doi: 10.6046/zrzyyg.2022164

引用格式: 刘立,董先敏,刘娟. 顾及地学特征的遥感影像语义分割模型性能评价方法[J]. 自然资源遥感,2023,35(3):80-87. (Liu L, Dong X M, Liu J. A performance evaluation method for semantic segmentation models of remote sensing images considering surface features[J]. Remote Sensing for Natural Resources,2023,35(3):80-87.)

顾及地学特征的遥感影像语义分割模型性能评价方法

刘 立^{1,2}, 董先敏^{1,2}, 刘 娟¹

(1. 自然资源部第三地理信息制图院,成都 610100; 2. 自然资源部数

字制图与国土信息应用重点实验室,成都 610100)

摘要:深度语义分割目前已经被广泛应用于国土遥感监测和遥感解译生产领域,针对现有语义分割结果质量评价 方法无法反映语义分割结果在空间几何特征上保持情况的客观问题,文章从遥感解译与测绘生产的实际需求出 发,提出了一种顾及地学特征的遥感影像语义分割模型性能评价方法——连通相似性指数(connectivity similarity, CSIM),从遥感地物图斑连通相似性层面,将地学特征嵌入语义分割模型性能评价体系。该方法可以定量评估遥感 影像语义分割结果与实际样本标签的图斑连通相似性程度,准确描述预测分类结果中图斑完整性的保持情况,从 而更加客观地判断预训练模型是否适用于测绘生产的遥感解译工作。经过大量实践证实,该评价方法可以更好地 实时监测和控制模型训练,有效地指导从预训练模型集合中选取最优性能的模型,准确地评估顾及地物几何特征 的遥感影像预测结果的真实质量,对深度学习赋能遥感解译与测绘生产具有重要作用。

关键词:性能评价;语义分割;地学特征;遥感解译;深度学习

中图法分类号: TP 79 文献标志码: A 文章编号: 2097 - 034X(2023)03 - 0080 - 08

0 引言

近年来,深度学习因特征提取能力强和识别精 度高等优点已在计算机视觉和图像处理的语义分割 领域取得重大进展并得到广泛的应用^[1]。在国土 遥感监测和遥感解译生产领域,将深度学习规模应 用于地表覆盖分类、要素提取更新和地物变化分析 等生产实践,已经显著提高了遥感影像图载信息的 提取效率^[2]。深度学习在遥感图像中的应用不同 于自然图像,遥感图像通常具有更复杂多样的模式 和更丰富的时空光谱信息,因此对语义分割结果的 评价方法提出了更高的要求^[3]。

目前,混淆矩阵作为语义分割模型性能评价的 重要工具,作为高级统计指标的标准输入^[4],计算 准确率、精确率、召回率和 F1 Score 等各种评价指 标,从而绘制 ROC 曲线和 CAP 曲线等,更全面地评 价和比较语义分割模型的性能。国内外学者将形态 轮廓(morphological profiles, MPs)、微分形态剖面 (differential morphological profiles, DMPs)、马尔可夫 随机场(Markov random field, MRF)、局部二进制模 式(local biary pattern, LBP)、像素形状指数(pixel shape index. PSI)、对象相关指数(object correlative index,OCI)和形状大小指数(shape size index,SSI) 等方法用于影像分类精度的评价^[5]。Zhang 等^[6]提 出了一种基于光谱相似性的图像对象的属性 OCI, 来描述分类结果的空间特征:段嘉鑫^[7]提出了图像 分割结果的逐像素误差谱的概念,通过解码器预测 细粒度的分割结果误差谱,实现了分割结果的局部 质量信息获取,以及像素级的分割质量评价;党 宇^[8]以图幅为评价基本单元,对单一地类图斑评价 有效性及混合图斑的评价策略进行研究及实验验 证:郑明国等^[9]将包括目标语义信息和地学先验知 识的地学知识嵌入遥感影像深度语义分割引导网络 训练,以提高分割结果的可解释性和可靠性。从国 内外的研究现状来看,顾及地学特征的深度学习语 义分割模型性能评价的方法研究较少,性能评价维 度缺少空间关系特征,尚难以有效支撑时空异质的 遥感影像语义分割结果的评价,主要问题表现在2 个方面: ①评价对象是像素粒度,像素只关注最低

基金项目:四川省重点研发计划项目"面向国土空间规划的自然资源三维智能化管控关键技术研究与应用"(编号:2022YFS0450)和 自然资源科技创新发展项目"自然资源技术融合研究与应用示范"(编号:121204007000204101)共同资助。

第一作者:刘立(1989-),男,本科,工程师,主要从事自然资源调查监测与地理信息工程应用研究。Email: 274114486@ qq. com。

通信作者:董先敏(1994-),女,研究生,工程师,主要从事行业地理信息服务应用研究。Email: 1390470924@qq.com。

收稿日期: 2022-04-27;修订日期: 2022-07-13

层次的局部特征信息,无法满足高层次面向目标对 象的遥感地物解译需求;②评价准则是像素分类是 否正确,没有考虑到空间位置因素的影响,无法直接 反映预测结果在空间几何特征上的保持情况。

遥感影像智能解译是遥感应用落地的核心和关 键技术,而语义分割性能的准确评价是实现遥感影 像智能解译的基础^[10-13]。性能评价体系应满足像 素和对象等不同层级的需求,遵循位置精度与逻辑 完整性准则^[14-16]。本文从时空异质遥感解译和测 绘生产的角度出发,在分析现有模型评价体系的基 础上,提出一种顾及地学特征的遥感影像语义分割 模型性能评价方法——连通相似性指数(connectivity similarity, CSIM),在充分考虑像素归类是否正确 的基础上,以连通图斑(即像素群)为粒度,聚焦空 间位置和空间关系,以真实标签和预测标签的连通 序列相似程度来进行语义分割性能评价,以期为模 型训练的实时监测、预训练模型的最优选择和遥感 影像预测结果的质量评价提供可靠的技术支持。

1 模型性能评价问题分析

语义分割是深度学习领域一种典型的面向像素 级别的计算机视觉技术,其性能评价指标是建立在 表征图像预测分类结果的情形分析表即混淆矩阵的 基础上^[17-18]。混淆矩阵元素是预测分类结果的统 计值。在评价语义分割性能时,标记图像至关重要, 为性能提供了逼近的标准。表1为目前常用的语义 分割性能评价指标的内容与意义。评价指标将混淆 矩阵中数量的结果表征为0~1之间的比率,便于进 行标准化的衡量。然而在遥感影像应用于测绘生产 的实践过程中,预期解译结果需要由聚集的像素群 转换为保持地物空间特征的矢量化面状图斑,如聚 集成块的建筑群和保持带状连通的道路。

表 1 常用评价指标 Tab. 1 Common evaluation indicator

指标	意义
准确率	正确预测的样本占总样本的比例
精确率	"正确被预测为正"占所有"实际被预测为正"的比例
召回率	"正确被预测为正"占所有"应该被预测为正"的比例

以遥感图像中建筑物的语义分割为例,在图1 (a)中,共有8×8个格网,每个格网代表一个像素, 共64个像素,其中灰色像素格网表示目标遥感地物 边界,共32个像素。经过卷积神经网络训练后得到 具备识别该遥感地物能力的模型,对遥感影像进行 预测,呈现3类预测结果:图1(b)预测的遥感地物 轮廓保持较好的完整性,图1(c)遥感地物轮廓不完 整,但是地物的空间位置保持比较准确,图1(d)遥 感地物部分像素虽然预测正确,但完整性极差。3 种预测标签的准确率、精确率和召回率指标的计算 结果如表2所示。可以发现,虽然3种预测标签形 状不同,但同一评价指标的数值相同,无法定量地评 估预测标签的质量。



Fig. 1 Comparison chart between real label and predicted labels

表 2	预测标签评价指标结果	
-----	------------	--

Tab. 2	Evaluation	results	of	predicted	labels
--------	------------	---------	----	-----------	--------

		-	
指标	图 1(b)	图 1(c)	图 1(d)
准确率	0.78	0.78	0.78
精确率	1	1	1
百回率	0.56	0.56	0.56

从遥感解译后处理的角度来看,图1(b)经过孔 洞填充可以达到较为理想的结果,图1(c)经过小图 斑过滤可以较好地保留空间位置,而图1(d)结果几 乎没有实用价值,所以期望结果的顺序应该为:图1 (b)>图1(c)>图1(d)。综上,在遥感应用场景的 特殊需求下,常规性能评价方法很难满足实际要求, 已经无法有效地指导遥感影像预测结果的质量 评价。

2 CSIM 设计

本文的目的是针对语义分割模型预测遥感影像,探索出一种评价预测标签真实质量的方法,并非 是要替代现有的评价指标体系,所以本文在现有的 评价体系中增加图斑连通相似性评价,以期在现有 的评价指标无法区分的区间进一步提高对遥感解译 质量的辨识度,实现性能评价指标值与主观质量分 数保持一致。CSIM 计算数据输入为模型的预测标 签和真实标签,数据输出为连通相似性指数。其方 法流程包括连通图斑特征提取、连通序列空间校准、 连通序列相似度评价和连通相似性指数计算。图 2 显示了现有评价方法和 CSIM 指标值对比,随着训 练时期的推移现有评价指标难以体现出性能差异, 在训练末期,CSIM 指标差异性明显,可以有效地评 价语义分割性能。



Fig. 2 Comparison of model performance evaluation methods

2.1 连通图斑特征提取

本文定义真实标签与预测标签中水平或垂直方 向上像素值相同的相邻像素集合为连通图斑。连通 图斑识别采用四邻域两遍扫描法,由左向右、由上向 下对真实标签和预测标签进行逐行扫描搜索,将同 一连通区域的所有像素点标记同一连通索引值,扫 描完成输出一个标记所有连通图斑像素集合的连通 序列。图 3 为连通图斑示意图,图中①,②,③均为 连通图斑。



连通图斑的特征有2类表示方法:一类是区域 特征,一类是轮廓特征。其中区域特征采用区域内 像素总数量来定量描述;轮廓特征主要针对图斑的 内外边界。经过逐一对连通图斑的特征提取,最后 输出记录所有连通对象特征信息的真实标签连通序 列 A 和预测结果连通序列集合 B。其中真实标签连 通序列 A = $\{a_1, a_2, \dots, a_i, \dots, a_m\}$, a_i 表示第 i 个连 通图斑的轮廓信息和像素点总数量的二元组, 预测 结果连通序列集合 B = $\{B_1, B_2, \dots, B_k, \dots, B_y\}$, B_k = $\{b_1, b_2, \dots, b_h, \dots, b_l\}$ 表示第 k 个预测标签连通序 列。

2.2 连通序列空间校准

连通图斑的识别方式为逐行扫描,因此连通序 列的存储顺序是基于连通图斑的识别先后顺序,这 就导致真实标签的连通序列 A 和预测标签的连通 序列 B_k 中的顺序难以保持近似的对应关系。为了 保证连通序列 A 和 B_k 匹配最优,本文基于连通图斑 的轮廓特征进行空间相交运算确定连通序列 A 中 的每个连通图斑与预测序列 B_k 中连通图斑的空间 对应关系,将 B_k 中具备空间关系的连通图斑按照 A 的索引顺序重新排列,将 B_k 中没有对应空间关系的 连通图斑作为冗余删除,图 4 显示 a_i 图斑与 b_h 和 b_{h+1} 具备空间相交的关系。



图 4 空间相交关系 Fig. 4 Spatial intersection relation

按照此方式,就可以依次建立连通序列 $A = B_k$ 的空间关联,并按照A序列顺序重组得到连通序列 $C_k = \{c_1, c_2, \dots, c_j, \dots c_n\}$,输出连通序列集合 $C = \{C_1, C_2, \dots, C_k, \dots, C_y\}$ 。图5为空间校准示意图。



2.3 连通序列相似度评价

在连通序列中,需要比较相似性的2个连通图 斑的区域特征指标一般都不相等,不同连通序列可 能存在连通图斑序列轴上的位移,亦即在还原位移 的情况下,2个连通序列是一致的。考虑到这些复 杂情况,本文通过把连通图斑序列进行延伸和缩短, 在连通图斑序列轴上进行局部的缩放和对齐,使其 形态姿势保持相对一致性,从而计算2个连通序列 的相似性。

图 6 的 2 条实线代表 2 个语义分割模型对同一 景遥感影像预测标签的连通序列,连通序列之间的 虚线代表 2 个连通区序列之间的相似的点。2 个连 通序列在形态上非常相似,但是这些形态特征点 (波峰、波谷)在时间上不能一一对齐。如果使用类 似欧氏距离的常规方法来计算 2 个序列的相似性, 无法符合实际认知。通过对连通图斑序列进行匹配 性的伸缩变形,匹配效果会大幅度增强。本文采用 动态连通区规整算法,算法在匹配 2 个序列时,序列 中的点不再是一一对应关系,而是一对一、一对多和 多对一等不同映射关系。算法的核心在于基于动态 规划匹配连通序列,使得序列相似点之间的距离和 最小。该距离即归整路径距离,用于衡量 2 个连通 序列之间的相似性。



Fig. 6 Dynamic connected area regularity

真实标签连通序列 $A = \{a_1, a_2, \dots, a_i, \dots, a_m\}$,预处理后预测结果连通序列 $C_k = \{c_1, c_2, \dots, c_j, \dots, c_n\}$,序列长度分别为 |A| = m, $|C_k| = n$, A 和 C_k 相似度计算方法如下:

1)构建一个 $m \times n$ 的相似性矩阵网格 **P**,矩阵 **P**(*i*,*j*) 位置的元素表示 a_i 和 c_j 这 2 点间的欧式距 离 $distance(a_i, c_j)$ 。

2)定义搜索起点为(1,1),开始搜索目标点 (*i*,*j*)的上方(*i*-1,*j*)、左方(*i*,*j*-1)和左上方(*i*-1,*j*-1),确定两者之间具有最小矩阵元素之和的 路径,此时路径的元素总数就是连通区归整路径距 离,也就是点*a*_i和*c*_j的欧式距离与可以到达该点的 最小的邻近元素的累积距离之和,即

 $d(i,j) = distance(a_i,c_j) + \min [d(i-1,j),d(i,j-1),d(i-1,j-1)]_{\circ} (1)$

3)重复以上步骤,直至完成从矩阵要素起点 (1,1)到终点(m,n)最短路径的搜索,此时路径的 元素总数就是连通区归整路径距离 d(m,n),即A 和 C_k的连通序列相似度。

2.4 CSIM 计算

CSIM 计算是将所有预测标签对应的连通序列

相似度归一化整合,映射到同一尺度下。本文采用 最值归一化方法将连通相似度集合中的数据映射到 0~1之间。该方法虽然受边界影响较大,但归一化 的值具备较高的辨识度,符合本文预期目的,其计算 方法为:

$$d_{k_{\rm csim}} = \frac{d_k - d_{\min}}{d_{\max} - d_{\min}} , \qquad (2)$$

式中: *d*_k为第 *k* 个预测标签的连通序列相似度值; *d*_{min}和 *d*_{max}分别为所有预测标签的连通序列相似度 的最小值和最大值; *d*_{k_csim}表示连通相似性指数值, 值越大表示第 *k* 个预测标签的连通性保持越好,其 对应的模型语义分割性能更优,更加符合测绘生产 的实际需求。

3 实验分析

3.1 实验数据

为了验证顾及地学特征的遥感地物图斑连通相 似性评价体系应用于语义分割性能评价的可行性, 实验选取具备几何结构多样性和尺度差异性的水体 作为研究对象,实验数据来源于全球测图工程项目 的芬兰西北片区,影像源为资源三号(ZY-3)影像, 空间分辨率为2m。时相为2019年6月24日,光谱 可见光波段为红光、绿光和蓝光。由于实验涉及模 型训练和模型评估2部分,因此实验数据样本包括 训练样本、验证样本和测试样本。从目标影像上分 别裁剪10幅5000像素×5000像素大小的影像切 片,其中7幅作为训练样本,2幅作为验证样本,1幅 作为测试样本。影像对应标签均来源于手工标定, 其中测试图像与其真实标签如图7所示。



(a) ZY - 3 B3(R), B2(G),
 B1(B)波段合成影像
 (b) 人工标定标签
 图 7 测试影像与真实标签



3.2 实验流程

通过将实验区域的小样本数据集进行数据增强 预处理,输入至卷积神经网络框架经过超参数调整 循环训练输出一系列模型集合,应用本文提出的 CSIM,对比现有精度评价指标,进行语义分割性能 评价,选择适配于测绘实际生产的最优模型。

针对模型训练,实验搭建在 TensorFlow 深度学 习框架平台上,基于残差学习机制的 Res – UNet 卷 积神经网络,训练数据集、验证数据集和训练参数作 为输入数据,输出数据为预训练网络模型和模型性 能参数,经过循环迭代,通过判断输出模型的预测值 f(x)与真实值 Y 的不一致程度,发送结束训练的信 号。其训练参数设置如表 3 所示。由于卷积神经网 络的拟合能力非常强,在训练数据集上的错误率降 到最低时,容易造成模型过拟合,从而在验证数据集 上的错误率会相对较高,本文希望获取一个最优泛 化性能的网络模型。为了避免训练模型的过拟合状 态出现,本文应用数据增强模式、Adam 梯度下降优化 算法和提早终止训练等方法来限制模型的复杂度。

表 3 模型训练实验参数

1 ab. 3	Experimental	parameters	01	model	training

参数	数值
样本数量/个	20 000
样本大小/像素	256×256
通道	RGB 3 通道
网络模型	Res – UNet
批尺寸/个	16
预设训练时期/次 ^①	200

①预设训练时期代表训练数据集中所有数据送入网络中,完成 了一次前向计算+反向传播的过程。

针对模型评估,实验基于模型训练的输出成果 集合,通过损失率和传统性能指标优先筛选出目标 模型集合。遍历目标模型集合逐一预测测试影像, 输出预测标签集合,应用本文提出的 CSIM 设计方 法,识别每幅预测标签的连通图斑,对由于影像质量 等其他因素带来的独立像素图斑进行过滤预处理操 作,然后逐一与真实标签进行对比反演,归一化整合 计算 CSIM。为更好地验证语义分割性能评估的有 效性,计算预测标签的现有精度指标,并进一步进行 各类评价指标的数值分析。

3.3 实验结果及分析

模型训练过程中损失函数值演变过程如图 8 所示,在模型训练的不断迭代过程中,损失函数在训练数据集上初期表现出稳步下降的趋势,后期则呈现 波浪状起伏之势,而在测试数据集上表现则一直处 于跌宕起伏状态。损失函数越小,常规意义上代表 模型的鲁棒性越好。为了能获得最好的泛化性能, 很好地拟合训练数据与验证数据,早停法停止训练 的标准取决于泛化错误的变化,即定义当泛化错误 在连续 10 个周期内增长的时候停止模型的继续训 练。由图 8 可以看出模型在第 100 时期时在验证数 据集上得到的损失率最低,此后连续 10 个时期内损 失率相对第 100 时期的指标都处于增长状态,因此 决定在第 109 时期停止训练。





为了验证 CSIM 指标的有效性,模型评估实验 选取第 100 时期前后连续 10 个时期的模型构成目 标模型集合进行验证。图 9 展示以测试影像真实标 签为参考基于空间地理位置匹配的准则,在训练数 据集上相同精确率与相同召回率的不同模型的预测 标签与真实标签的连通图斑集合对比情况。



图 9-1 不同时期模型预测标签连通图斑对比情况





Fig. 9 – 2 Comparison of connected map spots in predicted label by models in different periods

传统指标评价以像素为粒度,CSIM 以像素群为 粒度,两者在时间复杂度上保持一致。图9可以看 出,不同时期模型预测标签连通图斑序列的表现与 真实标签相比整体变化趋势比较相近,这也是精确 率和召回率无法更好地判别语义分割性能差异性的 验证。相似图斑出现的序列存在一定的差异性,差 异性越小,越接近于期望模型。应用 CSIM 评估连 通图斑的差异性,差异性表征 CSIM 指标归一化分 布在[0,1],并与传统评价指标对比(图 10)。图 10 (a) 描述目标模型集合在准确率、精确率和召回率 上的表现力,由于测试影像水体面积占比8.7%,背 景值占比较大,而准确率计算同时考虑到背景值的 预测情况,因此准确率平均高于精确率和召回率,集 中分布在[0.993 2,0.997 7],精确率分布集中在 [0.985 4,0.996 1], 召回率集中分布在[0.953 9, 0.979 0], 召回率在第 100 时期和第 104 时期分别 最大和最小,精确率在第90时期和第97时期分别 最大和最小。图 10(b)展示目标模型集合在 CSIM 表现力,归一化后指数区间分布均匀,第 97 时期指 标值最大,表示该时期模型在测试影像上连通性表 现最完整,而第 106 时期指标值最小,表示该时期模 型在测试影像预测结果连通性表现相对来说差强人 意。对比不同时期模型序列的现有指标和 CSIM,两 者结果呈现弱相关性。







(a) 真实标签
 (b) 第 97 时期 CSIM = 1.0
 (c) 第 100 时期 CSIM = 0.734 (d) 第 90 时期 CSIM = 0.727
 图 11 - 1 不同时期测试影像局部区域的预测二值图像
 Fig. 11 - 1 Predicted label of local area of test image in different periods



(e) 第93 时期 CSIM =0.702 (f) 第94 时期 CSIM =0.433 (g) 第104 时期 CSIM =0.341 (h) 第106 时期 CSIM =0.0
 图 11 −2 不同时期测试影像局部区域的预测二值图像

Fig. 11 - 2 Predicted label of local area of test image in different periods

真实标签、CSIM 较高的 4 个模型与 CSIM 较低的 3 个模型预测的二值图像结果。第 97 时期 CSIM 指 标为 1,理论上表示该模型在水体连通性上保持能 力最佳,实际预测结果显示①号湖泊没有零星图斑, 完整性保持最好,局部边界有一点偏差。第 106 时 期 CSIM 指标值为 0,表征其连通相似性在目标模型 集合中表现能力是最不好的,实际预测结果显示① 号湖泊零碎图斑较多,没有保持湖泊连通性的几何 特征,而且几乎丢失湖中心岛的空间位置信息。以 上说明了 CSIM 指标的可靠性,具有很高的置信度, 与预测标签连通性保持着高度的一致性。

4 结论

本文针对遥感解译和测绘生产智能解译需求, 分析了现有语义分割性能评价体系没有涉及到空间 关系特征保持,难以有效支撑时空异质的遥感影像 语义分割结果的评价,因此提出了一种顾及地学特 征的遥感影像语义分割模型性能评价方法——连通 相似性指数 CSIM,基于现有的模型评价指标,构建 复合指标体系,可以更好地实时监测和控制模型训 练;面向不同地理地域特征的不同时相的不同影像 源,可以有效地指导预训练模型的最优选取;同时 实现在像素粒度的基础上顾及地物几何特征的遥感 影像预测结果真实质量的准确评估。

本方法还有待改进的地方在于性能评价因子涉 及到的地学特征更多地是关注自身,包括像素是否 归类正确、像素群是否区域连通,没有涉及到邻近特 征的考量。针对这一问题,考虑在下一步工作中将 地学先验知识纳入评价因子中,进一步丰富完善遥 感影像语义分析模型性能评价方法。

参考文献(References):

[1] 龚健雅,许越,胡翔云,等. 遥感影像智能解译样本库现状与研究[J]. 测绘学报,2021,50(8):1013-1022.

Gong J Y, Xu Y, Hu X Y, et al. Status analysis and research of sample database for intelligent interpretation of remote sensing im-

age[J]. Acta Geodaetica et Cartographica Sinica, 2021, 50(8): 1013 – 1022.

- [2] Rawat W, Wang Z H. Deep convolutional neural networks for image classifycation: A comprehensive review [J]. Neural Computation, 2017,29(9):2352 - 2449.
- [3] Yuan Q Q, Shen H F, Li T W, et al. Deep learning in environmental remote sensing; Achievements and challenges [J]. Remote Sensing of Environment, 2020, 241:111716.
- [4] Zhang L P, Zhang L F, Du B. Deep learning for remote sensing data: A technical tutorial on the state of the art[J]. IEEE Geoscience and Remote Sensing Magazine, 2016, 4(2):22 - 40.
- [5] 洪 亮,冯亚飞,彭双云,等. 面向对象的多尺度加权联合稀疏 表示的高空间分辨率遥感影像分类[J]. 测绘学报,2022,51
 (2):224-237.

Hong L, Feng Y F, Peng S Y, et al. Classification of high spatial resolution remote sensing imagery based on object – oriented multiscale weighted sparse representation [J]. Acta Geodaetica et Cartographica Sinica, 2022, 51(2):224 – 237.

- [6] Zhang P L, Lyu Z Y, Shi W Z. Object based spatial feature for classification of very high resolution remote sensing images [J].
 IEEE Geoscience and Remote Sensing Letters, 2013, 10(6):1572 – 1576.
- [7] 段嘉鑫.基于上下文信息的图像分割结果质量评价方法研究
 [D].成都:电子科技大学,2021.
 Duan J X. Research on evaluation method of image segmentation results quality based on context information[D]. Chengdu:University of Electronic Science and Technology,2021.
- [8] 党 宇. 基于卷积神经网络的地表覆盖分类自动质量评价方法研究[D]. 徐州:中国矿业大学,2018.
 Dang Y. Study on automatic evaluation method of land cover classification based on convolutional neural network[D]. Xuzhou: University of Mining and Technology of China,2018.
- [9] 郑明国,蔡强国,秦明周,等.一种遥感影像分类精度检验的新 方法[J]. 遥感学报,2006,10(1):39-48. Zheng M G,Cai Q G,Qin M Z,et al. A new approach to accuracy assessment of classifications of remotely sensed data[J]. Journal of Remote Sensing,2006,10(1):39-48.
- [10] 欧阳松. 地学知识嵌入的遥感影像深度语义分割方法研究
 [D]. 武汉:武汉大学,2021.
 Ouyang S. The methodology research on geographic knowledge guided deep semantic segmentation of remote sensing imagery[D].
 Wuhan; Wuhan University,2021.
- [11] 陶 超,阴紫薇,朱 庆,等. 遥感影像智能解译:从监督学习到 自监督学习[J]. 测绘学报,2021,50(8):1122-1134.

Tao C, Yin Z W, Zhu Q, et al. Remote sensing image intelligent interpretation: From supervised learning to self – supervised learning [J]. Acta Geodaetica et Cartographica Sinica, 2021, 50(8):1122 – 1134.

- [12] 龚健雅.人工智能时代测绘遥感技术的发展机遇与挑战[J]. 武汉大学学报(信息科学版),2018,43(12):1788-1796.
 Gong J Y. Chances and challenges for development of surveying and remote sensing in the age of artificial intelligence[J]. Geomatics and Information Science of Wuhan University,2018,43(12): 1788-1796.
- [13] 叶利华,王 磊,张文文,等. 高分辨率光学遥感场景分类的深度度量学习方法[J]. 测绘学报,2019,48(6):698-707.
 Ye L H, Wang L, Zhang W W, et al. Deep metric learning method for high resolution remote sensing image scene classification[J].
 Acta Geodaetica et Cartographica Sinica,2019,48(6):698-707.
- [14] 常亮,邓小明,周明全,等.图像理解中的卷积神经网络[J]. 自动化学报,2016,42(9):1300-1312.
 Chang L, Deng X M, Zhou M Q, et al. Convolutional neural networks in image understanding[J]. Acta Automatica Sinica,2016, 42(9): 1300-1312.

[15] 曹玉东,刘海燕,曹 旭,等. 基于深度学习的图像质量评价方

法综述[J]. 计算机工程与应用,2021,57(23):27-36. Cao Y D,Liu H Y. Cao X, et al. Overview of image quality assessment method based on deep learning[J]. Computer Engineering and Applications,2021,57(23):27-36.

[16] 钱晓亮,李 佳,程 塨,等.特征提取策略对高分辨率遥感图像
 场景分类性能影响的评估[J].遥感学报,2018,22(5):758 - 776.

Qian X L, Li J, Cheng G, et al. Evaluation of the effect of feature extraction strategy on the performance of high – resolution remote sensing image scene classification[J]. Journal of Remote Sensing, 2018,22(5):758-776.

- [17] 李琳,余胜生.基于深度学习模型的图像质量评价方法[J]. 华中科技大学学报(自然科学版),2016,44(12):70-75.
 Li L, Yu S S. Image quality evaluation method based on deep learning model [J]. Journal of Huazhong University of Science and Technology(Natural Science Edition),2016,44(12):70-75.
- [18] 薛 廉,周春兰.基于监督分类分区域的特征因子提取[J].地 理空间信息,2009,7(2):100-102.
 Xue L, Zhou C L. Feature factor extraction based on supervised classification subregions[J]. Geospatial Information,2009,7(2): 100-102.

A performance evaluation method for semantic segmentation models of remote sensing images considering surface features

LIU Li^{1,2}, DONG Xianmin^{1,2}, LIU Juan¹

(1. The Third Geographical Information Mapping Institute of Natural Resources Ministry, Chengdu 610100, China; 2. Key Laboratory on Digital Mapping and Land Information Application, Natural Resources Ministry, Chengdu 610100, China)

Abstract: Deep semantic segmentation has been widely applied in land monitoring and interpretation based on remote sensing images. However, existing quality evaluation methods cannot reflect the preserved spatial geometric features of semantic segmentation results. Based on the practical demand for remote sensing image interpretation, surveying, and mapping, this study proposed a method for the performance evaluation of semantic segmentation models for remote sensing images considering geoscience features: the connectivity similarity index (CSIM). From the perspective of the connectivity similarity of surface feature spots in remote sensing images, the CSIM method embedded the surface features into the performance evaluation system of semantic segmentation models. The CSIM method allows for quantitatively evaluating the connectivity similarity of spots between the semantic segmentation results of remote sensing images and the actual sample labels, thus accurately describing the preserved spot integrity in the predicted classification results. Therefore, the CSIM method can objectively determine the applicability of a pre – training model to remote sensing image interpretation in surveying and mapping production. As substantiated by a lot of practice, the CSIM method can monitor and control the model training in real time, effectively guide the selection of the optimal pre – training model, and accurately evaluate the quality of remote sensing image interpretation, surveying, and mapping.

Keywords: model performance evaluation; semantic segmentation; geoscience feature; remote sensing interpretation; deep learning