

基于轻量语义分割网络的遥感土地覆盖分类^①

朱婉玲, 贾 渊

(西南科技大学 计算机科学与技术学院, 绵阳 621010)

通信作者: 贾 渊, E-mail: 114997152@qq.com



摘 要: 高分辨率遥感图像有丰富的空间特征, 针对遥感土地覆盖方法中模型复杂, 边界模糊和多尺度分割等问题, 提出了一种基于边界与多尺度信息的轻量化语义分割网络. 首先, 使用轻量化的 MobileNetV3 分类器, 采用深度可分离卷积来减少计算量. 其次, 使用自顶向下和自底向上的特征金字塔结构来进行多尺度分割. 接着, 设计了一个边界增强模块, 为分割任务提供丰富的边界细节信息. 然后, 设计了一个特征融合模块, 融合边界与多尺度语义特征. 最后, 使用交叉熵损失函数和 Dice 损失函数来处理样本不平衡的问题. 在 WHDL D 数据集的平均交并比达到了 59.64%, 总体精度达到了 87.68%. 在 DeepGlobe 数据集的平均交并比达到了 70.42%, 总体精度达到了 88.81%. 实验结果表明, 该模型能快速有效地实现遥感图像土地覆盖分类.

关键词: 高分辨率遥感图像; 土地覆盖分类; 轻量化语义分割; 多尺度; 边界增强; 卷积神经网络

引用格式: 朱婉玲, 贾渊. 基于轻量语义分割网络的遥感土地覆盖分类. 计算机系统应用, 2024, 33(2): 134-142. <http://www.c-s-a.org.cn/1003-3254/9404.html>

Remote Sensing Land Cover Classification Based on Lightweight Semantic Segmentation Network

ZHU Wan-Ling, JIA Yuan

(School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang 621010, China)

Abstract: High-resolution remote sensing images have rich spatial features. To solve the problems of complex models, blurred boundaries, and multi-scale segmentation in remote sensing land cover methods, this study proposes a lightweight semantic segmentation network based on boundary and multi-scale information. First, the method uses a lightweight MobileNetV3 classifier and depthwise separable convolutions to reduce computation. Second, the method adopts top-down and bottom-up feature pyramid structures for multi-scale segmentation. Next, a boundary enhancement module is designed to provide rich boundary detail information for the segmentation task. Then, the method designs a feature fusion module to fuse boundary and multi-scale semantic features. Finally, the method applies cross-entropy and Dice loss functions to deal with the sample imbalance. The mean intersection over union of the WHDL D dataset reaches 59.64%, and the overall accuracy reaches 87.68%. The mean intersection over union of the DeepGlobe dataset reaches 70.42%, and the overall accuracy reaches 88.81%. The experimental results show that the model can quickly and effectively realize the land cover classification of remote sensing images.

Key words: high-resolution remote sensing image; land cover classification; lightweight semantic segmentation; multiscale; border enhancement; convolutional neural network (CNN)

土地覆盖指的是地表上的自然物体和人工建造的物体, 获取准确的土地覆盖类别, 有助于人们管理与保护

土地资源^[1]. 土地覆盖分类研究, 目前主要使用遥感数据源, 可以从遥感图像中及时获取准确的土地覆盖信息.

① 基金项目: 国家自然科学基金 (NSFC62076209)

收稿时间: 2023-08-12; 修改时间: 2023-09-28; 采用时间: 2023-10-09; csa 在线出版时间: 2023-12-18

CNKI 网络首发时间: 2023-12-19

传统的遥感土地覆盖分类方法,如目视解译法,该方法简单,但花费时间长且分类精度不高.机器学习的方法,如支持向量机^[2]和决策树分类方法^[3]等方法,无法满足时效性和端到端的解译要求,而且无法满足高分辨率遥感图像中丰富信息的解译要求.

基于像素的土地覆盖分类方法,是按照地物的类别,划分遥感图像中的每一个像素.遥感技术不断发展,高分辨率遥感图像能充分呈现地物的信息,但也加大了土地覆盖分类任务的难度.基于深度学习的卷积神经网络能自适应地提取图像的特征,端到端的语义分割技术可用于土地覆盖分类任务.Zhang等人使用遥感图像的空间和光谱信息,为土地覆盖分类任务提供端到端的深度学习方法,实验结果优于传统机器学习的方法^[4].Chen等人使用并行的空洞卷积来扩大感受野,以应对多尺度分割问题^[5].Zheng等人提出前景建模的方法来处理遥感图像中前景与背景不平衡的问题^[6].Yu等人提出了一种高效的双路径模型来处理空间和分类语义^[7].Xu等人提出了三支的实时语义分割网络,在推理时间和准确性之间实现了合理平衡^[8].刘明威等人提出了一种基于U-Net改进的遥感图像语义分割模型,实现了土地覆盖分类任务的自动化^[9].何直蒙等人提出一种基于空洞卷积的深度学习网络,来减少提取高分辨率遥感图像的建筑物时的细节损失^[10].遥感图像中有多种地物,每种地物所需要的分割尺度有差异,所以遥感图像中多尺度,是需要改善的问题.而且在基于深度学习的语义分割方法中,模型复杂,分割结果的边界模糊,也是需要关注的问题.

本文提出了一种基于边界与多尺度信息的轻量化语义分割网络(lightweight semantic segmentation network based on boundary and multi-scale, LSSN-BM),用于土地覆盖分类任务.主要工作如下:(1)首先,本文使用轻量化的MobileNetV3模型^[11]作为分类器,并且在之后的分割模块中也采用深度可分离卷积来减少计算量.(2)其次,多分支编码器模块中使用了PAN结构^[12]来进行多尺度分割,它包含自顶向下和自底向上的特征金字塔结构,将高层特征与低层特征充分融合.(3)接着,设计了一个边界增强模块,它通过学习边界信息,为分割任务提供丰富的边界细节信息.(4)然后,设计了一个特征融合模块,融合边界特征与多尺度语义特征.(5)最后,对损失函数的使用进行探究,本文使用组合损失函数,语义分支使用多分类交叉熵损失函数,而边界增强模块使用二分类交叉熵损失函数和

Dice 损失函数的结合.

1 LSSN-BM 网络整体结构

本文提出一种基于边界与多尺度信息的轻量化语义分割网络(LSSN-BM),来改善遥感土地覆盖分类的模型复杂,边界模糊和多尺度分割等问题.

整体结构如图1,分类器采用MobileNetV3-Large版本的网络结构,本文将MobileNetV3的4个不同尺寸大小的特征图作为多分支编码器的输入.图1(a)多分支编码器使用PAN结构,它在FPN的自顶向下的结构基础上,增加了自底向上的结构,能充分融合高层与低层特征.图1(b)是边界增强模块,可以为遥感土地覆盖分类任务提供丰富的边界细节信息,其中边界模块使用二分类交叉熵损失函数和Dice损失函数.图1(c)是特征融合模块,解码多尺度语义特征,再和边界特征融合.图1中,C表示通道数,up2表示上采样两倍,down2表示下采样两倍,up4表示上采样4倍.

2 LSSN-BM 网络详细介绍

2.1 MobileNetV3 分类器

MobileNetV3模型^[11]是一个轻量化的网络,它使用深度可分离卷积来减少模型复杂度,同时使用瓶颈结构,增加特征图的通道数,以增加非线性表现力.

MobileNetV3首先通过一个传统的卷积块提取特征,再使用多个瓶颈模块继续提取特征,最后使用 1×1 卷积调整通道数.本文的分类器去掉了MobileNetV3-Large的平均池化和全连接层,将4个不同尺寸大小的特征图作为多分支编码器的输入.它们分别是主干网络4倍下采样、8倍下采样、16倍下采样和32倍下采样之后的特征图,输出通道数分别为24、40、112和960.

2.2 深度可分离卷积

MobileNetV1网络^[13]中使用了深度可分离卷积(depthwise separable convolution, DSC)来减少计算量,它是深度卷积(depthwise convolution, DW)和逐点卷积(pointwise convolution, PW)的组合.PW用来进行通道的融合,以实现改变特征图的通道数.DW在每个通道上进行卷积,将单个滤波器应用到每个通道.深度可分离卷积与普通卷积的计算量比值的计算如下:

$$s = \frac{1}{N} + \frac{1}{D_k^2} \quad (1)$$

其中, N 表示输出通道数,卷积核大小 D_k^2 .即使用卷积

核大小为 3×3 的深度可分离卷积, 其计算量下降到普通卷积的 1/9~1/8, 精度仅略有下降^[13].

本文在特征融合模块和边界增强模块中均使用了深度可分离卷积, 即图 1 中的 DSC 和 UP_DSC 模块.

DSC 模块的详细结构如图 2 所示, 先使用 3×3 的深度卷积提取特征, 然后是 1×1 的逐点卷积改变特征图的通道数. 使用 DSC 模块来替换 3×3 的普通卷积, 能减少模型计算量.

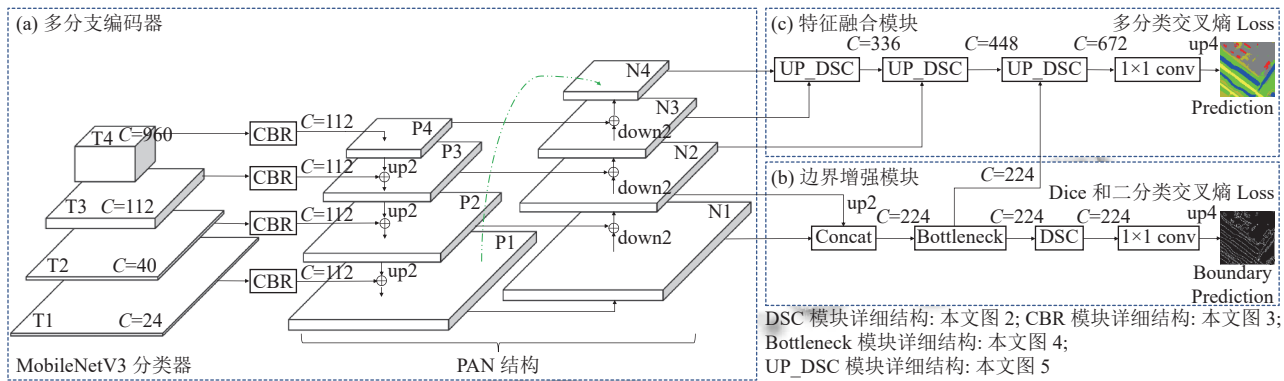


图 1 LSSN-BM 网络结构

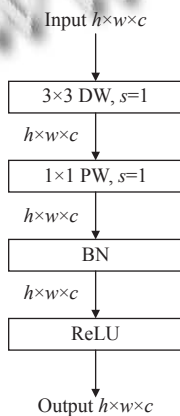


图 2 DSC 模块

2.3 多分支编码器

文中的多分支编码器使用 PAN 模块^[12], 它在 FPN 的自顶向下的结构基础上, 再加入了自底向上的路径增强, 用低层特征图中的精确的定位信息来增强特征金字塔. 这样的结构既包含较强的高层语义特征, 又包含低层的高分辨率特征. 为了将低层特征快速流动到顶层特征, 还加入了跳连接. 多分支编码器中的 CBR 模块如图 3 所示.

多分支编码模块把 MobileNetV3 分类器输出的 $\{T_i | i = 1, 2, 3, 4\}$ 特征图作为输入. 如图 1(a) 所示, 在自顶向下的特征融合路径, 首先使用 CBR 模块, 以改变特征图的通道数. 然后再把该特征图上采样两倍后, 与低一层的特征图相加融合, 得到 $\{P_i | i = 1, 2, 3, 4\}$ 特征图.

CBR 模块中有两种结构, 适应的场景将在第 3.3 节讨论. CBR (A) 是为了改变特征图的通道数, 而 CBR (B) 受空洞空间卷积池化金字塔^[5]的启发, 使用并行的不同膨胀率的空洞卷积来增加感受野, 这样可以适应多尺度的地物分割. 同时还加入了全局平均池化, 以获得更多的全局信息. 最后再使用一个 3×3 的卷积继续提取特征. 为了减少计算量, CBR (B) 模块中都采用深度可分离卷积.

在自底向上的结构中, 特征图被下采样两倍再与高一层的特征图相加融合, 得到 $\{N_i | i = 1, 2, 3, 4\}$ 特征图. 其中 N_4 特征图的获取还加入了跳连接, 使低层 P_1 的特征快速流动到高层.

2.4 边界增强模块

本文设计了一个边界增强模块, 为语义分支提供边界细节信息, 以解决分类结果边界模糊的问题. Canny 边缘检测算子^[14], 具有较高的准确率和稳定性, 可以获取图像的边缘. 边界的真实标记图, 是使用 Canny 算法从已知的真实标记语义图中获取的.

He 等人^[15]为减轻深层网络中的梯度消失等问题, 设计了具有跳跃连接的残差结构. 瓶颈残差模块如图 4 所示, 它是先降低通道数, 然后使用 3×3 卷积继续提取特征, 最后再调整输出特征图的通道数.

由于低层特征图有更高的分辨率和较多的细节信息, 所以我们选取较低层的特征图来预测边界. 如图 1(b) 所示, 首先使用 Concat 操作实现 N_1 和 N_2 特征

图的拼接, 再使用瓶颈残差模块^[15]和 DSC 模块继续提取特征, 然后使用 1×1 卷积调整通道数后, 最后把特征

图上采样到与 Canny 边缘检测后的边界标记图尺寸一致.

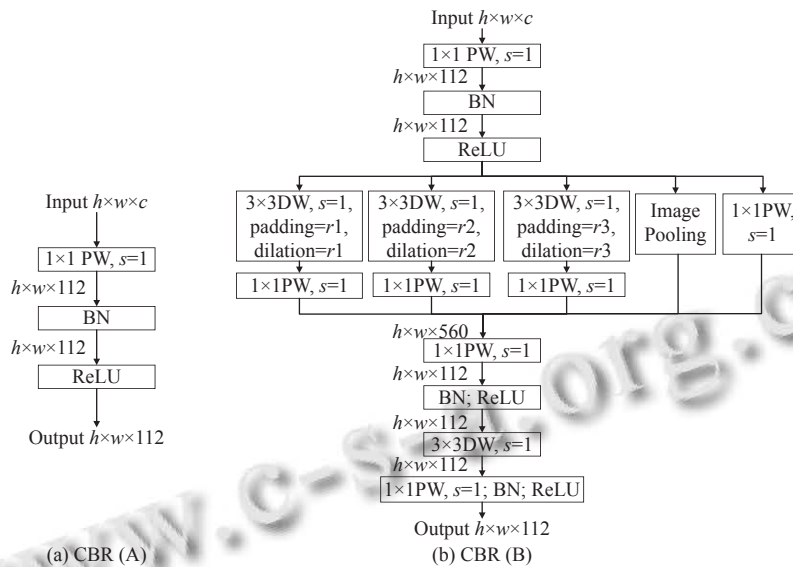


图3 CBR 模块

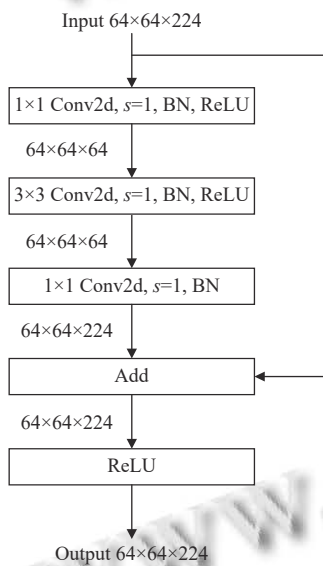


图4 瓶颈残差模块

2.5 特征融合模块

本文设计了一个特征融合模块, 能充分融合边界增强模块的输出特征与多尺度分支编码器输出的语义特征, 以提高分类精确度. 图 1(c) 是特征融合模块, 它使用多个 UP_DSC 模块来融合多尺度特征. 本文设计的 UP_DSC 模块如图 5 所示, 它有两个输入特征图, 先把低分辨率特征图进行两倍上采样, 再把低分辨率特征图与高分辨特征图 Concat 连接, 最后使用深度可分离卷积, 即 DSC 模块提取特征.

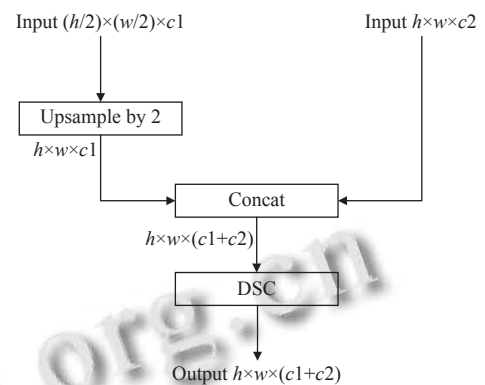


图5 UP_DSC 模块

多尺度分支编码器的输出特征图, 分别使用 UP_DSC 模块融合多尺度语义特征, 得到丰富的语义特征图. 最后再使用 UP_DSC 将语义特征图与边界增强模块的输出特征图进行融合, 经过 1×1 卷积调整通道数和 4 倍上采样得到最终的预测图.

2.6 组合损失函数

本文设计了一个组合损失函数, 语义分支使用多分类交叉熵损失函数, 边界增强模块使用二分类交叉熵损失函数和 Dice 损失函数的结合. 在文献[6]中介绍了推广到像素级别的 Focal 损失函数, 以解决样本不平衡的问题. 文献[16]中介绍了用于在线困难样本挖掘的 OHEM 损失函数. 通过实验, 语义分支选取了交叉熵损失函数.

文献[17]总结了多种可用于语义分割的损失函数,其中包含多分类交叉熵、二分类交叉熵和 Dice 损失函数等. 交叉熵损失函数被广泛用于分类目标, 由于分割其实是像素级别的分类, 所以它能应用到语义分割上. 多分类交叉熵损失计算如下:

$$L_{CE}(y, p) = -\frac{1}{N} \sum_{i=0}^{n-1} \sum_{k=0}^{k-1} (y_{i,k} \log(p_{i,k})) \quad (2)$$

其中, y 是真实标签, p 是预测标签, n 表示样本数, k 是标签数, $p_{i,k}$ 表示将第 i 个样本预测为第 k 个标签的概率.

Dice 系数用于计算两张图像间的相似性, 可以缓解样本不平衡的问题. 二分类交叉熵虽然计算简单, 但不能很好地处理数据类别不平衡的情况. 由于边界增强模块中, 在对边界和背景二分类时, 边界与背景极其不平衡, 所以本文还引入了 Dice 损失函数来处理该问题, 如式 (3). 但是 Dice 损失容易带来训练不稳定的情况, 所以在边界模块选取二分类交叉熵和 Dice 损失. 式 (4) 表示二分类交叉熵, 其中 p 是预测值, y 是真实值. 本文的组合损失函数, 是边界增强模块的损失函数和语义分支的损失函数组合, 如式 (5) 所示.

$$L_{Dice}(y, p) = 1 - \frac{2yp + 1}{y + p + 1} \quad (3)$$

$$L_{BCE}(y, p) = -(y \log(p) + (1 - y) \log(1 - p)) \quad (4)$$

$$L_{total} = L_{CE} + L_{Dice} + L_{BCE} \quad (5)$$

3 实验分析

3.1 数据集

WHDL D 数据集^[18]是遥感图像多分类数据集, 可用于遥感土地覆盖分类任务. WHDL D 数据集的图像共为 4940 张, 空间大小为 256×256 , 分辨率为 2 m 的 RGB 图像. WHDL D 数据集是按像素进行标记, 标记为 6 类区域, 分别是裸土、建筑物、人行道、道路、植被和水. 在本次实验中, WHDL D 数据集被划分为训练集有 2968 张、验证集有 992 张和测试集有 980 张.

DeepGlobe 土地覆盖分类数据集^[19]有 803 张图像, 其空间大小为 2448×2448 , 分辨率为 0.5 m. DeepGlobe 数据集是按像素标注为 7 类区域, 包括城市、农业、牧场、森林、水、贫瘠和未知区域, 其中未知区域在分类时不需要考虑. DeepGlobe 数据集划分为训练集 644 张和测试集 159 张.

3.2 实验设置及评价指标

实验使用 Python 3.7.11 和 PyTorch 1.11.0 的环境, Ubuntu 20.04 操作系统和具有 10 GB 显存的 GeForce RTX 3080 显卡. 所有骨干网络都使用 ImageNet 训练后的权重. 在实验中, 优化器使用 SGD 策略, 并将权重衰减设置为 0.000 1, 将动量设置为 0.9. 学习策略使用 Poly 动态调整学习率, 初始学习率为 0.01, 动力为 0.9. 训练的迭代次数为 15 000 次. WHDL D 数据集批量大小为 22. DeepGlobe 数据集批量大小为 4. 由于计算机的显存有限, 所以 DeepGlobe 数据集在训练时, 将输入图片大小调整为 1024×1024 . 在实验中的训练的时候采用了数据增强, 主要使用随机水平翻转、随机垂直翻转和随机旋转 90° 方法.

实验使用 3 个遥感土地覆盖分类任务中的评价标准: 平均交并比 (mean intersection over union, $mIOU$)、总体精度 (overall accuracy, OA) 和 $F1$ 分数 ($F1$ -score). 具体计算公式如下:

$$mIOU = \frac{1}{k+1} \sum_{i=0}^k \frac{TP}{FN + FP + TP} \quad (6)$$

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

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

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

$$F1\text{-score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (10)$$

其中, TP 表示预测成功的正例数目, FP 表示预测失败的正例数目, FN 表示预测失败的负例数目, TN 表示预测成功负例的数目. $F1$ -score 是综合评价精确率 ($Precision$) 和召回率 ($Recall$) 的指标.

模型的复杂度可以用参数量 (Parameters) 和网络推理中的累计乘加次数 (Mult-adds) 来评价. Parameters 用来评价模型的空间复杂度, 而 Mult-adds 用来评价模型的时间复杂度, 并且能反映出模型的分割速度. 模型推理的实时性还可以采用 FPS 来评价, 即每秒传输帧数.

3.3 消融实验结果及分析

WHDL D 数据集的多分支编码器均采用 CBR (A) 结构. 模型 LSSN-BM, 在边界增强模块使用二分类交叉熵和 Dice 损失的基础上, 对语义分支的损失函数选取进行了实验, 如表 1 所示. LSSN-BM_A 的语义分支

使用 Focal 损失函数, LSSN-BM_B 使用 OHEM 损失, LSSN-BM_C 使用交叉熵损失. 从表 1 中可以看出,

LSSN-BM_C 的实验结果表现更好, 即语义分支选取交叉熵损失函数.

表 1 WHDL D 测试集上的损失函数选取实验

Algorithm	Focal	OHEM	交叉熵损失	Parameters (M)	Mult-adds (G)	<i>mIOU</i> (%)	<i>OA</i> (%)	<i>F1-score</i> (%)
LSSN-BM_A	有	无	无	4.34	4.61	59.18	87.24	72.24
LSSN-BM_B	无	有	无	4.34	4.61	59.04	87.51	71.99
LSSN-BM_C	无	无	有	4.34	4.61	59.64	87.68	72.57

在语义分支使用交叉熵损失函数的基础上, 对每个模块的有效性进行了实验, 如表 2 所示. LSSN-BM_D, 只使用多分支编码器. LSSN-BM_E 继续加入了特征融合模块, *mIOU* 提高了 1%, *OA* 提高了 0.43%, *F1-score* 提高了 0.81%, 表明特征融合模块的有效性. LSSN-BM_C

是继续加入了边界增强模块, 更多保留了边界细节信息, 为语义分支提供边界细节信息. 从表 2 可以看出, 相比于 LSSN-BM_E, LSSN-BM_C 的 *mIOU* 再提高了 1.63%, *OA* 提高了 0.72%, *F1-score* 提高了 1.41%, 虽然模型复杂度略有变大, 但仍然在可接受范围内.

表 2 WHDL D 测试集上的消融实验

Algorithm	多分支编码器	特征融合	边界增强	Parameters (M)	Mult-adds (G)	<i>mIOU</i> (%)	<i>OA</i> (%)	<i>F1-score</i> (%)
LSSN-BM_D	有	无	无	3.36	1.17	57.01	86.53	70.35
LSSN-BM_E	有	有	无	3.48	1.3	58.01	86.96	71.16
LSSN-BM_C	有	有	有	4.34	4.61	59.64	87.68	72.57

最终 WHDL D 数据集的模型选择 LSSN-BM_C 的设计, 选择 MobileNetV3 分类器, 同时使用了多分支编码器、特征融合模块和边界增强模块, 语义分支选择交叉熵损失函数.

由于 DeepGlobe 数据集的图片尺寸大于 WHDL D 数据集的图片尺寸, 对于 DeepGlobe 数据集来说, 获得更多全局信息对分类任务有益, 所以引入了 CBR (B) 结构. 表 3 中第 1 行表示全部使用 CBR (A) 结构. 表 3 中第 2 行表示多分支编码器中低层的 T_1 和 T_2 特征图

采用 CBR (A) 结构, 而高层的 T_3 和 T_4 特征图采用 CBR (B) 结构, 且空洞卷积的膨胀率为 $r=[12, 25, 37]$. 与第 2 行相比, 表 3 中第 3 行是把 CBR (B) 中空洞卷积的膨胀率改为 $r=[6, 13, 19]$. 由此可以看出, 膨胀率 $r=[6, 13, 19]$ 的空洞卷积, 表现结果最好, 相比于全部使用 CBR (A) 结构, *mIOU* 提高了 6.69%, *OA* 提高了 4.09%, *F1-score* 提高了 5.27%. 所以 DeepGlobe 数据集的模型选择了 LSSN-BM_C2 的设计, 即膨胀率为 $r=[6, 13, 19]$.

表 3 DeepGlobe 测试集上的膨胀率选取实验

Algorithm	全用 CBR (A)	$r=[12, 25, 37]$	$r=[6, 13, 19]$	Parameters (M)	Mult-adds (G)	<i>mIOU</i> (%)	<i>OA</i> (%)	<i>F1-score</i> (%)
LSSN-BM_C	有	无	无	4.34	73.77	63.73	84.72	76.43
LSSN-BM_C1	无	有	无	4.49	74.44	69.27	88.62	80.68
LSSN-BM_C2	无	无	有	4.49	74.44	70.42	88.81	81.70

3.4 对比实验结果及分析

为进一步验证本文方法的有效性, 将本文模型与目前主流的语义分割模型 Farseg 模型^[6]、BiSeNetV2 模型^[7]和 PIDNet^[8]进行对比. WHDL D 的测试集上的模型复杂度对比如表 4 所示. WHDL D 的测试集上的实验结果对比如表 5 所示. LSSN-BM_C 网络相比其他方法, *mIOU*、*OA* 和 *F1-score* 均有提升, *mIOU* 达到了 59.64%, *OA* 达到了 87.68%, *F1-score* 达到了 72.57%.

在 WHDL D 的测试集上的结果显示, 相比于 Farseg

模型, LSSN-BM_C 的 *mIOU* 提高了 2.55%, *OA* 提高了 1.23%, *F1-score* 提高了 2.28%. 而且对比 Farseg 模型, LSSN-BM_C 模型的参数量减少了 86.17%, 累计乘加次数减少了 60.73%, 且 FPS 更优. 实验结果表明, 本文方法更容易部署且具有更快的推理速度.

表 4 WHDL D 测试集上的模型复杂度对比

Algorithm	Parameters (M)	Mult-adds (G)	FPS (Hz)
Farseg	31.37	11.74	67.25
BiSeNetV2	5.2	3.07	108.45
PIDNet	37.31	8.98	63.73
LSSN-BM_C	4.34	4.61	79.02

表5 WHDL D 测试集上的实验结果对比 (%)

Algorithm	OA	F1-score	mIOU	IOU					
				裸土	建筑物	人行道	道路	植被	水
Farseg	86.45	70.29	57.09	33.09	56.76	31.80	51.23	85.35	84.31
BiSeNetV2	85.74	67.74	54.48	34.10	53.04	24.35	47.89	84.81	82.71
PIDNet	85.77	68.95	55.68	37.38	53.96	25.43	49.14	84.86	83.32
LSSN-BM_C	87.68	72.57	59.64	39.09	56.16	33.58	54.71	86.65	87.66

WHDL D 测试集上, 相比于轻量级 BiSeNetV2 模型, LSSN-BM_C 的 *mIOU* 提高了 5.16%, *OA* 提高了 1.94%, *F1-score* 提高了 4.83%。虽然 LSSN-BM_C 的累计乘加次数要略优于 BiSeNetV2 模型, 但 LSSN-BM_C 模型的参数量更少。实验结果表明, 本文方法在满足模型轻量化同时, 能取得更好的分类效果。

在 WHDL D 的测试集上的结果显示, 相比于 PIDNet 实时语义分割模型, LSSN-BM_C 的 *mIOU* 提高了 3.96%, *OA* 提高了 1.91%, *F1-score* 提高了 3.62%。而且对比 PIDNet 模型, LSSN-BM_C 模型的参数量减少了 88.37%, 累计乘加次数减少了 48.66%, FPS 更优。实验结果表明, 本文方法在满足实时性要求的同时, 分类效果更好。

IOU 值反映出不同类型地物的分类效果, LSSN-BM

网络对裸土、人行道、道路、植被和水的 *IOU* 值均有提升。且水和植被的分类结果优于其他类别。

在 DeepGlobe 测试集上的模型复杂度对比如表 6 所示。在 DeepGlobe 的测试集上的实验结果对比如表 7 所示。本文方法 LSSN-BM_C2 的 *mIOU* 达到了 70.42%, *OA* 达到了 88.81%, *F1-score* 达到了 81.70%。DeepGlobe 输入网络的图片尺寸为 1024×1024, 图片尺寸越大对模型轻量化有着更高的要求。

表6 DeepGlobe 测试集上的模型复杂度对比

Algorithm	Parameters (M)	Multi-adds (G)	FPS (Hz)
Farseg	31.37	187.85	28.52
BiSeNetV2	5.2	49.12	79.79
PIDNet	37.31	143.72	51.21
LSSN-BM_C2	4.49	74.44	47.73

表7 DeepGlobe 测试集上的实验结果对比 (%)

Algorithm	OA	F1-score	mIOU	IOU					
				农业	水	牧场	森林	城市	贫瘠
Farseg	89.13	81.08	69.85	89.79	69.21	38.88	80.71	76.48	64.05
BiSeNetV2	84.29	71.36	58.31	85.05	44.81	24.96	76.54	69.80	48.69
PIDNet	75.95	57.93	45.64	76.55	22.48	10.88	74.99	62.34	26.61
LSSN-BM_C2	88.81	81.70	70.42	88.84	69.60	43.49	82.00	76.05	62.51

DeepGlobe 的测试集上的结果显示, 相比于 Farseg 模型, LSSN-BM_C2 的 *mIOU* 提高了 0.57%, *F1-score* 提高了 0.62%。虽然 LSSN-BM_C2 的 *OA* 略低于 Farseg, 但参数量减少了 85.69%, 累计乘加次数减少了 60.37%, 且 FPS 更优。由此可以看出 LSSN-BM_C2 方法运行速度更快且更容易部署, 综合结果更优。

DeepGlobe 的测试集的结果, 相比于轻量级 BiSeNetV2 模型, LSSN-BM_C2 的 *mIOU* 提高了 12.11%, *OA* 提高了 4.52%, *F1-score* 提高了 10.34%。虽然其累计乘加次数要高于 BiSeNetV2 模型, 但模型参数量更少。在遥感土地覆盖分类应用场景中, 速度与精度的平衡是需要考虑的问题, LSSN-BM_C2 模型能在一定轻量化的情形下, 取得更好的分类效果。

DeepGlobe 的测试集上的结果, 相比于 PIDNet 实时语义分割模型, LSSN-BM_C2 的 *mIOU* 提高了 24.78%, *OA* 提高了 12.86%, *F1-score* 提高了 23.77%。LSSN-BM_C2 模型的部署和推理时间更快, 而且对比

PIDNet 模型, LSSN-BM_C2 模型的参数量减少了 87.97%, 累计乘加次数减少了 48.2%。而且 FPS 基本接近, LSSN-BM_C2 方法满足实时性要求, 在该数据集上能大幅提高遥感土地覆盖分类任务精度。

在遥感土地覆盖分类应用场景中, 速度与精度的平衡是需要考虑的问题, LSSN-BM_C2 模型可以在一定轻量化的情形下, 取得更好的分类效果。DeepGlobe 测试集上不同地物的 *IOU* 值反映出, 在所有土地覆盖类别中, 农业和森林的分类结果优于其他类别。

为了探寻更大尺寸图像的效率和精度, 在表 8 中, 我们将 DeepGlobe 数据集的图片尺寸调整为 1152×1152。此时本文 LSSN-BM_C2 模型的 *mIOU* 为 69.91%, *OA* 为 88.37%。与表 6 相比, 表 8 中的图片尺寸增大后, 每个模型的累计乘加次数有所上升, FPS 速度有所下降, 从中可以看出图片尺寸越大, 处理每张图片会需要更多的时间。而且由于计算机硬件显存的限制, 此时

Farseg 模型已经显存不足,无法进行训练,即图片尺寸越大对模型轻量化有着更高的要求。

图 6 是 WHDL D 的测试集中的一张图像的分割效果,可以看出 LSSN-BM 模型更接近于真实标签,而且对人行道等小目标细节的划分更优,表明边界增强模

块是有效的。图 7 是 DeepGlobe 的测试集中的一张图像的分割效果,可以看出 LSSN-BM 模型明显优于 BiSeNetV2 和 PIDNet 模型,且与复杂 Farseg 模型的结果基本相当。LSSN-BM 模型为遥感图像的土地覆盖分类任务提供了一个轻量级的解决方案

表 8 DeepGlobe 调整尺寸后的实验结果对比

Algorithm	OA (%)	F1-score (%)	mIOU (%)	Parameters (M)	Mult-adds (G)	FPS (Hz)
BiSeNetV2	84.21	71.50	58.22	5.2	62.16	62.07
PIDNet	73.78	61.04	48.11	37.31	181.89	41.42
LSSN-BM_C2	88.37	81.35	69.91	4.49	94.21	37.38

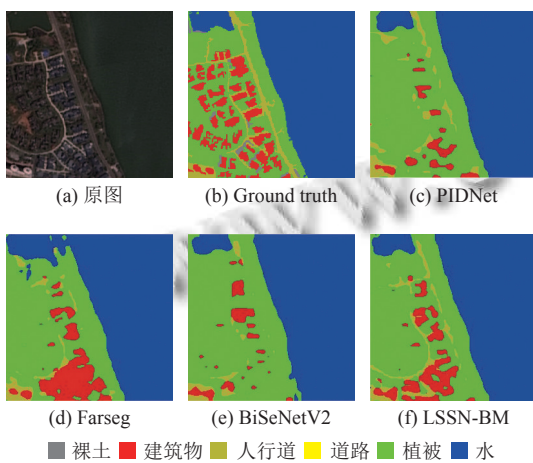


图 6 WHDL D 测试集语义分割结果

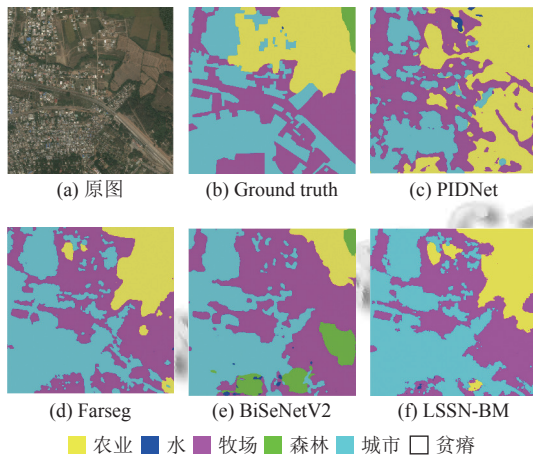


图 7 DeepGlobe 测试集语义分割结果

4 结论与展望

高分辨率遥感图像提供了丰富的纹理和空间分布等信息,给土地覆盖分类任务带来了更大的挑战。针对遥感土地覆盖分类任务的模型复杂,边界模糊和多尺度等问题,本文提出了一种基于边界与多尺度信息的

轻量化语义分割网络 LSSN-BM。采取轻量 MobileNetV3 分类器,之后的分割中也使用深度可分离卷积来减少计算量;多分支编码器采用 PAN 结构来进行多尺度分割;设计了边界增强模块来提供边界细节特征;设计了一个充分融合边界特征与多尺度语义特征的特征融合模块;并且使用组合损失函数。LSSN-BM 网络在 WHDL D 数据集上的 mIOU 达到了 59.64%, OA 达到了 87.68%, F1-score 达到了 72.57%,在 DeepGlobe 数据集上的 mIOU 达到了 70.42%, OA 达到了 88.81%, F1-score 达到了 81.70%。LSSN-BM 网络实现了推理时间和准确性的平衡,且模型更加轻量化,为土地覆盖分类提供了一个可行方案。此外,对土地覆盖分类的变化检测也是未来工作重要的研究方向。

参考文献

- 童心仪. 面向广域土地覆盖的高分辨率遥感影像分类方法研究[博士学位论文]. 武汉: 武汉大学, 2020.
- Khatami R, Mountrakis G, Stehman SV. A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research. Remote Sensing of Environment, 2016, 177: 89-100. [doi: 10.1016/j.rse.2016.02.028]
- Yang C, Wu GF, Ding K, et al. Improving land use/land cover classification by integrating pixel unmixing and decision tree methods. Remote Sensing, 2017, 9(12): 1222. [doi: 10.3390/rs9121222]
- Zhang X, Han LX, Han LH, et al. How well do deep learning-based methods for land cover classification and object detection perform on high resolution remote sensing imagery? Remote Sensing, 2020, 12(3): 417. [doi: 10.3390/rs12030417]
- Chen LC, Zhu YK, Papandreou G, et al. Encoder-decoder

- with atrous separable convolution for semantic image segmentation. Proceeding of the 15th European Conference on Computer Vision (ECCV). Munich: Springer, 2018. 833–851.
- 6 Zheng Z, Zhong YF, Wang JJ, *et al.* Foreground-aware relation network for geospatial object segmentation in high spatial resolution remote sensing imagery. Proceeding of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 4095–4104.
- 7 Yu CQ, Gao CX, Wang JB, *et al.* BiSeNet V2: Bilateral network with guided aggregation for real-time semantic segmentation. International Journal of Computer Vision, 2021, 129(11): 3051–3068. [doi: [10.1007/s11263-021-01515-2](https://doi.org/10.1007/s11263-021-01515-2)]
- 8 Xu JC, Xiong ZX, Bhattacharyya SP. PIDNet: A real-time semantic segmentation network inspired by PID controllers. Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver: IEEE, 2023. 19529–19539.
- 9 刘明威, 方静, 詹曙. 基于深度学习的土地覆盖遥感图像分割方法. 合肥工业大学学报(自然科学版), 2022, 45(6): 753–759.
- 10 何直蒙, 丁海勇, 安炳琪. 高分辨率遥感影像建筑物提取的空洞卷积 E-Unet 算法. 测绘学报, 2022, 51(3): 457–467. [doi: [10.11947/j.AGCS.2022.20200601](https://doi.org/10.11947/j.AGCS.2022.20200601)]
- 11 Howard A, Sandler M, Chen B, *et al.* Searching for MobileNetV3. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2019. 1314–1324.
- 12 Liu S, Qi L, Qin HF, *et al.* Path aggregation network for instance segmentation. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 8759–8768.
- 13 Howard AG, Zhu ML, Chen B, *et al.* MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861.
- 14 Canny J. A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986, PAMI-8(6): 679–698. [doi: [10.1109/TPAMI.1986.4767851](https://doi.org/10.1109/TPAMI.1986.4767851)]
- 15 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770–778.
- 16 Shrivastava A, Gupta A, Girshick R. Training region-based object detectors with online hard example mining. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 761–769.
- 17 Jadon S. A survey of loss functions for semantic segmentation. Proceedings of the 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). Via del Mar: IEEE, 2020. 1–7. [doi: [10.1109/CIBCB48159.2020.9277638](https://doi.org/10.1109/CIBCB48159.2020.9277638)]
- 18 Shao ZF, Zhou WX, Deng XQ, *et al.* Multilabel remote sensing image retrieval based on fully convolutional network. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13: 318–328. [doi: [10.1109/JSTARS.2019.2961634](https://doi.org/10.1109/JSTARS.2019.2961634)]
- 19 Demir I, Koperski K, Lindenbaum D, *et al.* DeepGlobe 2018: A challenge to parse the earth through satellite images. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Salt Lake City: IEEE. 2018. 172–181.

(校对责编: 孙君艳)