



计算机科学

COMPUTER SCIENCE

基于MLUM-Net的高分遥感影像土地利用多分类方法

胡绍凯, 赫晓慧, 田智慧

引用本文

胡绍凯, 赫晓慧, 田智慧. 基于MLUM-Net的高分遥感影像土地利用多分类方法[J]. 计算机科学, 2023, 50(5): 161-169.

HU Shaokai, HE Xiaohui, TIAN Zhihui. [Land Use Multi-classification Method of High Resolution Remote Sensing Images Based on MLUM-Net](#) [J]. Computer Science, 2023, 50(5): 161-169.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于多级多尺度特征提取的CNN-BiLSTM模型的中文情感分析](#)

Chinese Sentiment Analysis Based on CNN-BiLSTM Model of Multi-level and Multi-scale Feature Extraction

计算机科学, 2023, 50(5): 248-254. <https://doi.org/10.11896/jsjcx.220400069>

[基于多事件语义增强的情感分析](#)

Sentiment Analysis Based on Multi-event Semantic Enhancement

计算机科学, 2023, 50(5): 238-247. <https://doi.org/10.11896/jsjcx.220400256>

[基于情感知识的双通道图卷积网络的方面级情感分析](#)

Aspect-based Sentiment Analysis Based on Dual-channel Graph Convolutional Network with Sentiment Knowledge

计算机科学, 2023, 50(5): 230-237. <https://doi.org/10.11896/jsjcx.220300008>

[残差学习与循环注意力下的SSD目标检测算法](#)

SSD Object Detection Algorithm with Residual Learning and Cyclic Attention

计算机科学, 2023, 50(5): 170-176. <https://doi.org/10.11896/jsjcx.220400085>

[基于Swin Transformer和三维残差多层融合网络的高光谱图像分类](#)

Hyperspectral Image Classification Based on Swin Transformer and 3D Residual Multilayer Fusion Network

计算机科学, 2023, 50(5): 155-160. <https://doi.org/10.11896/jsjcx.220400035>

基于 MLUM-Net 的高分遥感影像土地利用多分类方法

胡绍凯¹ 赫晓慧² 田智慧²

1 郑州大学信息工程学院 郑州 450001

2 郑州大学地球科学与技术学院 郑州 450052

(18538256320@163.com)

摘要 针对高分辨率遥感影像土地利用多分类结果中地块结构不完整、边界质量差的问题,提出了基于 MLUM-Net 模型的遥感影像土地利用多分类方法。该方法利用多尺度空洞卷积和通道注意力机制构造 MDSPA 编码器,提高了网络多尺度特征提取能力与地块位置定位的准确性,并通过空间注意力机制自适应增强了多尺度特征表达;为消除上采样语义损失和减少分类结果噪声,设计了混合池化上采样优化模块,用于优化分类结果并消除网络分类误差;根据土地利用多分类数据集类别占比不均衡的特点和地块结构的相似性指数设计混合损失函数,消除数据类别占比产生的影响,提高地块结构完整性和精细化分类边界。在多个数据集上进行了实验验证,总体精度和 kappa 指标均有明显提高,其分类结果结构完整且边缘划分准确,在土地利用多分类领域具有较好的实用价值。

关键词: 遥感影像;土地利用分类;MLUM-Net;注意力机制;多尺度特征

中图法分类号 TP391.4

Land Use Multi-classification Method of High Resolution Remote Sensing Images Based on MLUM-Net

HU Shaokai¹, HE Xiaohui² and TIAN Zhihui²

1 School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China

2 School of Earth Science and Technology, Zhengzhou University, Zhengzhou 450052, China

Abstract Aiming at the problems of incomplete land plot structure and poor boundary quality in high-resolution remote sensing image land use multi-classification, a multi-classification method of remote sensing image land use based on MLUM-Net model is proposed. This method uses the multi-scale hole convolution and channel attention mechanism to construct the MDSPA encoder, which improves the network multi-scale feature extraction ability and the accuracy of the parcel's location and adaptively enhances the multi-scale feature expression through the spatial attention mechanism. To eliminate the semantic loss of upsampling and reduce the noise of classification results, a hybrid pooling upsampling optimization module is designed to optimize the classification results and eliminate the network classification errors. According to the characteristics of unbalanced classification ratio of multi-classification data set of land use and the similarity index of plot structure, this paper designs a mixed loss function to eliminate the influence of data category ratio. This function improves the structural integrity of the block and refines the classification boundary. Experimental verification has been carried out on multiple data sets, and the overall accuracy and kappa index have been significantly improved. The classification result has a complete structure and accurate edge division, which has good practical value in the land use multi-classification.

Keywords Remote sensing image, Land use classification, MLUM-UNet, Attention mechanism, Multi-scale features

1 引言

近年来,利用高分辨率遥感影像进行多类地物分类已被成功应用在城市规划^[1]、土地资源管理^[2]、环境评估等众多领域。但从高分影像中进行土地利用多分类仍存在以下问题:

1)多样性,多类地物类型包括农田、森林与水域等,具有多尺度的形状特征;2)光谱干扰,地物在复杂区域如水域和农田相隔的滩涂、森林与草地中间的灌木区域常表现为与周围相似的光谱度,导致模棱两可的边缘分类结果;3)地物干扰,地表场景复杂区域多类地物聚集,易造成遮挡与覆盖,导致地块

到稿日期:2022-03-10 返修日期:2022-04-29

基金项目:河南省重大科技专项(201400210900);第二次青藏高原综合科学考察研究项目(2019QZKK0106)

This work was supported by the Key Technologies for the Construction and Service in Henan Province(201400210900) and Second Tibetan Plateau Scientific Expedition and Research(STEP) Program(2019QZKK0106).

通信作者:赫晓慧(hexh@zzu.edu.cn)

结构不完整并产生噪点。因此,自动、准确地进行高分影像多类土地利用分类仍具有一定挑战。

深度学习因其高效的特征提取能力,被广泛应用于目标检测^[3]、图像检索、医学图像分割^[4]等领域,并取得了丰硕的研究成果。国内外研究人员将深度学习方法应用到土地利用多分类中,通过卷积神经网络提取健壮、高级、深层的抽象特征,以进行精确的多类地物分类。Marcos 等^[5]通过旋转卷积核的方式构建卷积神经网络 RotEqNet,并以编码向量场的形式传递整个网络的特征信息流,提升网络对不同尺度地块的敏感性以加强网络特征提取的能力。Memon 等^[6]以迁移学习思想,学习低分辨率与高分辨率遥感影像的衍生特征,提高地物特征的稳健性,并探究了卷积网络层数和卷积核的数量对遥感影像多类地物分类的影响。Xia 等^[7]采用条件随机场(CRF)模型和卷积网络相结合的方式地进行地物分类,并在连通区域加入区域限制以增强分类结果的一致性,得到了较为优秀的分类结果。

研究人员对深度学习不断进行探索,编码器-解码器网络(U-Net, Link-Net)和特征金字塔解析网络(PSPNet)等多种结构更为复杂、特征提取能力更强的深度卷积神经网络(Deep Convolutional Neural Network, DCNN)被用于遥感影像多类地物分类中。例如, Yao 等^[8]采用特征通道切分的方法进行卷积,构建了一种更关注空间特征的特征提取模块 Coord-Conv,用于改善空间信息,并基于此模块提出了 DCCN 网络,用于多类地物分类,以提高高分影像多分类结果的准确性; Cai 等^[9]使用快速特征嵌入卷积结构,结合 PSPNet 优秀的上下文场景解析能力,以湖北省不同年份的遥感数据构建数据集进行实验并得到优异的分类结果; Meng 等^[10]通过不同尺寸卷积核构建深度神经网络,并进行遥感影像场景分类,有效提升了分类精度并探究了不同尺寸的卷积核对多类地物分类结果的影响。DCNN 凭借更深的网络结构,虽能捕获更多的深层抽象特征以表达不同地块,但这些方法通过堆叠卷积与池化层的操作对无规则地块的提取效果较差,且对于复杂区域具有较低的光谱抗干扰性,因此很难在高聚集复杂的场景下进行多类地物分类。其次,由于高分影像中同一类别覆盖地块常具有不同尺度、光谱和纹理信息,DCNN 网络难以使用单一尺度的特征表示复杂地块。针对此问题,研究人员开始利用尺度语义特征表达地表复杂地块,并通过构建更大感受野的土地多分类网络,来提升复杂地块地物多分类的准确性。Zhang 等^[11]以 U-Net 为基础网络引入空洞卷积,在不增加网络参数的同时增大网络感受野,以更好地提取不同尺度的地块特征,提升网络地物多分类性能。Yang 等^[12]将多尺度空洞卷积模块 ASPP 和 Dense-Net^[13]网络相结合,提出了能提取更丰富多尺度特征的 DenseAPP 网络,更好地提取了不同尺度的多类地块。Aihichre 等^[14]通过预训练的网络提取地块基础特征,然后使用注意力加权机制抑制与分类无关区域的语义表达,提高网络对正确区域的关注度,最后将加权后的特征逐层添加到第二个网络,以提高网络对多尺度特征的提取能力,进而得到准确的分类结果。上述方法虽在多类地物分类中充分考虑了单一地块多尺度的特点,但仍不能解决复杂场景下光谱干扰造成的信息损失影像最终边缘划分

问题。同时,现有网络大量使用交叉熵作为损失函数,在多类地物中存在地块结构缺失、边缘划分模糊的分类结果,影响网络最终分类性能。

为解决上述问题,并进一步提高遥感影像土地利用多分类的准确度,本文提出了 MLUM-Net(Multi-scale Land Use Multi-classification Network)土地利用分类方法,首先利用多尺度空洞卷积与混合空间金字塔注意力机制设计 MDSPA 模块(Multi-scale Dilated Hybrid Spatial Pyramid Attention, MDSPA),将其作为编码器构建网络下采样过程,提高网络多尺度特征提取能力与地块位置定位的准确性,通过空间注意力机制能够自适应增强特征的通道相关性,抑制冗余特征表达,通过多级特征的融合聚合多尺度上下文信息,保留丰富的多类地块细节特征;其次,为避免上采样造成语义损失并改善信息流,提出了 HPP 优化模块(Hybrid Pooling Pyramid Upsampling),通过多尺度池化获取丰富的全局上下文信息,加强多尺度特征表达并消除网络分类误差,提高分类结果的准确性;最后,为解决土地利用数据集中常存在类别占比不均衡的问题,结合多类地块结构多样性的特点,设计了混合损失函数,提高了网络对噪声及区域级别地块的错误分类关注,进一步细化了分类结果。为验证本文方法的有效性,设计了多种实验,以验证网络模型与损失函数的性能。

2 MULM-Net 方法

2.1 MULM-Net 土地利用分类网络

为了提高高分影像中多类地物分类结构的完整性和边界划分的准确性,本文构建了 MLUM-Net 网络结构,如图 1 所示。该网络主要包含 MDSPA 下采样模块和 HPP 上采样优化方法。MLUM-Net 基于 MDSPA 模块构建编码器输出初始的多类地物分类结果,其结果与各层级 MDSPA 模块提取多尺度特征的结果相结合并将其作为 HPP 上采样的输入,进一步结合全局与局部上下文信息,细化分类结果并提高最终分类结果的准确性。

网络分为特征提取和分类结果优化两个阶段,特征提取网络基于 U-Net 网络^[15]进行改进。首先使用两个 3×3 的卷积对原始输入数据进行浅层的特征处理,得到输入数据的浅层特征信息。网络的下采样过程采用 5 组 MDSPA 模块构建编码器,每个 MDSPA 模块对其输入的特征图像进行通道的切分操作,并在每个分支上使用扩张率为 $r=1, 2, 3$ 的空洞卷积替代普通卷积进行特征提取,通过多个堆叠的 MDSPA 模块,对所有特征信息建立通道方向上的长依赖关系,提高网络对地物位置定位的准确性。高分影像中相邻地块常具有相似的地物轮廓特征,为提高网络对轮廓划分及位置定位的准确度,在多尺度空洞卷积后使用混合空间金字塔注意力同步进行通道方向和水平空间维度的特征处理,使网络自适应平衡水平和纵向权重并进行特征细化,以提高多类地物分类的准确性。由于 MDSPA 模块通道切分操作需要通道数保持为 4 的倍数,MDSPA 模块最后均使用 1×1 卷积控制输出特征维度。输入预测网络的图像尺寸为 $256 \times 256 \times 4$,经过两层浅层卷积操作和 5 组 MDSPA 模块后的特征图尺寸分别为 $256 \times 256 \times 16, 256 \times 256 \times 32, 128 \times 128 \times 64, 64 \times 64 \times 128, 32 \times$

$32 \times 512, 16 \times 16 \times 512, 8 \times 8 \times 512$, 每个 MDSPA 输出特征进行前向传播的同时, 还将特征结果进行侧向输出以作为优化网络的输入。为提高网络的全局信息流, 使用跳跃连接将低层

高分辨率和高层低分辨率的特征相连, 聚合全局和局部上下文信息。网络下采样使用 5 组反卷积来还原特征信息, 最后一层上采样接 HPP 上采样模块进行优化分类。

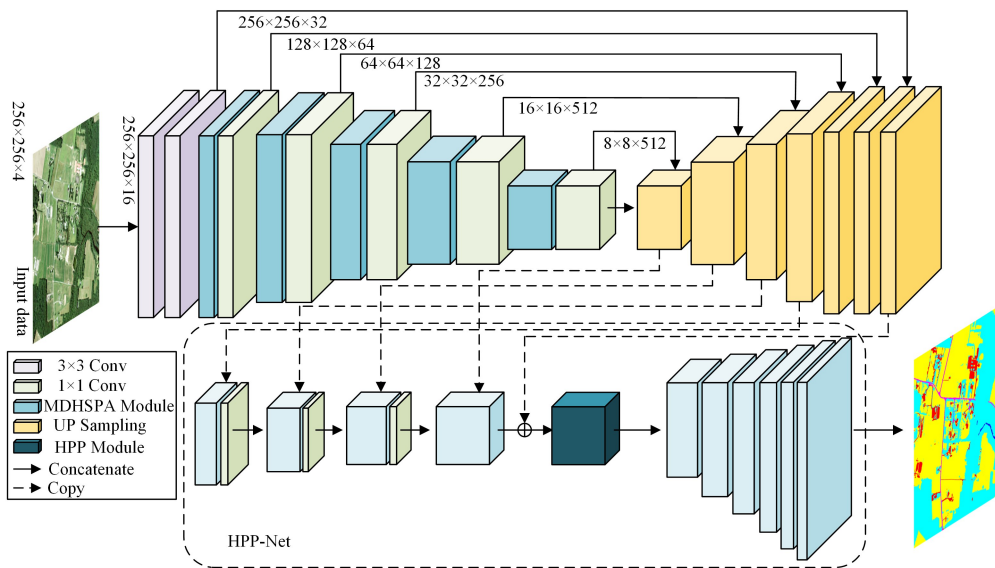


图 1 MHL-Net 网络的结构

Fig. 1 MHL-Net network structure

由于高分遥感影像存在边缘轮廓相似、同类地块尺度不规则和聚集区域光谱复杂等问题, 网络设计全局 HPP 上采样方法在特征提取网络的基础上进行分类结果的特征细化。首先将特征提取网络输出特征与下采样各层级高级抽象特征相结合, 使网络加强局部多尺度特征与全局高级抽象特征的非线性映射。其次利用多尺度池化提取全局池化特征, 提高网络多尺度特征的表达能。最终通过反卷积进行特征还原, 以更好地适用于聚集区域多类地物分类的目标, 提高地物多分类的分类精确度。

2.2 多尺度空洞混合空间金字塔注意力模块

MDSPA 模块由多尺度空洞卷积和混合空间金字塔注意力两部分组成, MDSPA 模块的结构如图 2 所示。

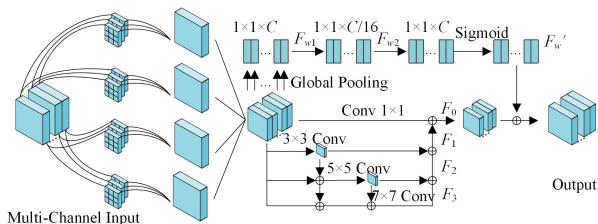


图 2 多尺度空洞混合空间金字塔注意力模块

Fig. 2 Multi-scale dilated hybrid spatial pyramid attention module

多尺度空洞卷积使用多种尺度的空洞卷积来扩大网络感受野并加强多尺度特征的提取能力。相比普通卷积, 空洞卷积使用扩张率控制卷积核的扩张间隔, 可以在不增加计算量的前提下扩大网络感受野。空洞卷积实际感受野的计算式如式(1)所示:

$$K = k + (k-1)(r-1) \quad (1)$$

其中, K 为实际感受野的大小, k 为原始卷积核的大小, r 为卷积扩张率。为了有效地从通道特征图中提取多尺度的空间信息, 该模块将输入特征以通道切分的方式分为 4 个支路同步

进行特征处理。每个支路依次使用扩张率 $r=1, 2, 3$ 的空洞卷积进行串联。其次为获得更多局部上下文信息, 每条支路均利用局部短连接的方式, 将空洞卷积串联。支路间以并行的方式独立学习多尺度空间信息, 最终将 4 个支路输出特征进行融合, 得到通道方向上的多尺度特征。支路 F_i' 经过 3 次空洞卷积输出特征可表示为:

$$F_i' = D_{K, r_i} \left(\sum_{i,j} Q[i, j] \times y[y_{i-1}, y_{j-1}, \dots, y_0] \right) \quad (2)$$

其中, D_{K, r_i} 表示空洞卷积操作, K 表示空洞卷积核大小, r_i 为扩张率, $Q[i, j]$ 为当前空洞卷积核, $y[y_{i-1}, y_{j-1}, \dots, y_0]$ 表示当前空洞卷积的输入, 也表示之前所有进行跨连接输出的特征结果。当前空洞卷积核的感受野的计算式为:

$$k' = k'_{i-1} + Q_{i-1} \times \prod_{i=1}^{i-1} S \quad (3)$$

其中, k'_{i-1} 为上一步空洞卷积的感受野, Q_{i-1} 为上一步空洞卷积的卷积核大小, S 为步长。最后多尺度空洞卷积模块将多支路上多尺度通道特征进行融合, 表示为:

$$F = \text{Concat}(F'_{i-1}, F'_{i-2}, \dots, F'_0) \quad (4)$$

其中, Concat 代表将不同支路输出特征进行特征融合, F_i' 为不同支路输出的特征结果。该模块利用通道上的特征预处理, 既可以有效捕捉局部信息, 也能更精准地合并上下文多尺度信息, 建立特征的跨维度交互通道, 提高地块结构的定位准确性, 为下一步混合空间注意力模块提供更准确的特征信息。

混合空间金字塔注意力处理分为两个阶段。首先, 使用通道注意力提取由多尺度空洞卷积模块输出的全局上下文先验信息, 自适应增强特征的通道相关性。其次, 使用不同尺度的卷积支路同步处理多尺度空洞卷积模块输出特征, 既可避免连续空洞卷积的栅格化影响, 也可实现连续的多尺度信息流。最终将多尺度空间特征信息与通道信息进行融合, 建立起通道的长依赖性并为高级特征产生更准确的像素级关注, 多尺度空洞卷积如图 3 所示。

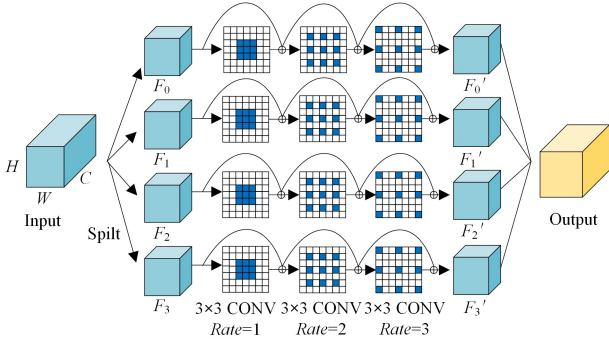


图3 多尺度空洞卷积模块

Fig. 3 Multi-scale dilated convolution module

多尺度空间特征的处理采用大小为 $3 \times 3, 5 \times 5, 7 \times 7$ 的卷积,以分支形式进行特征提取,并采用混合连接的方式将提取的不同尺度特征 F_1, F_2, F_3 进行特征融合,接着将多尺度空洞卷积输出特征 F 进行 1×1 卷积处理并与多尺度支路进一步融合成多尺度混合空间特征 F_0 , 1×1 的卷积操作是为了调整特征的通道并减少冗余特征,其表达式为:

$$F_0 = \text{Conv}_{1 \times 1} \oplus (F_1 \oplus F_2 \oplus F_3) \quad (5)$$

其中, $\text{Conv}_{1 \times 1}$ 表示对特征 F 的 1×1 卷积操作, \oplus 表示特征融合操作, F_1, F_2, F_3 表示多尺度卷积提取的支路特征。

通道注意力分支首先将多尺度空洞卷积提取特征 F 进行全局平均池化,得到 $1 \times 1 \times C$ 的特征结果,然后通过两个全连接层进行通道的挤压和激励,使网络自适应地权衡每个通道的重要程度,从而产生更加精准的位置定位特征。全局平均池化操作可以表示为:

$$F_c = \frac{1}{H \times W} \sum_{i=1, j=1}^{H, W} x_c(i, j) \quad (6)$$

其中, F_c 表示全局平均池化操作后的特征结果, H 和 W 分别表示输入特征的长和宽, Σx 表示逐像素的平均池化操作。

使用 Sigmoid 函数重新校准加权后通道方向的注意向量,通道注意力分支的输出特征可表示为:

$$F_{\omega'} = \sigma(F_{\omega_2} \partial(F_{\omega_1}(F))) \quad (7)$$

其中, $F_{\omega'}$ 表示通道注意力分支的输出特征, σ 表示 Sigmoid 函数, F_{ω_1} 和 F_{ω_2} 分别表示挤压和激励全连接层操作, ∂ 表示 Relu 激活函数, F 为多尺度空洞卷积输出特征。

最后,将两个支路的特征结果进行融合,得到多尺度特征信息更加丰富的细化特征,并将其作为该模块的输出特征。融合操作可表示为:

$$F' = F_0 + F_{\omega'} \quad (8)$$

其中, F_0 为多尺度混合空间特征, $F_{\omega'}$ 为通道注意力加权后的特征, F' 为 MDHSPA 模块最终输出的多尺度空洞混合空间特征结果。

2.3 混合池化金字塔上采样模块

卷积神经网络能够使用高度概括性的语义表达特征进行遥感影像分类,但使用卷积操作不可避免地会将真实影像中的特征信息放大和缩小,导致最终分类的预测结果边缘部分存在几何失真^[13]。同时,上采样过程高级抽象语义信息逐步被还原为预测对象的轮廓、结构、纵横比等特征表达,这会不可避免地造成语义损失和冗余信息干扰,分类结果存在结构

不完整和边界划分不准确的问题。为了进一步提高分类结果的准确性并消除网络误差,提出了混合池化金字塔上采样方法,该方法的核心设计 HPP 模块如图 4 所示。

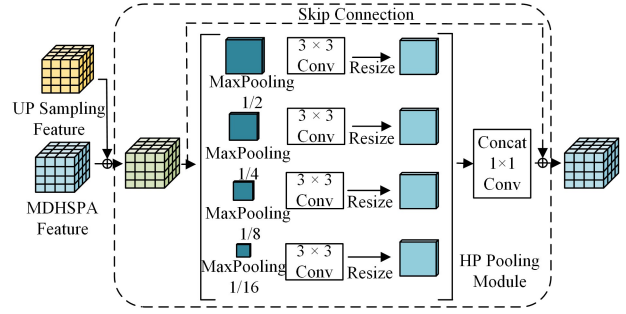


图4 混合池化金字塔上采样模块

Fig. 4 Hybrid pooling pyramid upsampling module

首先将特征提取网络上采样各层级特征的尺寸进行扩展并进行通道方向叠加,与最后一层上采样特征进行特征融合,生成复合特征并将其作为 HPP 上采样的输入。为避免上采样特征聚合过程造成的信息冗余,通道叠加过程使用一组权重向量来自适应控制上采样特征,权重参数参与到网络训练中以自适应学习各层级权重值。其次,使用 4 种不同尺度的全局池化支路提取池化特征信息,每一支路池化操作的卷积核大小分别为 $2 \times 2, 4 \times 4, 8 \times 8, 16 \times 16$,支路中采用 1 个 3×3 卷积来增加非线性映射并后接 BN 和 ReLU 激活函数,最后将 4 条支路特征进行尺寸调整并融合。为保持全局特征权重分配,使用 1×1 卷积调整通道数量后进行特征输出。

2.4 损失函数

遥感影像土地多分类数据集中常存在待分类样本数据类别分布不平衡的问题,且网络上采样易造语义损失,从而影响最终分类结果的准确性。因此,根据两个网络的特点与土地多分类数据存在的问题,本文结合多分类 Focal loss 函数^[16]和 SSIM 损失函数^[17]构建 MLUM-Net 的复合损失函数 L_{MLUM} 。Focal loss 损失函数用于提升多类地物提取的完整性与边缘分类的准确性,SSIM 损失函数用于进一步优化分类结果。

多分类 Focal loss 函数是基于交叉熵损失函数的改进版本。交叉熵损失函数的计算式如式(9)所示:

$$L_{\text{Cross-Entropy}} = -\frac{1}{n} \sum_i (y_i \log a_i + (1 - y_i) \log(1 - a_i)) \quad (9)$$

其中, y_i 为输入样本 x_i 的真实类别概率, a_i 为预测输入样本属于 x_i 的概率, a_i 的取值范围为 $(0, 1)$, a_i 越趋近于 1, 正样本的概率就越大。伴随着训练过程,反向传播不断修正权重参数值,使 $Loss$ 值趋近 0。但交叉熵损失函数每一次的梯度回传对每一个类别都具有相同的关注度,因此极易受到类别不平衡的影响。

多分类 Focal loss 函数的计算式如式(10)所示:

$$L_{\text{Focal}} = -\frac{1}{n} \sum_i (\alpha y_i \log(1 - \alpha_i)^\gamma \log \alpha_i + (1 - \alpha)(1 - y_i) \alpha_i^\gamma \log(1 - \alpha_i)) \quad (10)$$

相比交叉熵函数, Focal loss 多了 α 和 γ 两个超参, α 是取值为 0 到 1 之间的平衡因子,主要对数据量大的单类别样本

概率起到抑制作用, r 参数起到了平滑损失函数的作用,对易训练样本进行概率惩罚,使网络更关注难训练样本。因此,在训练过程中调节 α 和 γ 的值,可调节网络对不同类别的关注度,找到低数据量、难训练样本和高数据量、易训练样本的平衡点,最终得到准确的多分类结果。

图像质量 SSIM 是一种衡量重建图像与原图相似性的性能指标,将其引入损失函数中评估土地分类结果的准确性。较高的 SSIM 有着更为准确的地块分类结果,较低的 SSIM 可加强网络对地块尺度、纹理等信息的学习并消除噪点。由于 SSIM 损失函数使用均值和方差计算误差,但在单幅图中遥感土地利用影像的均值和方差变化剧烈,因此利用滑动窗口将图像分割为 N 幅小图片进行平均损失的计算, $x = \{x_i; 1, 2^2, \dots, N^2\}$ 与 $y = \{y_i; 1, 2^2, \dots, N^2\}$ 分别表示土地多分类预测概率图和标签图对应块的像素值,SSIM loss 函数的计算式为:

$$L_{SSIM} = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1) + (\sigma_x^2 + \sigma_y^2 + C_2)} \quad (11)$$

其中, μ_x 和 μ_y 分别表示 x 与 y 的平均值, σ_{xy} 表示 x 与 y 的平均值的协方差, C_1 与 C_2 为避免分母为零的常数, $C_1 = 0.01^2$, $C_2 = 0.03^2$ 。

在训练开始时,SSIM 损失函数对多类地块边缘区域分配较高权重,有助于多类地块的边界划分。随着网络训练地块边界的 SSIM 损失减少,地块完整性结构损失为主导,网络整体梯度偏向学习地块完整性,直到预测地块与真实地块非常相似,损失值迅速下降为 0。

L_{MLUM} 复合损失函数结合 Focal loss 与 SSIM loss 的特点并利用两个加权参数 λ_1 与 λ_2 优化两者的损失权重。设计的 L_{MLUM} 网络的损失函数的计算式为:

$$L_{MLUM} = \lambda_1 L_{Focal} + \lambda_2 L_{SSIM} \quad (12)$$

为避免因超参数过多而增加网络训练的难度,对式(13)进行公式转换,计算式为:

$$L_{MLUM} = \lambda_1 \left(L_{Focal} + \frac{\lambda_2}{\lambda_1} L_{SSIM} \right) \quad (13)$$

转换后的式(10)中 λ_1 为正参数,其大小不会影响损失 $\left(L_{Focal} + \frac{\lambda_2}{\lambda_1} L_{SSIM} \right)$ 的最小化,以 $\lambda = \frac{\lambda_2}{\lambda_1}$ 替代参数 λ_1 与 λ_2 起到减少超参数个数的目的,降低网络训练的复杂度。 $\lambda = 1$ 代表两个损失函数对网络的影响一样, λ 越接近于 0, L_{Focal} 起到的作用就越大, λ 越大大于 1, L_{SSIM} 对网络的影响效果就越大。

3 实验

为了验证高分遥感影像多类地物分类 MLUM-Net 方法的有效性,使用高分遥感影像数据对网络进行训练、测试与验证。实验与测试均基于 TensorFlow 深度学习框架进行设计并实现,运行环境中 CPU 为 HYGON C86 7465 24-CORE Processor, GPU 为 ASPEED Graphics Family 16GB×2。

3.1 实验数据

本文实验数据选用 Chesapeake Land Cover^[18], GID (Gaofen Image Dataset, GID)^[19] 多类地物高光谱数据集进行实验。其中,Chesapeake Land Cover 数据集包含 607 张原始数据,分辨率均为 6000×7000 左右,影像空间分辨率为 1m,

共包含 7 类土地利用遥感场景。原始数据共包含 29 个通道,其中有 RGB-NIR 图像数据、低分辨率 Landsat 8 的多光谱图像、来自 USGS NLCD 2011 数据集的低分辨率土地覆盖标签、Microsoft Bing 的高分辨率建筑足迹掩码等。

GID 数据集是由武汉大学基于高分 2 号卫星遥感数据于 2019 年发布的一个用于土地覆盖和土地利用分类的大型数据集,它包含来自中国 60 多个不同城市的 150 幅高质量高分二号图像,空间分辨率为 4m。数据集内部具有较高类内多样性和较低的类间可分离性,可作为遥感影像土地利用分类的基准数据集,共包含 6 类遥感土地场景,影像的分辨率均为 6800×7200。

为了保证实验公平性并对 MLUM-Net 分类方法做出客观评价,将所有训练、测试、验证数据切割为 1280×1280 的大小,人工筛选去除数据集中误差较大的影像和标签,并按照 70%, 20%, 10% 的比率设置训练集、验证集、测试集。此外,为避免类别与类别之间差异过大,对训练集中数据量较少类别的影像进行数据增强,主要为水平、纵向方向的翻转和旋转,将该部分数据的数据量扩增至原来 4 倍。

3.2 评价指标

本文采用评估遥感影像多类地物分类方法质量的常用评价指标,即总体精度 (Overall Accuracy, OA)、Kappa 系数作为总体的评价指标,制图精度 (Producer Accuracy, PA) 和用户精度 (User Accuracy, UA) 为单类别的评价指标。其计算式如式(14)~式(17)所示:

$$PA = \frac{X_i}{Y_{ii}} \times 100\% \quad (14)$$

$$UA = \frac{X_i}{X_{ii}} \times 100\% \quad (15)$$

$$OA = \frac{1}{N} \sum_{i=1}^k X_i \times 100\% \quad (16)$$

$$Kappa = \frac{N \sum_{i=1}^k X_i - \sum_{i=1}^k (X_{ii} Y_{ii})}{N^2 - \sum_{i=1}^k (X_{ii} Y_{ii})} \quad (17)$$

其中, N 为像素总数, n 为预测样本类个数, X_i 指某一类别 i 被正确分类的数值, X_{ii} 为预测数据中该类别的像素总数, Y_{ii} 为标签中该类别的像素总数。

3.3 实验结果与分析

为验证本文提出的 MLUM-Net 网络在遥感影像土地利用多分类任务中的有效性,本文进行了多项实验与测试工作。首先为验证本文所提混合损失函数的有效性,在 Chesapeake 数据集上进行了不同损失函数的对比实验,并进行了实验参数的对比测试与详细分析。随后为验证网络中各模块的有效性与对整体方法的影响,进行了详细的消融实验。其次,选取 Link-Net^[20], Dense-Net, DenseASPP^[21] 和 DeepLabV3^[22] 网络作为对比网络,分别在 Chesapeake 数据集与 GID 数据集上进行详细的网络对比测试。最后为探讨本文方法精度提升与算法效率的关系,在 Chesapeake 数据测试集上进行了检测用时测试工作,并与对比方法进行了对比测试和详细分析。

3.3.1 不同损失函数的实验对比与分析

在 MLUM-Net 网络结构的基础上,分别利用 L_{Focal} 损失函数、 L_{SSIM} 和混合损失函数 L_{MLUM} 对网络进行训练,测试结果如表 1 所列。

表 1 不同损失函数分类结果对比

Table 1 Comparison of classification results of different loss functions

Loss	OA/%	Kappa
L_{Focal}	79.79	0.803
L_{SSIM}	77.27	0.789
$L_{MLUM, \lambda=0.5}$	83.59	0.817
$L_{MLUM, \lambda=0.8}$	84.17	0.824
$L_{MLUM, \lambda=1.0}$	83.94	0.821
$L_{MLUM, \lambda=1.2}$	83.42	0.812
$L_{MLUM, \lambda=1.5}$	83.21	0.810

由表 1 可知, L_{Focal} 倾向于关注地块光谱、纹理和纵横比等易提取特征, 多类地物光谱度相似的复杂区域得不到有效的边界划分, 易造成分类错误且产生更多噪点, 最终的总体精度仅为 79.79%。 L_{MLUM} 损失函数通过图像局部区域结构的相似性, 对边界区域施加更大损失, 优化复杂区域的边缘及地块中的亮度与对比度, 同时也存在过度损失的现象, 影响正确分类, 总体精度为 77.27%。损失函数 L_{MLUM} 结合了两者的优点, 且使用超参数控制两者对网络的影响程度, 能明显减少地块的结构缺失, 同时优化复杂区域的错误划分, 整体分类结果更加准确。混合损失函数 L_{MLUM} 与使用单个 L_{Focal} 和 L_{SSIM} 损失相比, 当 $\lambda=1$ 时, OA 和 Kappa 分别提升了 4.15%, 0.018 和 6.67%, 0.032。经过参数调整, 当 $\lambda=0.8$ 时, L_{MLUM} 损失函数在 Chesapeake 达到最优, OA 和 Kappa 分别为 84.17%, 0.824。

3.3.2 消融实验

为进一步验证本文所提 MDSPA 模块与 HPP 上采样优化模块对 MLUM-Net 方法的有效性, 设置模块对比消融实验。实验分别使用 MDSPA 模块与 HPP 模块对网络进行训练, 测试结果如表 2 所列。

表 2 不同模块之间的消融实验

Table 2 Ablation experiment between different modules

Model	OA/%	Kappa
U-Net	75.62	0.763
U-Net+ MDSPA	79.52	0.801
U-Net+ HPP	77.37	0.761
MLUM-Net	84.17	0.824

表 3 Chesapeake Land Cover 数据集的测试结果

Table 3 Test results on Chesapeake Land Cover dataset

Method	Water (PA/UA)	Building (PA/UA)	Road (PA/UA)	Farmland (PA/UA)	Forest (PA/UA)	Meadow (PA/UA)	OA/%	Kappa
Link-Net	76.97/78.26	73.75/76.24	77.86/80.11	73.46/75.81	73.28/76.77	80.39/81.21	78.06	0.711
Dense Net	77.46/76.89	76.32/78.73	72.46/79.61	75.92/77.64	72.83/80.24	79.83/78.62	78.62	0.734
DenseASPP	86.81/83.24	79.25/80.34	78.65/80.34	77.37/78.72	75.68/78.54	81.67/80.41	80.26	0.764
DeepLabV3	83.67/ 84.94	80.65/ 80.94	80.37/82.56	79.36/77.34	80.61/ 83.97	82.34/80.74	81.63	0.782
MLUM-Net	85.27/83.06	81.92/78.64	81.54/83.84	81.26/80.79	82.35/82.67	83.14/83.64	82.11	0.816

由表 3 中的总体精度和 Kappa 系数两项精度测试结果可以得出, Dense Net 网络的结果虽然可用于遥感影像多类地物目标的分类, 但测试结果较差, 总体精度和 kappa 分别为 78.62% 和 0.734。建筑物、道路、森林、农田这种无规则地块的准确率较低, 这是由于 Dense Net 虽采用稠密残差块提取不同地块的光谱、纹理、尺度等特征, 但网络缺乏对目标特征的细节处理, 上采样过程会丢失大量细节信息, 从而造成地块结构缺失。Link-Net 虽在 U-Net 网络的基础上设计了更加优秀的全局上下文编码器来提取全局信息, 且使用关系嵌入模块区分不同特征, 但网络缺乏尺度轴上的特征提取能力, 地

由表 2 可知, 本文所提 MLUM-Net 方法中的 MDSPA 模块和 HPP 模块均能提高模型的性能, 且将两个模块与基础网络进行融合后的多类地物分类网络的分类结果明显优于基础网络 U-Net, 其中 OA 分别提高了 3.90% 和 1.75%。图 5 给出了本文模块与基础网络分别融合后, 在 Chesapeake 高分遥感影像多分类数据集上的分类性能对比结果图。由图可知, U-Net 缺乏多类地物特征提取能力以及对复杂区域细节部分进行处理的能力, 分类结果较差。加入 MDSPA 模块能有效扩大网络感受野, 提高对不同类别地块提取的能力且提取的地块结构较为完整, 一些离散分布的建筑物和不规则水域也能较为准确地与周围场景进行区分。添加 HPP 优化模块后能有效增强多尺度特征表达, 地块整体结构更准确且能提高地块之间边缘划分的准确性, 减少边缘特征信息的损失。

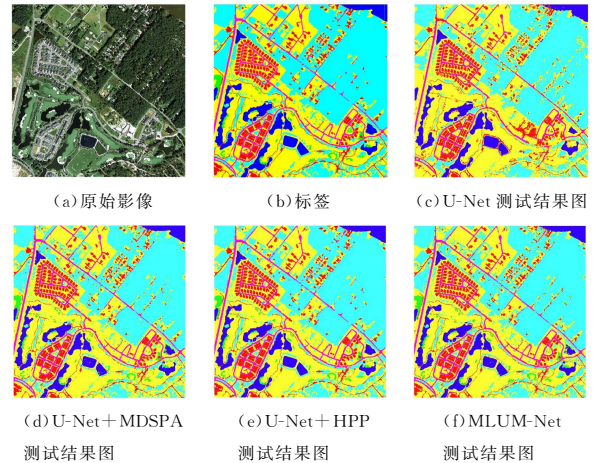


图 5 消融实验测试结果图

Fig. 5 Test results of ablation experiments

3.3.3 Chesapeake 数据集的实验结果与分析

本文的第一组整体网络对比实验在 Chesapeake Land Cover 数据集上进行, 该数据集共包含 7 种土地覆盖类别: 水域、草地、林地、农用耕地、建筑物群、道路和背景, 测试结果如表 3 所列。

位置定位准确率不高且易产生大量噪声, 影响整体结果, 总体精度只有 78.06%, 可见其模型结构并不能完全适用于复杂的遥感影像土地利用分类任务。从表 3 中的 DenseASPP 与 DeepLabV3 可以得出, 总体精度和 kappa 指标比 Dense Net 与 Link-Net 均有显著提高, 特别是 DeepLabV3 的总体精度较 Dense Net 提高了 3.01%, 表明通过串联和稠密连接的空洞卷积提取不同感受野的地块特征对多类地物分类是有效的, 同时 DenseASPP 还设计了稠密连接, 将多扩张率的空洞卷积进行连接, 能生成覆盖范围更大的接收野特征, 有利于不规则地块完整结构的提取, 农田、森林等具有明显无规则特点

的单类别地块相比 Dense Net 在制图精度上分别提高了 5.9%,7.33%。但 DenseASPP 与 Dense Net 对地块边缘的划分精确度不高,尤其是草地与森林、建筑物与道路等光谱与纹理相似的区域划分效果不佳。而本文提出的 MLUM-Net 方法利用通道多尺度切分方法,提高了网络对多类地块定位的准确性,并采用多尺度混合空间注意力模块提高了网络对高分影像中不规则地块的敏感性,且使用通道注意力机制增强网络对多类别分类特征的细节区分,在特征细节处理

上明显优于其他方法。网络最后使用 HPP 上采样优化方法优化地块结构并消除噪声,得到更加准确的分类结果,总体精度分别较 Link-Net,Dense Net,DenseASPP 与 DeepLabV3 提高了 4.05%,3.49%,1.85%,0.48%,离散分布的建筑物相比 DenseASPP 与 DeepLabV3 在制图精度上进一步提高了 2.67%,1.27%。各方法的多类地物分类结果如图 6 所示,其中图 6(a)~图 6(g)分别对应高光谱影像、标签、Dense Net、Link-Net、DenseASPP、DeepLabV3 的多类地物分类结果。

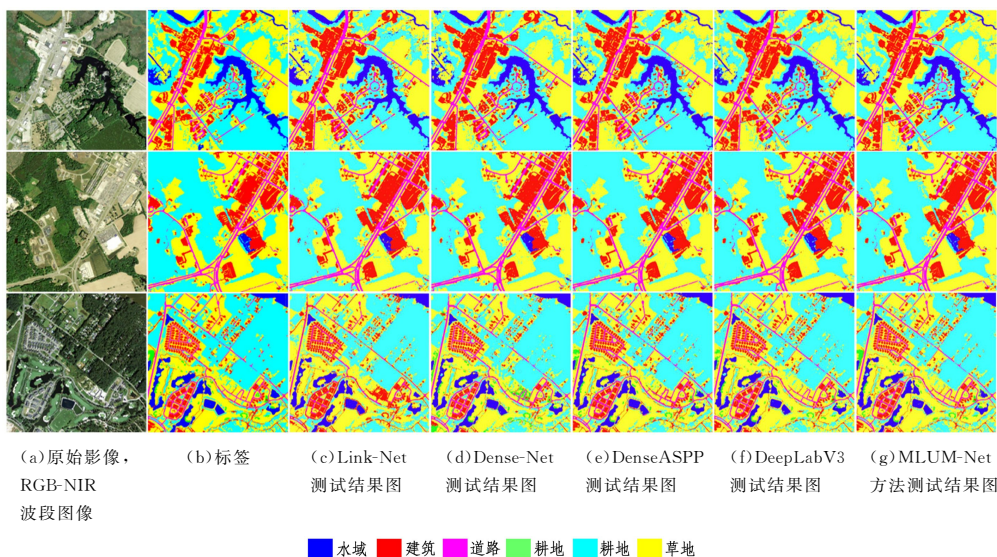


图 6 Chesapeake Land Cover 数据集测试结果图

Fig. 6 Test results on Chesapeake Land Cover dataset

由图 6 可知,Dense Net 方法受限于网络设计,缺乏对不同地块的特征细节区分,导致地块缺失比较严重,且地块周围存在较多噪声,无法有效地对复杂的多类地块的边缘进行区分。Link-Net 网络的分类结果虽噪声较少,但地块结构并不能完整提取与划分,出现了大量的模棱两可的分类结果。DenseASPP 与 DeepLabV3 方法的分类结果虽然整体较为准确,但在复杂场景下仍存在一定的地块结构缺失问题,从整体来看,多类地块边缘划分不够准确,对待复杂区域边缘部分光谱混淆的抗干扰性不及本文方法。本文提出的

MLUM-Net 方法的表现是最优异的,能够较准确地划分各种土地场景,即使对农田和建筑物这种离散分布的地块也能较好地划分出边缘轮廓。从各项评价指标与分类结果整体性来看,本文方法在 Chesapeake 数据集上的分类效果最好。

3.3.4 GID 数据集上的实验结果与分析

本文进行的第二项网络对比实验采用了 GID 数据集,该数据集共包含 6 类土地覆盖地块,即水域、草地、林地、农用耕地、建筑群与背景,其测试结果如图 7 与表 4 所示。

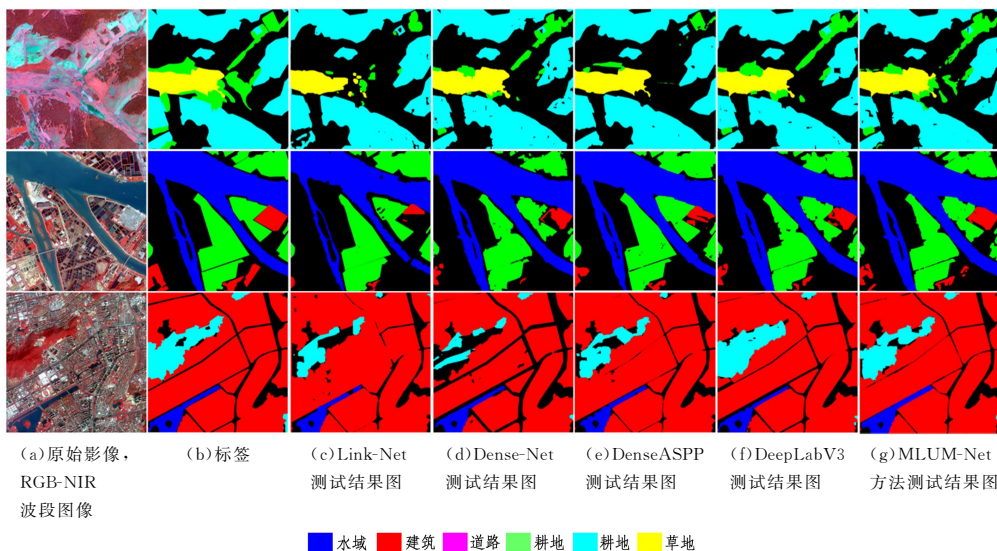


图 7 GID 数据集测试结果图

Fig. 7 Test results on GID dataset

表 4 GID 数据集测试结果
Table 4 Test results on GID dataset

Method	Water (PA/UA)	Building (PA/UA)	Farmland (PA/UA)	Forest (PA/UA)	Meadow (PA/UA)	OA/%	Kappa
Link-Net	74.27/78.19	76.17/76.45	74.29/76.13	75.42/75.58	78.83/79.27	77.12	0.711
Dense Net	77.05/79.76	75.67/81.53	76.17/78.35	77.34/74.68	80.67/83.56	78.62	0.726
DenseASPP	78.39/80.52	78.71/78.02	78.13/79.47	78.94/77.36	81.73/80.33	79.14	0.764
DeepLabV3	80.47/81.41	82.57/77.71	79.36/78.69	80.61/80.48	81.63/80.75	79.81	0.793
MLUM-Net	81.96/80.84	81.63/82.95	80.26/80.72	82.22/81.89	83.49/81.47	81.57	0.804

由图 7 可以看出,本文方法在地块整体结构完整性与地块边缘划分上的表现是最优秀的,尤其在复杂光谱场景下的林地地块与建筑物地块,都得到了较完整的地块提取结果,在测试结果中表现出了最少的地块缺失,如图 7 第一行林地与第二行建筑物分类结果。这得益于本文采用的多尺度空洞混合金字塔注意力模块,能够在多尺度特征提取的结果上与混合损失函数共同推动网络对缺失地块的学习与训练,并且消除了地块分类误差,带来了更加准确的地块分类结果。而 Link-Net 的地块缺失是最为严重的,并且受限于网络结构设计,在地块边缘还会出现模棱两可的分类结果。DeepLabV3 与 DenseASPP 网络均采用多尺度特征提取方法,地块缺失现象得到了较好的改善,分别取得了第二、第三的精度表现。

从表 4 所列的总体精度与 kappa 指标也能看出,本文方法在 GID 数据集上的性能表现是最优的,分别为 81.57%, 0.804; DeepLabV3 与 DenseASPP 次之,总体精度分别为 79.81%与 79.14%; Dense Net 与 Link-Net 表现最差,分别为 78.62%与 77.12%。从单类别各项精度来看,本文方法在离散分布的农用耕地、森林与草地相比对比网络提升明显,其中制图精度 PA 较第二名的 DeepLabV3 分别提升了 0.9%, 1.61%, 1.86%。但是在水域的分类结果中,本文方法与对比网络并没有表现出较大的性能差异,这是由于在该数据集中水体与周围地块具有较大的光谱差异。

3.3.5 算法效率实验与结果分析

为探讨本文方法在土地利用多分类任务中的算法效率,与对比网络进行了精度提升和测试用时的实验与分析。以对比网络中模型结构最简单的 Link-Net 模型为基础,从总体精度和整张测试图片的检测用时方面进行对比测试与结果分析,结果如表 5 所列。

表 5 算法效率对比测试结果

Table 5 Algorithm efficiency comparison test results

模型名称	OA/%	整张影像测试时间/s
Link-Net	78.06	18.164
Dense Net	78.62(+0.56%)	21.519(+18.5%)
DenseASPP	80.26(+2.2%)	23.681(+30.3%)
DeepLabV3	81.63(+3.57%)	22.369(+23.2%)
MLUM-Net	82.11(+4.05%)	24.286(+33.7%)

从表 5 中可以得出,Link-Net 相比其他模型时间消耗最少,但由于其网络结构的限制,其总体精度的表现也是最差的。DenseNet 的总体精度相比 Link-Net 有一定提升,但由于采用密集残差连接与多层特征融合策略,网络的检测速度相比 Link-Net 也有一定的提高。此外, DenseASPP 与 DeepLabV3 的精度分别较 Link-Net 提升了 2.2%, 3.57%, 虽然两个网络均采用 ASPP 模块进行特征提取,但 DenseASPP 与

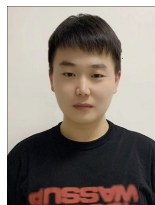
Dense Net 一样采用了过多的连接以进行特征叠加与融合计算,使整张图片的检测时长较 DeepLabV3 有一定的增加,较 Link-Net 模型增加了 30.3%。而本文方法在进行土地利用分类任务时,其整张影像检测时间较 Link-Net 增加了 33.7%,但总体精度也有 4.05%的提高。与 DenseASPP 和 DeepLabV3 相比, MLUM-Net 方法的测试时间分别增加了 0.605 s 和 1.917 s,总体精度分别提高了 1.85%和 0.48%,在控制影像检测时间损耗不过多增加的同时,提高了土地利用多分类任务的分类效果,有效改善了地块结构提取不完整和边缘划分不准确的问题。

结束语 本文提出了 MLUM-Net 土地利用多分类方法,该方法能较好地解决土地分类地块结构不完整、边缘划分精确度较低的问题。其中, MDSPA 模块通过多尺度空洞卷积和通道注意力机制来提取完整的多类地块区域,提高了地块位置定位的准确性,混合池化 HPP 模块在特征解码阶段利用多尺度特征,有效聚合全局特征优化分类结果。在此基础上,利用混合损失函数进一步细化提取特征,提升分类结果的精确度。实验结果表明, MLUM-Net 土地利用多分类方法的各项性能指标均优于其他方法,能够从复杂的高分辨率遥感影像中获得完整且精准的分类结果,在土地利用多分类任务中具有广泛的应用前景。但本文方法在一些地块分类结果的指标上没有达到最好的效果,且采用过多的特征融合与通道叠加计算提高了网络的时间复杂度。下一步工作可针对不同模块的特征融合算法进行改进,以提高网络效率,使其在土地利用分类领域发挥更好的作用。

参考文献

- [1] SU M, GUO R Z, CHEN B, et al. Sampling Strategy for Detailed Urban Land Use Classification: A Systematic Analysis in Shenzhen[J]. Remote Sensing, 2020, 12(9): 1497.
- [2] MEN J L, LIU Y Y, ZHANG B. High-scoring Image Land Use Classification Based on Feature Cascades of Multi-Structure Convolutional Neural Network[J]. Geomatics and Information Science of Wuhan University, 2019, 44(12): 1841-1848.
- [3] CHEN K Q, ZHU Z L, DENG X P, et al. Deep Learning for Multi-scale Object Detection: A Survey[J]. Journal of Software, 2021, 32(4): 1201-1227.
- [4] FENG F, LIU P X, LI L, et al. Study of FCM Fusing Improved Gravitational Search Algorithm in Medical Image Segmentation [J]. Computer Science, 2018, 45(S1): 252-254.
- [5] MARCOS D, VOLPI M, KELLENBERGER B, et al. Land cover mapping at very high resolution with rotation equivariant CNNs Towards small yet accurate models[J]. Isprs Journal of Photo-

- grammetry & Remote Sensing, 2018, 145PA(NOV.): 96-107.
- [6] MEMON N, PARIKH H, PATEL D, et al. Automatic Land Cover Classification of Multi-Resolution Dualpol Data using Convolutional Neural Network(CNN)[J]. Remote Sensing Applications Society and Environment, 2021(2): 100491.
- [7] XIA M, CAO G, WANG G Y, et al. Remote Sensing Image Classification Based on Deep Learning and Conditional Random Fields[J]. Journal of Image and Graphics, 2017, 22(9): 1289-1301.
- [8] YAO X, YANG H, WU Y, et al. Land Use Classification of the Deep Convolutional Neural Network Method Reducing the Loss of Spatial Features[J]. Sensors, 2019, 19(12): 2792-2807.
- [9] CHAI H B, YAN C, ZOU Y F, et al. Using PSP Net to Realize Land Cover Classification of Remote Sensing Imagery in Hubei Province[J]. Geomatics and Information Science of Wuhan University, 2021, 46(8): 1224-1232.
- [10] MENG Q X, DUAN X L. High-resolution Remote Sensing Image Scene Classification Based on DCNN[J]. Journal of Central China Normal University(Natural Sciences), 2019, 53(4): 568-574.
- [11] ZHANG W, TANG P, ZHAO L. Fast and accurate land cover classification on medium resolution remote sensing images using segmentation models[J]. International Journal of Remote Sensing, 2021, 42(9): 3277-3301.
- [12] YANG M, YU K, CHI Z, et al. DenseASPP for semantic segmentation in street scenes[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018: 3684-3692.
- [13] HUANG G, LIU Z, LAURENS V D M, et al. Densely connected convolutional networks[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). 2017: 2261-2269.
- [14] AIHICHRE H, ALSUWAYED A, BAZI Y, et al. Classification of Remote Sensing Images using EfficientNet-B3 CNN Model with Attention[J]. IEEE Access, 2021, 9(9): 14078-14094.
- [15] RONNERBERGER O, FISCHER P, BROX T. U-Net: Convolutional Networks for Biomedical Image Segmentation[J]. Medical Image Computing and Computer Assisted Intervention, 2015, 28(4): 234-241.
- [16] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal Loss for Dense Object Detection[C]// IEEE International Conference on Computer Vision(ICCV). 2017: 2999-3007.
- [17] WANG Z, BOVIK A C, SHEIKH H R, et al. Image Quality Assessment: From Error Visibility to Structural Similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.
- [18] ROBINSON C, HOU L, MALKIN K, et al. Large Scale High-Resolution Land Cover Mapping with Multi-Resolution Data [C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). IEEE, 2019.
- [19] TONG X Y, XIA G S, LU Q, et al. Land-cover classification with high-resolution remote sensing images using transferable deep models[J]. Remote Sensing of Environment, 2020, 237: 111322.
- [20] CHAURASIA A, CULURCIELLO E. Exploiting encoder representations for efficient semantic segmentation[C]// 2017 IEEE Visual Communications and Image Processing(VICIP). 2017: 1-4.
- [21] YANG M, YU K, ZHANG Z, et al. DenseASPP for Semantic Segmentation in Street Scenes[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018: 3684-3692.
- [22] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking Atrous Convolution for Semantic Image Segmentation [C] // Computer Vision and Pattern Recognition. 2017.



HU Shaokai, born in 1995, postgraduate. His main research interests include artificial intelligence, remote sensing image processing.



HE Xiaohui, born in 1978, Ph.D, professor. Her main research interests include artificial intelligence, computer vision, remote sensing image processing, and data mining.

(责任编辑:喻藜)