Kappa 系数进行定量评价, 结果如表 6 和表 7 所示。实验 1 将最后阶段输出的权重设为 1, 其余侧输出设为 0, 表示网络不采用多尺度监督; 实验 2 参考 Zhou 等^[16]评估深度监督对道路提取性能影响的实验, 对不同阶段的侧输出赋予不同的权重, 两个分辨率较高的侧输出赋予较

表 6 MS-MDeeplabV3+ 输出损失权重比较

Table 6 Comparison of different loss weights for MS-MDeeplabV3+

ID	Weight of side output				mIoU	OA	Kappa
	D ₁	D ₂	D ₃	D ₄			
1	0	0	0	1	0.6159	0.8214	0.7975
2	0.3	0.3	0.8	1	0.6317	0.8485	0.8283
3	1	1	1	1	0.6348	0.8502	0.8297

表 7 MS-XDeeplabV3+ 输出损失权重比较

Table 7 Comparison of different loss weights for MS-XDeeplabV3+

ID	Weight of side output				mIoU	OA	Kappa
	D ₁	D ₂	D ₃	D ₄			
1	0	0	0	1	0.6681	0.8827	0.8447
2	0.3	0.3	0.8	1	0.6933	0.9096	0.8629
3	1	1	1	1	0.6970	0.9122	0.8646

大的权重,分别为 1 和 0.8;实验 3 将 4 个侧输出赋予同样的权重,均为 1。实验结果表明,与不采用多尺度监督的实验 1 相比,实验 2 和实验 3 在所有评价指标上分别提高约 0.2 和 0.3,说明多尺度监督有助于模型训练过程,提高模型精度。

4 结 论

基于 DeeplabV3+ 网络,构建 4 种具有不同编解码结构的网络。为减少模型参数量和训练时间,在 DeeplabV3+ 网络的编码区引入 MobilenetV2 和 Xception_65 两种轻量级网络,构建 MDeeplabV3+ 和 XDeeplabV3+ 网络。在此基础上,为提高模型分类精度,在解码区引入逐层特征融合实现解码区上采样的细化,融合具有更多位置信息的浅层特征,引入通道注意力模块突出重要特征,减少信息冗余,引入多尺度监督实现感受野自适应,构建 MS-MDeeplabV3+ 和 MS-XDeeplabV3+ 网络。实验结果表明,编码器采用 Xception_65,解码器同时引入逐层特征融合、通道注意力模块和多尺度监督的 MS-XDeeplabV3+ 网络在减少模型参数量、加快模型训练速度的同时能更细化地物的边缘信息,提高分类精度。

参 考 文 献

- [1] 胡杰, 张莹, 谢仕义. 国产遥感影像分类技术应用研究进展综述[J]. 计算机工程与应用, 2021, 57(3): 1-13.
Hu J, Zhang Y, Xie S Y. Summary of research progress on application of domestic remote sensing image classification technology[J]. Computer Engineering and Applications, 2021, 57(3): 1-13.
- [2] 张乾坤, 蒙继华, 任超. 构建地块二维表征及 CNN 模型的作物遥感分类[J]. 遥感学报, 2022, 26(7): 1437-1449.
Zhang Q K, Meng J H, Ren C. Crop classification based on two-dimensional representation and CNN model from remote sensing[J]. National Remote Sensing Bulletin, 2022, 26(7): 1437-1449.
- [3] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(4): 640-651.
- [4] 谢锦莹, 丁丽霞, 王志辉, 等. 基于 FCN 与面向对象的滨海湿地植被分类[J]. 林业科学, 2020, 56(8): 98-106.
Xie J Y, Ding L X, Wang Z H, et al. Classification of coastal wetland vegetation utilizing FCN and object-oriented methods[J]. Scientia Silvae Sinicae, 2020, 56(8): 98-106.
- [5] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[M]//Navab N, Hornegger J, Wells W M, et al. Medical image computing and computer-assisted intervention-MICCAI 2015. Lecture notes in computer science. Cham: Springer, 2015, 9351: 234-241.
- [6] 李传林, 黄风华, 胡威, 等. 基于 Res-AttentionUnet 的高分辨率遥感影像建筑物提取方法[J]. 地球信息科学学报, 2021, 23(12): 2232-2243.
Li C L, Huang F H, Hu W, et al. Building extraction from high-resolution remote sensing image based on Res-AttentionUnet[J]. Journal of Geo-Information Science, 2021, 23(12): 2232-2243.
- [7] 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.
- [8] 张哲晗, 方薇, 杜丽丽, 等. 基于编码-解码卷积神经网络的遥感图像语义分割[J]. 光学学报, 2020, 40(3): 0310001.
Zhang Z H, Fang W, Du L L, et al. Semantic segmentation of remote sensing image based on encoder-decoder convolutional neural network[J]. Acta Optica Sinica, 2020, 40(3): 0310001.
- [9] Chen L C, Papandreou G, Kokkinos I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs[EB/OL]. (2014-12-22) [2022-06-08]. <https://arxiv.org/abs/1412.7062>.
- [10] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.
- [11] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation[EB/OL]. (2017-06-17) [2022-05-08]. <https://arxiv.org/abs/1706.05587>.
- [12] Chen L C, Zhu Y K, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 833-851.
- [13] 许泽宇, 沈占锋, 李杨, 等. 增强型 DeepLab 算法和自适应损失函数的高分辨率遥感影像分类[J]. 遥感学报, 2022, 26(2): 406-415.
Xu Z Y, Shen Z F, Li Y, et al. Classification of high-resolution remote sensing images based on enhanced DeepLab algorithm and adaptive loss function[J]. National Remote Sensing Bulletin, 2022, 26(2): 406-415.
- [14] Chollet F. Xception: deep learning with depthwise separable convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 1800-1807.
- [15] Sandler M, Howard A, Zhu M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 4510-4520.
- [16] Zhou M T, Sui H G, Chen S X, et al. BT-RoadNet: a boundary and topologically-aware neural network for road extraction from high-resolution remote sensing imagery

- [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 168: 288-306.
- [17] Huang J F, Zhang X C, Sun Y, et al. Attention-guided label refinement network for semantic segmentation of very high resolution aerial orthoimages[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 4490-4503.
- [18] 朱戎, 叶宽, 杨博, 等. 基于改进 DeeplabV3+ 的地物分类方法研究[J]. 计算机科学, 2021, 48(S2): 382-385.
- Zhu R, Ye K, Yang B, et al. Feature classification method based on improved DeeplabV3+ [J]. Computer Science, 2021, 48(S2): 382-385.
- [19] 徐志凡, 杜洪波, 韩承霖, 等. 基于 DeepLab V3+ 改进的图像语义分割模型[J]. 智能计算机与应用, 2021, 11(6): 181-184.
- Xu Z F, Du H B, Han C L, et al. Improved based on DeepLab V3+ network[J]. Intelligent Computer and Applications, 2021, 11(6): 181-184.