

DOI:10.19651/j.cnki.emt.2108505

# IGBP 土地覆盖升尺度及精度评价方法研究\*

陈施施<sup>1</sup> 谢涛<sup>1</sup> 王超<sup>2</sup>

(1.南京信息工程大学遥感与测绘工程学院 南京 210044; 2.南京信息工程大学电子与信息工程学院 南京 210044)

**摘要:** 作为陆地生态系统中的生物物理参量之一,土地覆盖是支撑城市动态监测等多种科学研究的基础变量。面向基于 IGBP 体系的土地覆盖产品,提出了一种基于语义邻近度和聚类规则集合的升尺度方法。并在此基础上,针对当前升尺度评价方法中缺少真实参考数据,设计了一种基于类内相似性的精度评价指标 ICS。基于 MODIS 的 MCD12C1 数据进行了实验,并分别使用占比法与 ICS 对不同升尺度方法结果开展了综合对比分析与讨论,其中所设计的语义邻近度方法和规则集合法在加权 ICS 中分别达到 84.1% 和 84.8%,验证了所提出方法的可行性与有效性。

**关键词:** 土地覆盖;升尺度;语义邻近度;类内相似性

**中图分类号:** TP79 **文献标识码:** A **国家标准学科分类代码:** 210.5

## Research on IGBP land cover upscaling and accuracy evaluation method

Chen Shishi<sup>1</sup> Xie Tao<sup>1</sup> Wang Chao<sup>2</sup>

(1. School of Remote Sensing &amp; Geomatics Engineering, Nanjing University of Information Science &amp; Technology, Nanjing 210044, China; 2. School of Electronic &amp; Information Engineering, Nanjing University of Information Science &amp; Technology, Nanjing 210044, China)

**Abstract:** As one of the biophysical parameters in terrestrial ecosystems, land cover is a basic variable that supports various scientific studies such as urban dynamic monitoring. For land cover products based on the IGBP system, proposes an upscaling method based on semantic proximity and clustering rule sets. And on this basis, in view of the lack of real reference data in the current upscaling evaluation method, an accuracy evaluation index ICS based on structural similarity is designed. The experiments based on the MCD12C1 data of MODIS were carried out, and the results of different upscaling methods were comprehensively compared, analyzed and discussed using the proportion method and ICS respectively. The designed semantic proximity method and rule set method reached 84.1% and 84.8% in weighted ICS, respectively, and the feasibility and effectiveness of the proposed method were verified.

**Keywords:** land cover; up-scale; semantic proximity; structural similarity

## 0 引言

土地覆盖作为影响陆地生态系统机能的一个重要物理参量,其数据在地理学、大气科学、农学、林学等领域具有重要应用<sup>[1-2]</sup>。随着空间技术的迅速发展,遥感数据因其覆盖范围广、多时相、多光谱等优势,目前已成为土地覆盖分类的重要数据源<sup>[3]</sup>。其中,传统的 MODIS 以及我国的 FY 系列土地覆盖产品均采用国际地圈生物圈计划(International Geosphere Biosphere Programme, IGBP)。作为目前主流的土地覆盖分类体系,IGBP 将土地覆盖分为 11 类自然植被、3 类开发过的土地以及 3 类非植被土地类型<sup>[4]</sup>。由于该体系具有的开放性与高效性,在全球及区域尺度的土地监测和制图上受到了用户和遥感研究者的积极关注与使用<sup>[5]</sup>。

尽管如此,当前现有可获取的遥感土地覆盖产品通常存在尺度单一的局限,无法满足全球或不同区域尺度的生态、水文、气象等研究<sup>[6]</sup>;此外,目前陆地标准产品的尺度变化研究,主要使用降尺度方法<sup>[7-8]</sup>,来弥补空间分辨率的不足,以满足实际应用的需求;最后,目前尚缺乏基于 IGBP 标准的地表覆盖分类产品的升尺度方法,同时也缺乏衡量升尺度效果的精度评价方法。因此,围绕这些问题展开研究非常必要,以提供高精度、高可靠的地表覆盖升尺度产品。

围绕遥感数据的升尺度方法,学者们已经取得了一系列的成果。根据对聚合窗口中类别值的处理方式的不同,常见方法有众数聚合法,随机抽样法和基于语义邻近度的

收稿日期:2021-12-04

\* 基金项目:国家自然科学基金面上项目(42176180)资助

方法等<sup>[9-10]</sup>。众数聚合法即在窗口中,用数量多的类别代替窗口中的其他所有类别。该方法能够更好地保留优势类别,但会导致弱势类别的减少甚至消失,从而导致类别空间结构变化。随机抽样法是在聚合窗口中随机选择一个类别作为聚合后的类别值,有利于保持原尺度各类别的面积比例,但会使空间结构趋于离散。基于语义邻近度的方法,能够削弱众数聚合法中高优先级类别更容易被保留下来的绝对优势,但其精度容易受设计的语义邻近层次图的影响,需要根据不同的分类规则去设计对应且合理的语义邻近层次图<sup>[11]</sup>。特别是,目前尚缺乏面向 IGBP 标准的语义邻近层次图,限制了此类方法在地表覆盖产品中的适用性。在精度评价方面,由于仅有升尺度前产品,而缺乏升尺度标准结果作为参照,因此通常仅通过比例变化来判断升尺度结果的优劣。这种指标虽然在一定程度上能够评判升尺度后结果相对原数据各类别的占比变化情况<sup>[12]</sup>,但比例是否变化无法直接反应升尺度结果的精度及可靠性。

鉴于此,本文设计了一套完整的 IGBP 土地覆盖升尺度及精度评价方法。首先,基于 IGBP 体系的语义邻近层次图,并成功应用在语义邻近度法中。此外,通过参考 MODIS 产品的聚类规则,提出了一种升尺度规则集合。在此基础上,提出了一种基于类内相似性<sup>[13-14]</sup> (structural similarity, SSIM) 的新的度量方法,并结合占比法对众数聚合、随机抽样、语义邻近度及规则集合法的性能进行了综合

评估,包括对不同时间段同一地区的土地覆盖数据的目视分析,用占比法测试得到的精度评价和采用类内相似性得到的精度评价。结果表明,基于 IGBP 体系所提出的语义邻近度和规则集合两种升尺度方法,均是可行且有效的。特别是语义邻近度方法,在总体和单类精度中均显著优于其他对比方法。

## 1 方法与实验

### 1.1 IGBP 分类体系

IGBP 是由国际科学联盟理事会 (international council for science, ICSU) 发起并组织的重大国际科学计划,其科学目标是为了阐明和了解整个地球系统的物理、化学和生物学过程,从而能够预测地球系统在未来十年至百年时间尺度上的变化<sup>[15]</sup>。由于 IGBP 采用严谨的国际机构指导全球变化科学,被广泛用于世界范围的国家科学计划,协助开展相应的科学研究。

IGBP 分类标准将土地覆盖分为 17 个地物类别,如表 1 所示。这 17 个类别按照覆盖成因,覆盖类型等的不同,又可将其分为 3 个大类。其中,表 1 中的常绿针叶林至永久湿地都属于自然植被类;农田、城市建筑区和农田与植被相交属于开发过的土地类;水体、冰与雪和贫瘠属于非植被土地类。这三大类中的各个小类之间均具有较高的语义邻近度,这也是本文构建语义邻近层次图的依据所在。

表 1 地物类别优先级规则

类别	水体	常绿针叶林	常绿阔叶林	落叶针叶林	落叶阔叶林	混交林	密集灌木	疏松灌木	多树荒原
序号	1	2	3	4	5	6	7	8	9
类别	荒原	草原	永久湿地	农田	城市建筑区	农田与植被相交	冰与雪	贫瘠	
序号	10	11	12	13	14	15	16	17	

### 1.2 算法的设计与实现

#### 1) 语义邻近度法

##### (1) 语义邻近层次图和语义邻近度计算

结合 IGPB 分类体系中不同类别的语义关系<sup>[16]</sup>,本文设计的语义邻近层次如图 1 所示。

在此基础上,定义每个类别的语义多层次细节表达中的语义特征项  $\{I_i: \langle Class_i, \{n_1, n_2, \dots, n_j\} \rangle, R_i\}$ , 其中,  $I_i$  表示层次图中第  $i$  个语义特征项,  $Class_i$  表示该项名称,  $\{n_1, n_2, \dots, n_j\}$  表示该项值域,  $R_i \in \{1, 2, 3\}$  表示该项层次的级别。如{覆盖成因, {自然营造, 人工建造}}。对任意土地覆盖数据实体对象  $E_i$ , 定义函数  $N_{IR}(E_i, I)$  代表其在第  $i$  个语义特征项上的取值。若  $E_i$  在语义特征项  $I$  上有语义特征, 则  $N_{IR}(E_i, I) = I.v$ ; 反之,  $N_{IR}(E_i, I) = 0$ 。综上, 土地覆盖数据实体对象  $E_i$  的语义特征表达函数可表示为:  $\{I_1.v, I_2.v, \dots, I_n.v, n = 15\}$ 。计算语义邻近度的具体步骤如下:

Step 1: 根据层次图对 17 个类别进行语义特征的多层次细节表达, 具体如下。

水体:  $E_1 = \{ '自然营造', '湿区', '其他用地', '已利用', 0, 0, 0, '较多', 0, 0, 0, 0, 0, 0, 0 \}$ 。

常绿针叶林:  $E_2 = \{ '自然营造', '植被', '其他用地', '未利用', '木本', '旱生', '非农作物', 0, 0, 0, '覆盖大于 60%', '大于 2 m', '常年绿色', '针叶', 0 \}$ 。

常绿阔叶林:  $E_3 = \{ '自然营造', '植被', '其他用地', '未利用', '木本', '水生', '非农作物', 0, 0, 0, '覆盖大于 60%', '大于 2 m', '常年绿色', '阔叶', 0 \}$ 。

落叶针叶林:  $E_4 = \{ '自然营造', '植被', '其他用地', '未利用', '木本', '旱生', '非农作物', 0, 0, 0, '覆盖大于 60%', '大于 2 m', '落叶周期', '针叶', 0 \}$ 。

落叶阔叶林:  $E_5 = \{ '自然营造', '植被', '其他用地', '未利用', '木本', '水生', '非农作物', 0, 0, 0, '覆盖大于 60%', '大于 2 m', '落叶周期', '阔叶', 0 \}$ 。

混交林:  $E_6 = \{ '自然营造', '植被', '其他用地', '未利用', '木本', '均有', '非农作物', 0, 0, 0, '覆盖大于 60%', '大于 2 m', '情况均有', '均有', 0 \}$ 。

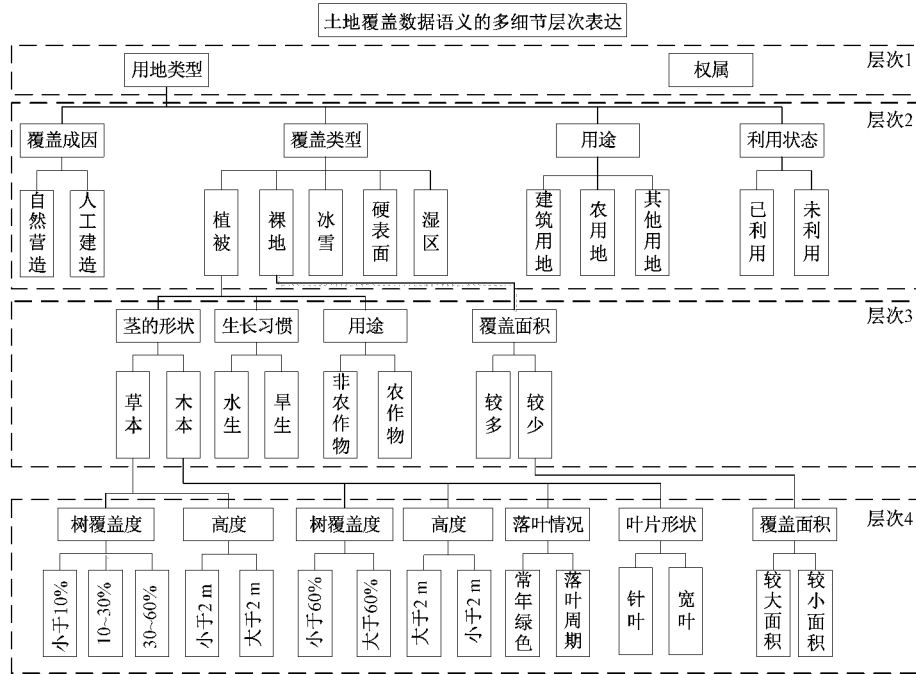


图 1 土地覆盖数据语义的多细节层次

密集灌木:  $E_7 = \{ '自然营造', '植被', '其他用地', '未利用', '木本', '水生', '非农作物', 0, 0, 0, '覆盖大于 60%', '小于 2 m', 0, 0, 0 \}$ 。

疏松灌木:  $E_8 = \{ '自然营造', '植被', '其他用地', '未利用', '木本', '旱生', '非农作物', 0, 0, 0, '10% \sim 60%', '小于 2 m', 0, 0, 0 \}$ 。

多树荒原:  $E_9 = \{ '自然营造', '植被', '其他用地', '未利用', '均有', '水生', '非农作物', 0, '30% \sim 60%', '大于 2 m', 0, 0, 0, 0 \}$ 。

荒原:  $E_{10} = \{ '自然营造', '植被', '其他用地', '未利用', '草本', '旱生', '非农作物', 0, '10% \sim 30%', '大于 2 m', 0, 0, 0, 0 \}$ 。

草原:  $E_{11} = \{ '自然营造', '植被', '其他用地', '已利用', '草本', '水生', '非农作物', 0, '从 0% \sim 10%', '小于 2 m', 0, 0, 0, 0, 0 \}$ 。

永久湿地:  $E_{12} = \{ '自然营造', '湿区', '其他用地', '已利用', 0, 0, 0, '较少', 0, 0, 0, 0, 0 \}$ 。

农田:  $E_{13} = \{ '人工建造', '植被', '农业用地', '已利用', '草本', '水生', '含农作物', 0, '不含有树木 0%', '小于 2 m', 0, 0, 0, 0, '较大面积' \}$ 。

城市与建成区:  $E_{14} = \{ '人工建造', '硬地', '建筑用地', '已利用', 0, 0, 0, 0, 0, 0, 0, 0, 0 \}$ 。

农田与天然植被相交:  $E_{15} = \{ '人工建造', '植被', '其他用地', '已利用', '均有', '水生', '含农作物', 0, '不含有树木 0%', '小于 2 m', 0, 0, 0, 0, '较小面积' \}$ 。

冰与雪:  $E_{16} = \{ '自然营造', '冰雪', '其他用地', '未利用', 0, 0, 0, 0, 0, 0, 0, 0, 0 \}$ 。

贫瘠:  $E_{17} = \{ '自然营造', '裸土', '其他用地', '未利用', 0, 0, 0, 0, 0, 0, 0, 0, 0 \}$ 。

Step 2: 根据上述各类别的语义特征计算  $G_{SED}(E_1, E_2)$  和  $G_{SUD}(E_1, E_2)$ 。其中函数  $G_{SED}(E_1, E_2)$  代表实体  $E_1$  和  $E_2$  语义特征相同部分的度量值,  $G_{SUD}(E_1, E_2)$  代表实体  $E_1$  和  $E_2$  语义特征不同的部分中属于  $E_1$  而不属于  $E_2$  的度量值, 则实体  $E_1$  和  $E_2$  的语义邻近度  $G_{SND}(E_1, E_2)$  可表达为公式:

$$G_{SND}(E_1, E_2) = \frac{G_{SED}(E_1, E_2)}{G_{SED}(E_1, E_2) + G_{SUD}(E_1, E_2) + G_{SUD}(E_2, E_1)} \quad (1)$$

其语义邻近度取值在 0~1 之间。

如: 对于实体  $E_2$  和  $E_3$ , 语义特征相同部分的度量值  $G_{SED}(E_1, E_2) = 2 + 2 + 2 + 2 + 3 + 3 + 4 + 4 + 4 = 26$ , 语义不同部分的度量值  $G_{SUD}(E_1, E_2) = G_{SUD}(E_2, E_1) = 3 + 4 = 7$ , 则其语义邻近度为  $G_{SND}(E_1, E_2) = \frac{26}{26 + 7 + 7} = 0.65$ 。

Step 3: 依据 Step2, 遍历所有类别, 可得 17 个类别的语义邻近度结果, 如表 2 所示。

(2) 方法实现流程

采用所提出的语义邻近度方法对土地覆盖数据进行升尺度的流程如图 2 所示。

实现步骤如下:

Step 1: 读取 HDF 文件, 计算在分类结果(uint8 类型)矩阵基础上, 分别统计滤波窗口(采用  $5 \times 5$  尺寸)内属于每种类别像素的个数。

Step 2: 根据式(2)计算当前窗口中每个类别的权重度

表 2 语义邻近度表

优先级	17	16	15	14	13	12	11	10
类别	水体	常绿针叶林	常绿阔叶林	落叶针叶林	落叶阔叶林	混交林	密集灌木	疏松灌木
水体	1	0.33	0.33	0.33	0.33	0.33	0.33	0.33
常绿针叶林	0.33	1	0.65	0.78	0.5	0.5	0.56	0.52
常绿阔叶林	0.33	0.65	1	0.5	0.78	0.5	0.72	0.39
落叶针叶林	0.33	0.78	0.5	1	0.65	0.5	0.56	0.52
落叶阔叶林	0.33	0.5	0.78	0.65	1	0.5	0.72	0.39
混交林	0.33	0.5	0.5	0.5	0.5	1	0.562 5	0.39
密集灌木	0.33	0.56	0.72	0.56	0.72	0.56	1	0.56
疏松灌木	0.33	0.51	0.39	0.52	0.39	0.39	0.56	1
多树荒原	0.33	0.48	0.7	0.48	0.7	0.48	0.7	0.48
荒原	0.33	0.7	0.47	0.7	0.47	0.48	0.48	0.7
草原	0.6	0.36	0.54	0.36	0.54	0.36	0.55	0.36
永久湿地	0.57	0.33	0.33	0.33	0.33	0.33	0.33	0.33
农田	0.14	0.06	0.17	0.06	0.17	0.06	0.17	0.06
城市与建成区	0.14	0	0	0	0	0	0	0
农田与天然植被相交	0.33	0.13	0.26	0.13	0.26	0.13	0.26	0.13
冰与雪	0.33	0.6	0.6	0.6	0.6	0.6	0.6	0.6
贫瘠	0.33	0.6	0.6	0.6	0.6	0.6	0.6	0.6
9	8	7	6	5	4	3	2	1
多树荒原	荒原	草原	永久湿地	农田	城市与建成区	农田与天然植被相交	冰与雪	贫瘠
0.33	0.33	0.6	0.57	0.14	0.14	0.33	0.33	0.33
0.48	0.7	0.36	0.33	0.06	0	0.13	0.6	0.6
0.7	0.48	0.55	0.33	0.17	0	0.26	0.6	0.6
0.48	0.7	0.36	0.33	0.06	0	0.13	0.6	0.6
0.7	0.48	0.55	0.33	0.17	0	0.26	0.6	0.6
0.48	0.48	0.36	0.33	0.06	0	0.13	0.6	0.6
0.7	0.48	0.55	0.33	0.17	0	0.26	0.6	0.6
0.48	0.7	0.36	0.33	0.06	0	0.13	0.6	0.6
1	0.42	0.32	0.33	0.11	0	0.25	0.6	0.6
0.43	1	0.32	0.33	0.11	0	0.09	0.6	0.6
0.32	0.32	1	0.6	0.39	0.14	0.35	0.33	0.33
0.33	0.33	0.6	1	0.14	0.14	0.33	0.33	0.33
0.11	0.11	0.39	0.14	1	0.33	0.45	0	0
0	0	0.14	0.14	0.33	1	0.14	0	0
0.25	0.09	0.35	0.33	0.45	0.14	1	0.14	0.14
0.6	0.6	0.33	0.33	0	0	0.14	1	0.6
0.6	0.6	0.33	0.33	0	0	0.14	0.6	1

指标  $\omega_i$ 。其中,  $i$  与表 1 中的类别对应, 如  $i = 1$  对应表中的水体, 优先级  $P_i$  最高, 值为 1; 以此类推,  $i = 17$  对应“贫瘠”类别, 优先级  $P_i$  最低, 值为 17。  $S(i, j)$  为本文所采用的语义邻近度。

$$\omega_i = N_i + \sum_{j \neq i} [N_j P_j S(i, j)] \quad (2)$$

Step 3: 根据权重度指标从大到小排序, 取权重度指标最大的类别为当前窗口的类别。若两种类别计算得到相同的权重度指标, 则根据类别优先级, 选择优先级高的为当前窗口的标签。

Step 4: 以  $5 \times 5$  窗口对 MCD12C1 产品进行升尺度采

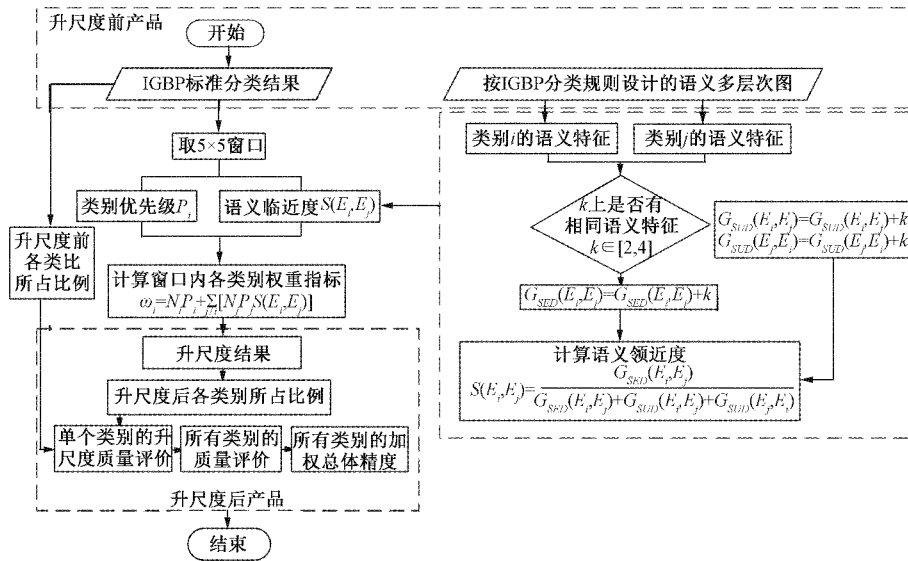


图 2 语义邻近度法流程

样,并按照 Step1~3 赋予每个窗口(即升尺度后的 1 pixel)对应的标签,从而获得最终升尺度结果。

2) 判别规则集合法

MODIS 说明文件“MCD12\_User\_Guide\_V6”中

“5.2 MCD12C1 IGBP Legend”给出了待升尺度产品生成过程中所采用的聚类策略。受此启发,本文设计了设计基于 IGBP 分类标准的升尺度判别规则集合如表 3 所示。依据该规则集合,可直接对 5×5 尺寸窗口赋予升尺度结果标签。

表 3 判别规则集合

类别	规则
水体	水体像素占比超过 60%
常绿针叶林	常绿针叶林像素数量最多+树林(常绿针叶+常绿阔叶+落叶针叶+落叶阔叶+混交)像素占比超过 60%
常绿阔叶林	常绿阔叶林像素数量最多+树林(常绿针叶+常绿阔叶+落叶针叶+落叶阔叶+混交)像素占比超过 60%
落叶针叶林	落叶针叶林像素数量最多+树林(常绿针叶+常绿阔叶+落叶针叶+落叶阔叶+混交)像素占比超过 60%
落叶阔叶林	落叶阔叶林像素数量最多+树林(常绿针叶+常绿阔叶+落叶针叶+落叶阔叶+混交)像素占比超过 60%
混交林	混交林像素最多+树林(常绿针叶+常绿阔叶+落叶针叶+落叶阔叶+混交)像素占比超过 60%
密集灌木	密集灌木像素最多+灌木像素(密集+疏松)占比超过 60%
疏松灌木	疏松灌木像素最多+灌木像素(密集+疏松)占比在 10%~60%之间
多树荒原	树林(常绿针叶+常绿阔叶+落叶针叶+落叶阔叶+混交)像素占比在 30%~60%之间
荒原	树林(常绿针叶+常绿阔叶+落叶针叶+落叶阔叶+混交)像素占比在 10%~30%之间
草原	草原像素最多
永久湿地	水体像素占比在 30%~60%之间+植被(树林+灌木+荒原+草原+农田+农田与天然植被相交)像素超过 10%
农田	农田像素占比超过 60%
城市与建成区	城市与建成区像素最多且占比超过 30%
农田与天然植被相交	农田与天然植被相交像素最多+植被像素在 40%~60%之间
冰与雪	冰与雪像素占比超过 60%
贫瘠	贫瘠像素占比超过 60%,且植被像素的占比小于 10%
无数据或未分类	不满足以上判别条件的均属于此类

### 1.3 升尺度结果评价指标

为了在无真实参考数据条件下对不同升尺度方案进行有效的定量分析,本文设计了一种基于 SSIM 的精度评价指标(inner class SSIM, ICS),其依据在于:升尺度之后,属于同一类别像素之间的结构自相似性越大,则每个类别像素之间的联系就越紧密(或相似度越高),升尺度结果中各个类别的划分就越准确。每个升尺度后像素都可以看成由  $5 \times 5$  窗口内的混合像元而来,因此每个像素实际可以看成是一个  $5 \times 5$  的图像,因此计算两个窗口之间的 ICS,能够反映被赋予相同类别标签像素之间的关联程度,也可以反映被赋予不同类别标签像素之间的差异大小,具体步骤如下。

Step1:在升尺度结果基础上,以全部属于水体的像素为例,遍历计算每两个水体像素之间的 ICS,得到水体像素之间的平均相似程度  $ICS_{water-water}$ ,依次类推,获得所有类别的内部像素  $ICS_{xxx-xxx}$ 。

Step2:对全部类别的  $ICS_{xxx-xxx}$  求均值,得到类内可分性精度指标:

$$ICS_{intra-class} = \frac{\sum_{i=1}^{17} ICS_{xxx-xxx}}{17} \quad (3)$$

Step3:进而求得加权类内可分性指标,其中  $P_i$  为升尺度结果中类别  $i$  所占比例:

$$ICS_{ratio-intra-class} = \sum_{i=1}^{17} (ICS_{xxx-xxx} \times P_i) \quad (4)$$

## 2 实验结果分析

### 2.1 数据选取

本文实验以 MODIS 土地覆盖产品 MCD12C1 为实验数据,空间分辨率为 5 km,存储为 HDF 格式,大小为 7 200 像元  $\times$  3 600 像元,数据覆盖全球。共对 3 组实验数据进行了实验,产品拍摄时间为 2002 年 1 月 1 日、2015 年 1 月 1 日、2018 年 1 月 1 日和 2019 年 1 月 1 日。

### 2.2 目视结果分析

本文采用  $5 \times 5$  的聚合窗口,分别采用众数聚合(方案 1)<sup>[17]</sup>、随机抽样(方案 2)<sup>[18]</sup>、本文提出的语义邻近度(方案 3)与判别规则集合法(方案 4)4 种方案进行实验。由于全球结果图,不利于细节展示,因此仅展示局部区域的原图及对应 4 种方案的升尺度结果图,如图 3~6 所示。

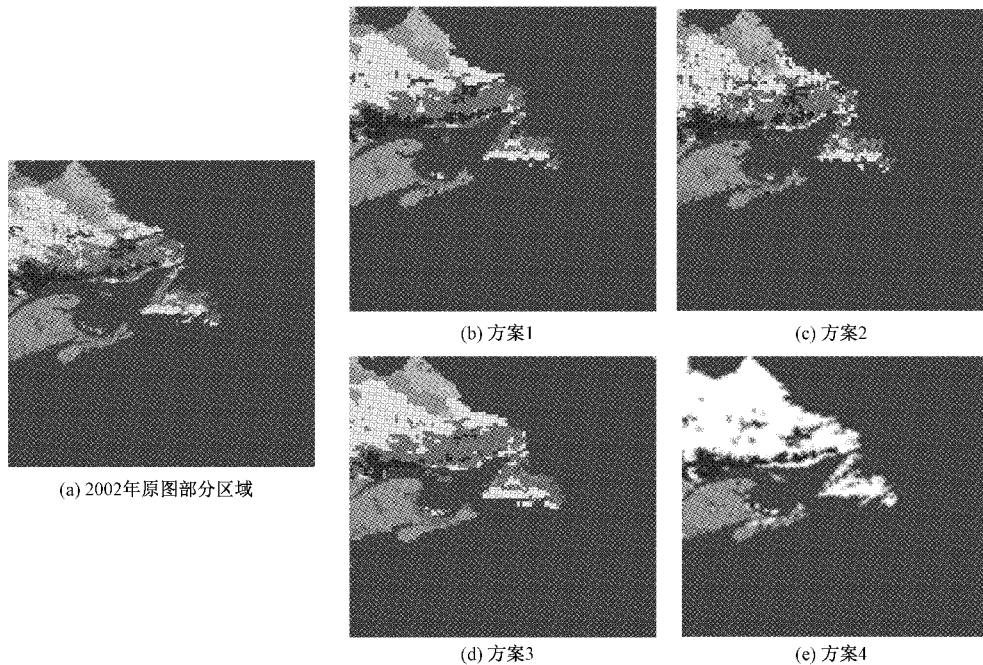


图 3 2002 年局部图及升尺度结果

通过目视分析可以看出,本文提出的语义邻近度(方案 3)在整体上要优于方案 1 和方案 2。方案 4 采用的是 MODIS 官方判别规则进行升尺度,相比 IGBP 分类规则要多出“无数据或未分类”这一大类,而在本文中,对原数据的可视化处理方式上采用的是 IGBP 分类规则,所以在目视分析中,与其他方案有明显差距。分别对图 3~6 中各方案结果分析,由图 3(b)众数聚合法的结果看,不同级别

类别同时出现时,高优先级类别会占据绝对优势,而级别低且像素数少的类别所占比例大大减少;由图 3(c)来看,部分小数量类别因随机抽样法的随机性,分布较之原图,变得不连续,但总体而言,较好保留了类别空间格局;从图 3(d)可看出,加入语义邻近度信息后,一定程度上减缓了众数聚类法中类比优先级高的绝对优势,且类别空间格局也没有出现大的变化;图 3(e)则为根据官方判别规则制

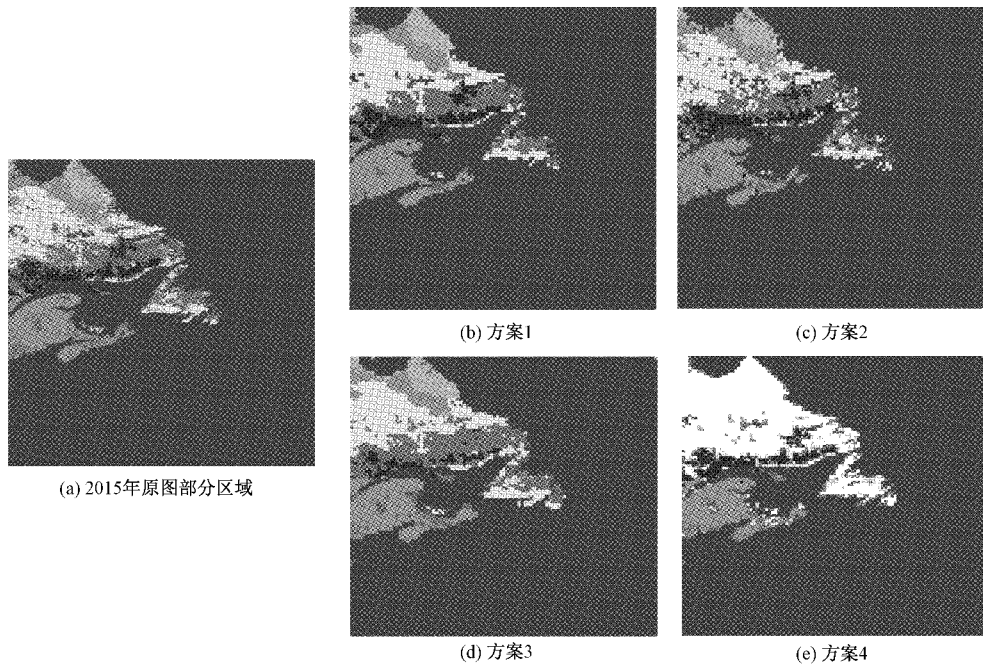


图 4 2015 年局部图及升尺度结果

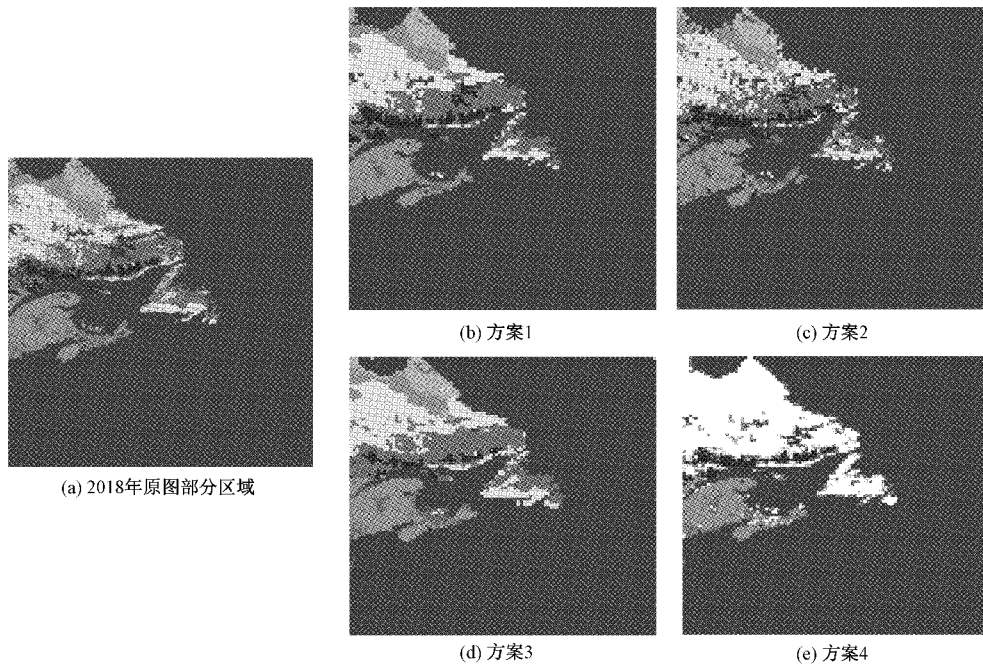


图 5 2018 年局部图及升尺度结果

定的分类方案,可见效果不佳,主要体现在由于分类规则过于严格导致出现了大量“无数据或未分类”类别。图 5~6 结果和图 3 呈相同趋势。因此,方案 3 相比方案 1,加入语义邻近度信息后,一定程度上减缓了方案 1 中优先级高的类别占据绝对优势的问题,且类别空间格局也没有出现大的变化。方案 2 相比于其他方案,因为采用了随机抽样的原则,较好地保持了整体的空间格局,但从单个类别来看,一些小数量的类别分布较之原图,变得不连续,破坏了单类别

的结构。方案 4 因为分类规则的不同,单从目视结果来看“无数据或未分类”占据多数,因此效果不佳。

### 2.3 定量评价

为了定量分析 4 种方案的优劣,本文通过占比法<sup>[19]</sup>和本文提出的 ICS 指标来进行评价。其中,占比法是利用的原理是在对土地覆盖数据进行升尺度处理前后,理想的升尺度算法应不改变每个类别像素所占比例,评价结果如表 4~7 所示。

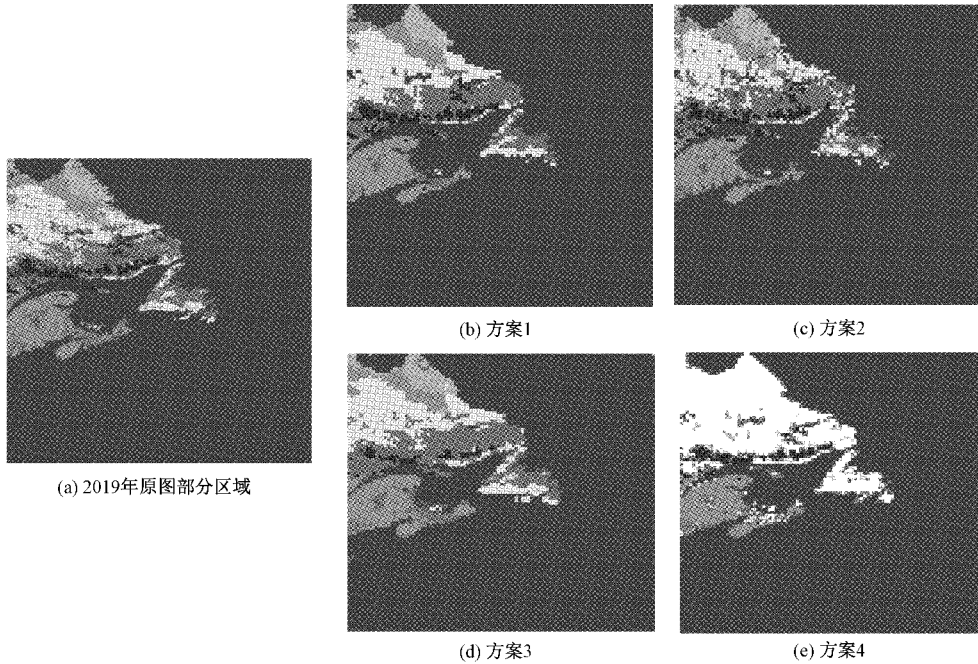


图 6 2019 年局部图及升尺度结果

表 4 2002 年数据占比法评价结果

精度/方案	方案 1	方案 2	方案 3	方案 4
评价标准	越大越好	越大越好	越大越好	越大越好
$ACC_{avg}$	0.899 645	0.899 964	0.897 924	0.887 208
$ACC_{weight}$	0.999 371	0.999 848	0.989 960	0.991 961
水体	0.999 274	0.999 799	0.986 534	0.994 865
常绿针叶林	0.999 893	0.999 979	0.997 154	0.998 714
落叶阔叶林	0.999 615	0.999 968	0.998 767	0.998 409
混交林	0.999 588	0.999 960	0.999 789	0.998 318
多树荒原	0.999 628	0.999 944	0.995 770	0.992 367
荒原	0.999 589	0.999 989	0.997 257	0.984 418
农田	0.999 530	0.999 955	0.997 423	0.995 463
城市与建成区	0.999 539	0.999 990	0.999 848	0.999 446
贫瘠	0.999 259	0.999 998	0.997 031	0.995 050

表 5 2015 年数据占比法评价结果

精度/方案	方案 1	方案 2	方案 3	方案 4
评价标准	越大越好	越大越好	越大越好	越大越好
$ACC_{avg}$	0.899 651	0.899 971	0.897 902	0.887 168
$ACC_{weight}$	0.999 368	0.999 977	0.989 912	0.991 921
水体	0.999 248	0.999 984	0.986 488	0.994 847
常绿针叶林	0.999 805	0.999 972	0.997 090	0.998 622
落叶阔叶林	0.999 612	0.999 966	0.998 748	0.998 354
混交林	0.999 522	0.999 973	0.999 773	0.998 344
多树荒原	0.999 772	0.999 947	0.995 726	0.991 752
荒原	0.999 620	0.999 895	0.997 296	0.985 157
农田	0.999 460	0.999 938	0.997 327	0.995 396
城市与建成区	0.999 511	0.999 974	0.999 862	0.999 405
贫瘠	0.999 354	0.999 951	0.997 150	0.995 296



表 6 2018 年数据占比法评价结果

精度/方案	方案 1	方案 2	方案 3	方案 4
评价标准	越大越好	越大越好	越大越好	越大越好
$ACC_{avg}$	0.899 636	0.899 975	0.897 917	0.887 160
$ACC_{weight}$	0.999 340	0.999 972	0.989 932	0.991 943
水体	0.999 216	0.999 974	0.986 518	0.994 875
常绿针叶林	0.999 803	0.999 965	0.997 119	0.998 638
落叶阔叶林	0.999 603	0.999 969	0.998 832	0.998 437
混交林	0.999 527	0.999 999	0.999 747	0.998 395
多树荒原	0.999 682	0.999 983	0.995 822	0.991 829
荒原	0.999 605	0.999 929	0.997 293	0.984 843
农田	0.999 457	0.999 992	0.997 330	0.995 417
城市与建成区	0.999 516	0.999 979	0.999 869	0.999 410
贫瘠	0.999 327	0.999 918	0.997 238	0.995 321

表 7 2019 年数据占比法评价结果

精度/方案	方案 1	方案 2	方案 3	方案 4
评价标准	越大越好	越大越好	越大越好	越大越好
$ACC_{avg}$	0.899 637	0.899 959	0.897 916	0.887 100
$ACC_{weight}$	0.999 327	0.999 973	0.989 919	0.991 932
水体	0.999 193	0.999 993	0.986 490	0.994 897
常绿针叶林	0.999 807	0.999 964	0.997 177	0.998 666
落叶阔叶林	0.999 594	0.999 947	0.998 854	0.998 457
混交林	0.999 534	0.999 933	0.999 726	0.998 403
多树荒原	0.999 643	0.999 839	0.995 788	0.991 507
荒原	0.999 602	0.999 858	0.997 336	0.984 812
农田	0.999 455	0.999 948	0.997 319	0.995 383
城市与建成区	0.999 503	0.999 997	0.999 866	0.999 391
贫瘠	0.999 341	0.999 909	0.997 150	0.995 260

从占比法的评价结果来看,不论是总体精度,加权总体精度还是各个类别的单类精度,4 个方案的精度都很高,且数据比较接近。这是因为土地覆盖数据的数据量大,很难从比例变化上去评判不同升尺度方案的好坏,也不能反

映升尺度后数据的结构完整度。但总体而言,方案 2 随机抽样法的精度要高于其他 3 个方案,是因为其随机性,能够较好地保留与原数据相似的各类别所占比例。采用 ICS 指标的评价结果如表 8~11 所示。

表 8 2002 年数据类内 SSIM 方案评价结果

精度/方案	方案 1	方案 2	方案 3	方案 4
评价标准	越大越好	越大越好	越大越好	越大越好
$SSIM_{avg}$	0.451 016	0.427 187	0.456 382	0.507 173
$SSIM_{weight}$	0.827 328	0.827 062	0.837 831	0.842 048
水体	0.950 568	0.947 831	0.990 505	0.965 730
常绿针叶林	0.186 805	0.178 881	0.301 502	0.210 300
落叶阔叶林	0.258 173	0.248 611	0.282 848	0.282 200
混交林	0.270 843	0.282 136	0.265 565	0.334 724
多树荒原	0.269 917	0.268 825	0.233 049	0.233 437
荒原	0.402 746	0.364 256	0.340 653	0.221 135
农田	0.363 465	0.399 480	0.328 915	0.508 496
城市与建成区	0.395 366	0.311 421	0.289 136	0.435 376
贫瘠	0.616 523	0.576 796	0.537 145	0.819 426

表 9 2015 年数据类内 SSIM 方案评价结果

精度/方案	方案 1	方案 2	方案 3	方案 4
评价标准	越大越好	越大越好	越大越好	越大越好
$SSIM_{avg}$	0.447 923	0.422 359	0.465 173	0.533 295
$SSIM_{weight}$	0.833 943	0.833 539	0.840 624	0.847 981
水体	0.953 726	0.953 386	0.991 647	0.970 108
常绿针叶林	0.200 882	0.188 804	0.264 874	0.204 449
落叶阔叶林	0.288 433	0.235 178	0.260 746	0.293 049
混交林	0.269 236	0.273 814	0.277 617	0.308 661
多树荒原	0.277 144	0.256 096	0.225 007	0.211 851
荒原	0.358 958	0.411 203	0.338 614	0.221 423
农田	0.411 609	0.402 144	0.377 714	0.545 288
城市与建成区	0.360 570	0.286 861	0.410 273	0.585 305
贫瘠	0.683 706	0.644 338	0.493 620	0.828 923

表 10 2018 年数据类内 SSIM 方案评价结果

方案/精度	方案 1	方案 2	方案 3	方案 4
评价标准	越大越好	越大越好	越大越好	越大越好
$SSIM_{avg}$	0.439 655	0.421 148	0.461 661	0.513 921
$SSIM_{weight}$	0.825 738	0.828 038	0.838 713	0.845 968
水体	0.946 903	0.949 771	0.988 079	0.968 402
常绿针叶林	0.177 127	0.177 284	0.236 895	0.197 106
落叶阔叶林	0.235 860	0.253 893	0.287 308	0.308 944
混交林	0.256 055	0.267 430	0.258 899	0.334 942
多树荒原	0.258 571	0.269 640	0.241 128	0.227 942
荒原	0.418 550	0.343 142	0.332 904	0.224 904
农田	0.370 976	0.444 644	0.361 155	0.555 599
城市与建成区	0.381 492	0.280 978	0.262 708	0.405 498
贫瘠	0.647 387	0.643 795	0.548 610	0.831 917

表 11 2019 年数据类内 SSIM 方案评价结果

方案/精度	方案 1	方案 2	方案 3	方案 4
评价标准	越大越好	越大越好	越大越好	越大越好
$SSIM_{avg}$	0.442 901	0.419 843	0.454 277	0.515 077
$SSIM_{weight}$	0.829 935	0.826 212	0.835 348	0.839 150
水体	0.949 464	0.948 703	0.990 736	0.959 945
常绿针叶林	0.202 134	0.190 783	0.285 362	0.201 975
落叶阔叶林	0.237 700	0.246 855	0.327 756	0.310 769
混交林	0.255 247	0.285 895	0.295 290	0.317 214
多树荒原	0.256 875	0.275 880	0.242 974	0.214 080
荒原	0.385 748	0.394 803	0.320 586	0.217 568
农田	0.409 204	0.424 961	0.359 973	0.610 554
城市与建成区	0.330 022	0.255 807	0.235 720	0.407 387
贫瘠	0.595 570	0.574 058	0.524 761	0.776 214

从表 8~11 中的数据结果可以看出,非加权类内可分性精度指标的表现形式为:方案 4>方案 3>方案 1>方案 2。加权类内可分性精度指标的表现形式:方案 4>方案 3>方案 1>方案 2(“>”代表优于)。方案 4 采用的是 MODIS 官方判别规则进行升尺度,与其他方案升尺度结果相比,分类规则要多出“无数据或未分类”这一类别,而该类别在计算 ICS 时值相对比较大影响了总体精度,因此其精度指标都最高。如果不把官方方案计算在内,则方案 3 的精度指标都优于其他两种方案,方案 3 的语义邻近度能够较好的保证同一类别升尺度窗口的结构一致性,因此表现最好。且对于单类精度而言,方案 1 在前面的 6 类的精度上普遍排在后两位,从第 6 类开始精度开始排在前列。而方案 3 则是在前 6 类的精度普遍显著高于其他方案。因此可以看出,语义邻近度能够有效抑制众数聚合中优先级高的类别占据绝对优势的缺点。而方案 2 因为其随机性,不同类别在精度上的表现时好时坏。因此总体来看,方案 3 的非加权类内可分性精度指标和加权类内可分性精度指标都优于其他两种方案,且对优先级高的类别的升尺度效果最好。可见本文设计的语义邻近度法,能够很好地削弱众数聚合中中高优先级的类别更容易被保留下来的这种绝对优势。

### 3 结 论

土地覆盖类别数据升尺度方法中,众数聚合法存在众数情况下弱势类别的减少甚至消失,导致类别空间结构变化的问题;随机抽样法存在着会使空间结构趋于离散的问题。针对以上问题,本文设计了一种面向 IGBP 分类规则的语义层次图,并按照类别优先级,经过计算得到各类别对应的语义邻近度,应用在了升尺度算法中的语义邻近度法中。同时,提出了一种升尺度规则集合。在此基础上,因为缺少真实参考数据作为对比,设计了一种基于类内相似性的精度指标 ICS,以对升尺度后的结果进行有效的评价。实验证明,本文设计的语义邻近度方法和规则集合法在加权 ICS 中分别达到 84.1%、84.8%,高于众数聚合法和随机抽样法约 1%~2%,是一种有效且可靠的土地覆盖升尺度方法。但本文所提方法中的类别优先级及语义层次图的设计不是唯一的,在不同的领域中需要结合实际进行设计,在后续的研究中可进一步对类别优先级和语义层次图进行优化,以提高方法的普适性与稳定性。

#### 参考文献

- [1] 范胜龙,郭亚淑,邱凌婧,等.福州市主城区土地覆盖变化对城市生态脆弱性的影响分析[J].福建师范大学学报(自然科学版),2018,34(2):92-98,116.
- [2] 董金玮,匡文慧,刘纪远.遥感大数据支持下的全球土地覆盖连续动态监测[J].中国科学:地球科学,2018,48(2):259-260.
- [3] 王亚伟,廖成,张东民,等.基于谷歌地球的二维电波传播预测[J].电子测量技术,2018,41(6):40-44.
- [4] 杜婷,焦继宗,段含明,等.中国科学院土地利用分类系统与 IGBP 分类系统的转换研究——以西北农牧交错带为例[J].兰州大学学报(自然科学版),2020,56(1):90-95.
- [5] 刘远,周买春.3种 IGBP 分类系统的土地覆盖数据在韩江流域的对比分析[J].遥感技术与应用,2017,32(3):575-584.
- [6] LIN B B, EGERER M H, LIERE H, et al. Local and landscape-scale land cover affects microclimate and water use in urban gardens[J]. Science of the Total Environment, 2018(610-611): 570-575.
- [7] 盛佳慧,饶鹏.基于风云气象卫星的土壤湿度数据降尺度方法研究[J].红外与毫米波学报,2021,40(1):74-88.
- [8] ARCHILLA B G, NOVO J, TITI E S. Uniform in time error estimates for a finite element method applied to a downscaling data assimilation algorithm for the Navier-Stokes equations[J]. SIAM Journal on Numerical Analysis, 2018, 58(1): 410-429.
- [9] 谭诗腾,王继成,徐柱,等.基于优先级和语义邻近度的地表覆盖数据升尺度方法[J].国土资源遥感,2016,28(1):50-56.
- [10] SUN P J, CONGALTON R G, PAN Y Z. Improving the upscaling of land cover maps by fusing uncertainty and spatial structure information [ J ]. Photogrammetric Engineering & Remote Sensing, 2018, 84(2): 87-100.
- [11] 陈彦彤,李雨阳,陈伟楠,等.基于深度语义分割的遥感图像海面舰船检测研究[J].仪器仪表学报,2020,41(1):233-240.
- [12] 任立成,刘勇,张建林,等.基于特征融合的双分支模板更新跟踪算法[J].国外电子测量技术,2021,40(5):14-21.
- [13] 刘英,吴立新,岳辉.基于梯度结构相似度的矿区土壤湿度空间分析[J].武汉大学学报(信息科学版),2018,43(1):87-93.
- [14] 魏义康,金聪.结合局部与非局部结构的相似性与多样性判别投影在人脸识别上的应用[J].电子测量技术,2019,42(14):91-97.
- [15] 张增祥,汪潇,温庆可,等.土地资源遥感应用研究进展[J].遥感学报,2016,20(5):1243-1258.
- [16] 贾小斌,艾廷华,彭子凤.基于特征匹配的土地利用数据语义邻近度计算[J].资源科学,2017,39(3):408-417.
- [17] 苗立志,焦东来,杨立君.面向地理标记语言空间数据的地理信息聚合[J].计算机应用,2014,34(6):1816-1818,1824.
- [18] PUY G, TREMBLAY N, GRIBONVAL R, et al.

- Random sampling of bandlimited signals on graphs[J]. Applied and Computational Harmonic Analysis, 2018, 44(2): 446-475.
- [19] AZIZ A S, KALE R, KALEKAR M G, et al. Determining relative percentage change as a new metrics in scaling and root planning therapy outcome in patients with chronic periodontitis[J]. Asian Journal of Research in Biochemistry, 2021, 8(2): 36-43.

### 作者简介

陈施施, 硕士研究生, 主要研究方向为遥感影像处理。

E-mail: 20201235001@nuist.edu.cn

谢涛, 教授, 主要研究方向为海洋微波遥感。

E-mail: xictao@nuist.edu.cn

王超(通信作者), 副教授, 硕士生导师, 主要研究方向为高分辨率遥感影像处理。

E-mail: chaowang@nuist.edu.cn