

地学知识图谱引导的遥感影像语义分割

李彦胜¹, 武康¹, 欧阳松¹, 杨坤², 李和平³, 张永军¹

1. 武汉大学 遥感信息工程学院, 武汉 430079;

2. 贵州省基础地理信息中心, 贵阳 550004;

3. 贵州省第一测绘院, 贵阳 550025

摘要: 尽管深度语义分割网络有效提升了遥感影像语义分割性能, 但其效果远未达到人类领域专家的目视解译水平。原因是人类视觉系统在进行遥感影像解译时, 往往会综合运用视觉特征、语义信息和先验知识。然而, 深度语义分割网络本质上是数据驱动的面向像素级损失反向优化的分类方法。这种基于像素级优化的深度语义分割网络, 一方面受限于像素空间尺度, 缺乏整体性目标线索挖掘; 另一方面难以跨越结构化数据和非结构化知识之间的鸿沟, 无法充分利用地学先验知识和空间语义信息。针对以上两方面的问题, 本文提出了地学知识图谱引导的遥感影像深度语义分割方法, 运用从地学知识图谱中抽取的地物目标语义信息和地学先验知识构建实体级连通约束和实体间共生约束, 引导深度语义分割网络训练。其中, 实体级连通约束以连通域实体而非像素单元计算损失, 得到实体级别的特征表示, 使得分割结果更具整体性, 边界模糊和随机噪声现象得到抑制。实体间共生约束将共生条件概率量化的空间共生知识嵌入到数据驱动的深度语义分割网络中, 实现空间语义信息和地学先验知识对实体空间分布的约束引导和自动优化。验证结果表明, 在实体级连通约束和实体间共生约束的引导下, 深度语义分割网络可以完成对实体级特征的学习并根据空间共生知识自动优化地物实体的空间分布, 有效改善了遥感影像语义分割性能。

关键词: 地学知识图谱, 深度语义分割网络, 实体级连通约束, 空间共生知识约束, 地学知识嵌入优化

中图分类号: TP701/P2

引用格式: 李彦胜, 武康, 欧阳松, 杨坤, 李和平, 张永军. 2024. 地学知识图谱引导的遥感影像语义分割. 遥感学报, 28(2): 455-469

Li Y S, Wu K, Ouyang S, Yang K, Li H P and Zhang Y J. 2024. Geographic knowledge graph-guided remote sensing image semantic segmentation. National Remote Sensing Bulletin, 28 (2) : 455-469 [DOI: 10.11834/jrs.20231110]

1 引言

遥感影像语义分割作为地学信息解译的基础性工作, 广泛地应用于土地覆盖制图、自然环境保护、灾害应急监测、城市空间规划等领域, 具有重要的应用价值 (Ma 等, 2019)。遥感影像是地表三维世界经过大气传输到达传感器的映射, 包含了丰富的地物光谱、目标形状特征和地理空间关系等信息。人类领域专家在语义分割过程中, 往往需要综合考虑目标语义信息和地学先验知识才能够有效克服遥感影像的“同谱异物、同物异谱”问题, 从而取得理想的分割结果 (Zhu 等, 2017)。

传统的图像语义分割方法包括最大似然法 (MLE)、随机森林 (RF)、决策树 (DT)、支持向量机 (SVM) 等基于浅层特征判别的监督分类方法 (Camps-Valls 等, 2014; 李楠 等, 2018)。这类方法一般先根据人工设计的特征描述子提取图像的光谱、纹理和几何结构特征, 然后进行特征分类 (肖春姣 等, 2020)。传统分割方法高度依赖于人工设计特征, 难以跨越底层图像数据与高层逻辑信息的语义鸿沟, 其鲁棒性和精度较差。随着人工智能理论与技术的快速发展, 深度学习方法广泛运用到了遥感图像处理任务中 (Li 等, 2018, 2020, 2021a)。基于深度学习的语义分割方法包

收稿日期: 2021-03-09; 预印本: 2023-08-24

基金项目: 国家自然科学基金(编号: 42030102, 41971284, 42371321)

第一作者简介: 李彦胜, 研究方向为耦合知识图谱和深度学习的遥感大数据智能解译。E-mail: yansheng.li@whu.edu.cn

通信作者简介: 李和平, 研究方向为遥感影像解译及应用。E-mail: 1596686651@qq.com

括全卷积网络 (FCN)、U型网络 (U-Net)、分割网络 (SegNet)、掩模区域卷积网络 (Mask R-CNN)、深度分割网络 (DeepLab)、分割 Transformer (SegFormer)、多视野融合网络 (MFV-Net)、PID网络 (PIDNet) 等深度语义分割网络方法 (Long 等, 2015; Ronneberger 等, 2015; Badrinarayanan 等, 2017; Zhao 等, 2017; He 等, 2018; Chen 等, 2018; Xie 等, 2021; Li 等, 2023; Xu 等, 2023)。深度语义分割网络通过端到端的学习机制对特征提取与特征分类一体化模型自动学习, 从而自适应完成分割工作, 使得分割的准确度大幅提高、分割的过程也更加智能化。然而, 深度学习是基于像素的数据驱动方法, 通过降低在每个像素上的损失来反向优化网络模型, 缺乏实体级别的学习, 不能有效提取出目标形状特征, 使得分割结果整体性缺失、边界模糊和随机噪声分布明显, 同时受制于结构化数据驱动方法的缺陷, 往往难以利用地学先验知识和实体间丰富的语义信息 (如空间关系) (Liu 等, 2020), 导致可解释性差。以上两点不足严重制约着深度语义分割网络性能, 亟需在分割过程中从实体尺度出发, 综合考虑地学先验知识和实体间语义信息。

先验知识和语义信息是对规则或事实的抽象化表达, 难以形式化建模。为充分利用先验知识, 国内外专家学者探索了诸多方法, 如辅助通道嵌入、物理模型建模、迁移学习等。辅助通道嵌入方法 (Wu 等, 2021) 通过将领域知识作为辅助的输入通道参与网络训练测试从而进行知识嵌入。物理模型建模方法 (Xu 等, 2022; Li 等, 2022) 将物理知识融入损失函数和模型结构中来利用先验知识。迁移学习方法 (Dash 等, 2022; 李发森等, 2022; 欧阳淑冰等, 2022) 通过预训练等方式将领域先验知识耦合网络模型中。然而这些方法大多只考虑到特定场景下的先验知识的入, 对语义信息的利用程度不够。在此背景下, 遥感科学应该得到本体论、知识图谱等知识表示技术的支持。本体 (Ontology) (Arvor 等, 2019) 作为对特定领域中概念及其相互关系的形式化表达, 具有很强的知识表示能力、基于认知语义学的推理能力和共享知识的能力。地学知识图谱作为语义网络, 描述了地物目标的属性以及目标之间的关系, 相较于其他建模方式, 能够更好的捕捉实体之间的语义关联和复杂的关系模式, 有助于更好

地理解和推理知识, 适用于地学知识的结构化表达。其中, 地学本体是地学知识图谱的骨架, 地物目标作为本体的实例化对象组成了地学知识图谱的基本单元。地学知识图谱的发展经历了专家智能解译系统 (Goodenough 等, 1987)、地学信息图谱 (张洪岩 等, 2020)、地理知识图谱 (Hogan 等, 2022) 3个阶段。传统的专家智能解译系统通过知识的规则化模仿专家的决策过程, 针对性强, 但适用范围小, 解译精度有限; 地学信息图谱是一种借鉴图谱思想构建的地球信息科学理论, 系统化和抽象化地表达了地学知识, 但由于面向的是整个地学领域, 并不能直接用于遥感影像解译; 地理知识图谱则是将当前的知识图谱理论引入到了地理科学领域, 完成了地学知识的结构化表达和推理。现有遥感影像解译方法缺乏地学知识的嵌入, 导致解译的可解释性和可靠性受限, 这使得地学知识图谱驱动下的遥感影像智能解译具有广阔的应用前景 (王志华 等, 2021; 李彦胜和张永军, 2022; 张永军 等, 2023)。通过引入地学知识图谱, 从中抽取符号化的地学先验知识和语义信息, 并借助知识推理以完成遥感影像解译, 从而提高解译结果的准确度和可解释性。在已有工作中, 建筑物本体模型 (Gui 等, 2016) 用于从 SAR 影像中提取建筑物, 地学知识图谱推理 (Gu 等, 2017) 被用于基于目标的高分辨率遥感影像语义分类方法, 旨在挖掘利用地学知识图谱推理理论对遥感影像解译的优势。地学知识图谱嵌入方法 (吴瑞 等, 2022) 被用于高光谱解混领域, 通过先验知识来进一步提高端元选择的可靠性, 从而提升解混的精度。地学知识图谱推理增强了分类结果的可解释性和可信度, 但是相比于深度学习方法, 其分类精度较差 (Li 等, 2021b)。联合深度学习与知识推理是协调数据驱动方法与知识驱动方法的重要途径 (Arvor 等, 2019; Li 等, 2022)。在该类工作中, Alirezaie 等 (2019) 实现了本体推理器与深度神经网络分类器在输入和输出端的交互。耦合深度语义分割网络和图卷积神经网络的遥感影像语义分割方法 DSSN-GCN (Ouyang 和 Li, 2021) 借助图卷积神经网络对节点依赖关系的强大建模能力, 在深度语义分割网络的特征提取的基础上引入了地物目标空间拓扑关系, 从而预先将空间拓扑先验知识嵌入到了网络中, 有效提高了网络的性能。虽然遥感领域已有研究引入知

识来推动遥感影像解译技术(范菁等, 2017), 但仍没有具体工作来探讨如何将地学知识图谱来引导优化深度网络。鉴于结构化数据和非结构化知识之间的鸿沟, 如何将地学知识嵌入深度语义分割网络中以自主引导网络训练仍极具挑战性。

基于上述分析, 为在深度语义分割网络分割过程中实现实体级特征自主学习以及充分利用空间语义信息与地学先验知识, 本文提出了地学知识图谱引导的遥感影像深度语义分割方法, 使用从地学知识图谱中抽取得到的地物目标语义信息和地学先验知识来构建实体级连通约束和实体间共生约束, 从而自主引导深度语义分割网络训练。实体级连通约束通过对分割结果的实体级约束, 保证了分割结果的整体性。实体间共生约束通过量化共生先验知识, 实现将非结构化的知识嵌入到数据驱动的深度语义分割网络中。验证结果表明, 本文提出的地学知识图谱引导的深度语义分割方法明显优于已有深度语义分割方法。

2 理论与方法

本文提出方法的总体流程图如图1。主要包括地学知识图谱构建与遥感解译先验知识提取模块、地学知识图谱引导的损失约束模块以及深度语义分割网络优化模块。地学知识图谱构建与遥感解译先验知识提取模块首先利用地学本体定义抽象类和属性关系, 再从数据集标签影像中抽取实体及其属性以实例化本体类, 从而完成地学知识图谱构建, 最后从图谱中提取空间共生先验知识。地学知识图谱引导的损失约束模块包括常规的像素级稠密约束、实体级连通约束和实体间共生约束。数据集中的标签包含大量的连通域, 这些连通域是地物目标的分割结果, 连通域之间的空间共生分布是领域分类知识的体现。空间共生等地学先验知识可以根据连通域的空间分布进行抽取。实体级连通约束和实体间共生约束均以连通域为处理单元, 前者计算每一个连通域单元而不是像素的损失, 以实现对实体的整体性约束; 后者借助空间共生先验知识完成邻域实体对中心实体的打分, 该分值代表中心实体在当前邻域空间分布下所属类别的置信度, 再根据分值计算损失, 从而实现对实体空间分布的约束。深度语义分割网络优化模块负责网络的训练和图像的语义分割。本文提出方法是一种地学知识图谱引导的

遥感影像深度语义分割方法, 深度语义分割网络优化模块作为其中一部分, 并不局限于特定的结构, 可基于不同的深度语义分割网络构建。为了验证方法的通用性, 本研究将地学知识图谱引导的深度语义分割方法作用在U-Net和DeepLab V3+这2种常见的网络中。深度语义分割网络通过优化加入约束的综合损失来调整网络模型, 从而学习到实体级特征表示和利用空间共生知识引导分割。

2.1 地学知识图谱构建及遥感解译先验知识提取

2.1.1 地学知识图谱构建

地学知识图谱建立在地理学本体的基础上, 通过实例化本体完成构建。地理学本体的引入可将地学知识符号化, 有助于提高知识的自主化运用水平。为建立地理学本体, 本文使用网络本体语言OWL(Web Ontology Language)描述本体。地理学本体的层次结构如图2所示。可见地物对象类(rs: GeoObject)为根类, 其余子类从中衍生: 一级子类主要包括水体(rs: Water)、植被(rs: Vegetation)、透水地面(rs: Ground)、农业用地(rs: Agricultural land)、城镇用地(rs: Urban land)、交通工具(rs: Vehicle)和其他(rs: Unknown); 二级子类在一级子类的基础上细分为湖泊(rs: Lake)、河流(rs: River)、海洋(rs: Sea)、草地(rs: Grass)、树木(rs: Tree)、裸地(rs: Bare land)、荒地(rs: Wasteland)、耕地(rs: Farmland)、牧场(rs: Rangeland)、建筑物(rs: Building)、道路(rs: Pavement)、车辆(rs: Car)、船只(rs: Ship)、飞机(rs: Airplane)。属性关系是本体类间或实体间联系的语义桥梁, 地理学本体的核心属性主要包括从属(oc: isA)等上下层属性, 相邻(geo: adjacentTo)、环绕(geo: surroundedBy)、方位(geo: hasDirectionOf)等空间关系属性以及多数类(geo: MaxClass)等统计属性。图2中上下层属性描述了实体间的从属关系, 比如:“耕地”是“农业用地”, 因此“耕地”和“农业用地”之间就包含上下层属性; 空间关系属性描述了实体间的空间关系, “道路”环绕“车辆”, 因此“车辆”和“道路”之间就包含环绕的空间关系属性; 统计属性描述实体间的统计性质和特征, 如某个类别在知识图谱中出现的频

率等统计属性。地学本体是地学知识图谱的骨架，
而实体作为本体类的实例化对象，组成了地学知

识图谱的基本单元（图2）。

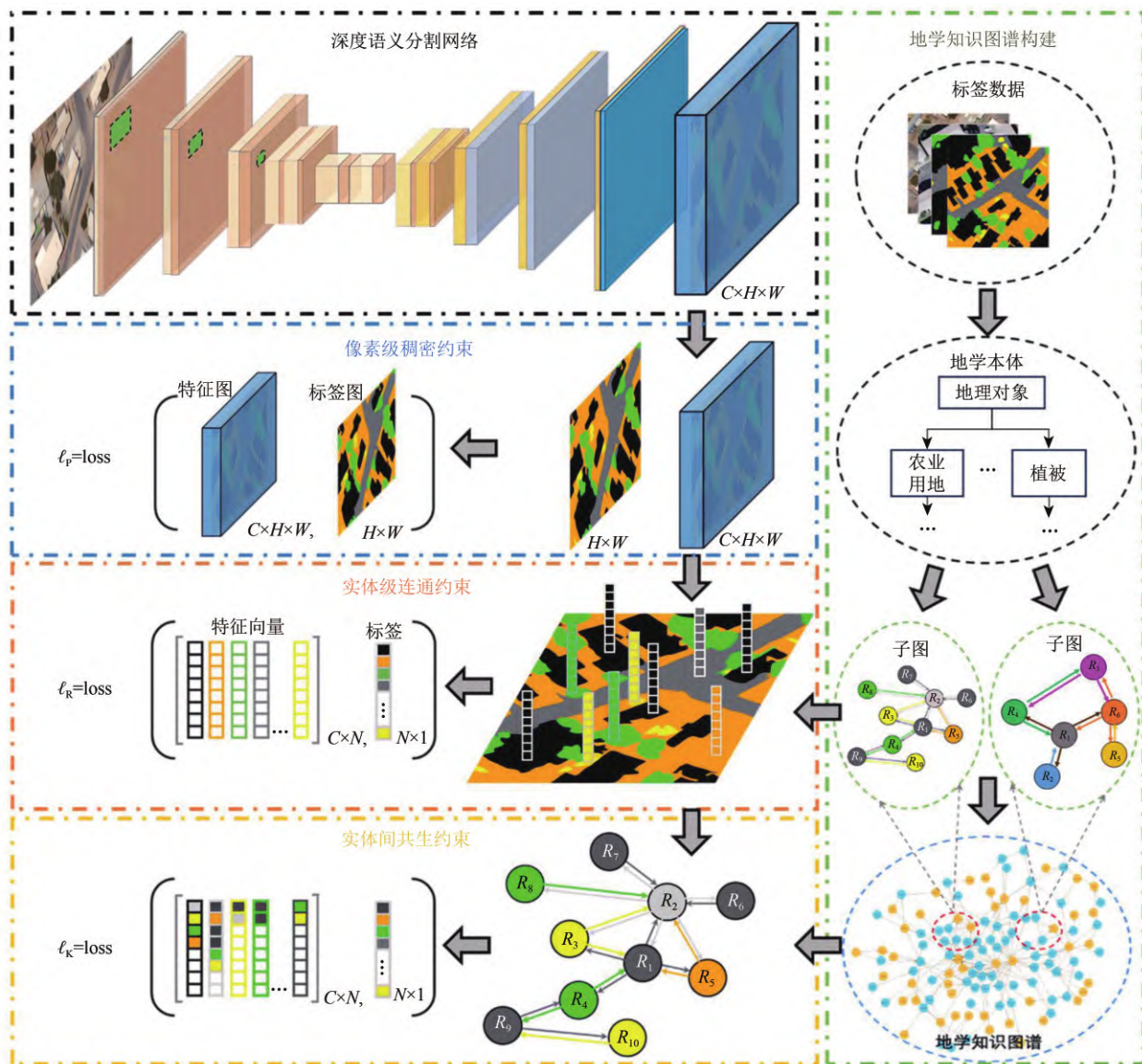


图1 本文提出方法的总体流程图
Fig. 1 The workflow of the proposed method

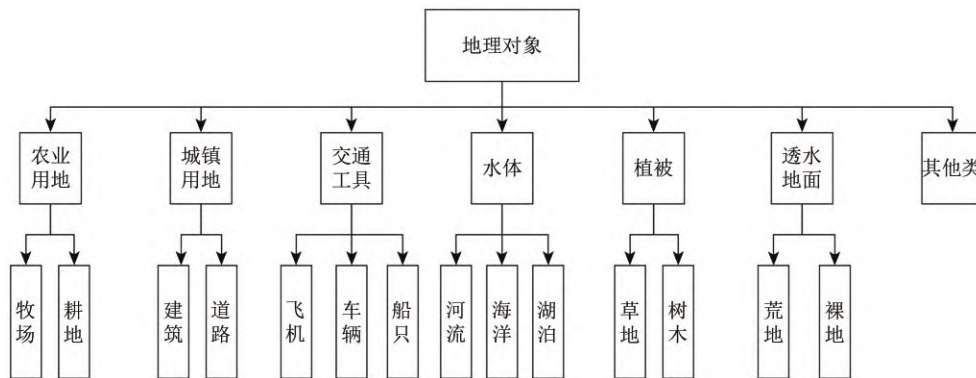


图2 地学本体层次结构
Fig. 2 The hierarchy of the geographic ontology

2.1.2 遥感解译先验知识提取

数据集的标签影像包含大量的地物目标信息，包括类别、空间分布等信息。地物目标的空间分布是地学先验知识的体现，因此可以从标签影像上地物目标的空间分布中抽取空间共生关系等先验知识。一般来说，超像素可以作为目标单元，因为受超像素分割算法的约束，影像中每个超像素是由一系列同质的像素组成的，同时超像素具有一致的形状大小，意味着形状较大的地物目标可以分割出更多的超像素，那么在获取目标单元的条件共生概率等统计属性时能够将地物目标的尺寸也一并考虑进来。鉴于地物目标尺寸信息的重要性，本文以原始影像超像素分割块对应于标签影像的区域作为本体类的实体。利用标签影像中的超像素（实体）实例化本体类，超像素中具有多数像素的地物类别作为对应实体的本体类类别，也即实体的多数类 geo: MaxClass 属性。超像素的空间关系属性和统计属性作为实体的属性。若实体 obj_1 和 obj_2 在空间上存在公共边，则两者具有相邻属性，以三元组 $(\text{obj}_1, \text{geo: adjacentTo}, \text{obj}_2)$ 表示。同理可得其他属性关系。本体类的空间共生属性是一种地学先验知识，可由统计概率表达，具体做法是在邻域内统计不同本体类实体出现的条件概率。以本体类 C_i 、 C_j 为例，统计地学知识图谱中所属 C_i 的实体出现的概率 $P(C_i)$ ，以及出现与该实体相邻的类别为 C_j 的实体的概率 $P(C_i, C_j)$ ，再根据下式即可计算出在本体类 C_i 实体出现的条件下邻域内出现本体类 C_j 实体的概率 $P(C_j|C_i)$ ，称 P 为共生条件概率。

$$P(C_j|C_i) = \frac{P(C_i, C_j)}{P(C_i)} \quad (1)$$

以类别 C_{building} 建筑类为例，首先统计出地学知识图谱中出现建筑类的实体的概率 $P(C_{\text{building}})$ 以及邻域内同时出现类别为建筑类和 C_{pavement} 道路类实体的概率 $P(C_{\text{building}}, C_{\text{pavement}})$ ，再根据式 (1) 即可计算出在建筑类实体出现的条件下邻域内出现道路类实体的概率 $P(C_{\text{pavement}}|C_{\text{building}})$ 。

2.2 地学知识图谱引导的深度语义分割网络优化目标函数

从遥感领域知识图谱中抽取目标实体关系知识用于自主引导深度语义分割网络训练。引导方

法主要包括实体级连通约束和实体间共生约束，约束构建以标签连通域为实体。实体级连通约束以实体而不是像素为单元计算损失，实现对实体的整体性约束。实体间共生约束通过实体间空间共生知识将非结构化的地学先验知识嵌入到数据驱动的神经网络中以约束分割目标的空间分布。深度语义分割网络通过优化总体目标函数来完成遥感领域知识图谱的嵌入。深度语义分割网络的总体优化目标函数 \mathcal{L} 包括像素级稠密约束的常规损失项 ℓ_p 、实体级连通约束损失项 ℓ_r 和实体间共生约束损失项 ℓ_k ，具体计算公式如下：

$$\mathcal{L} = \ell_p + \alpha \cdot \ell_r + \beta \cdot \ell_k \quad (2)$$

式中， α 和 β 为常数，用于调节 ℓ_r 和 ℓ_k 在总体损失中所占的比例。

2.2.1 像素级稠密约束

深度语义分割网络输出的分类置信度图 $F \in \mathbb{R}^{C \times H \times W}$ 及其标签影像 $Y \in \mathbb{R}^{H \times W}$ 计算像素级稠密损失，像素级稠密约束的常规损失项 ℓ_p 定义为

$$\ell_p = \text{loss}(F, Y) \quad (3)$$

式中， $F = \phi(I, W_0)$ ， $\phi(\cdot)$ 既代表深度语义分割网络的层次化映射函数，也代表深度语义分割网络模型， I 为输入图像， W_0 为深度语义分割网络的参数。 C 、 H 和 W 分别为图像的类别数、高度和宽度。

2.2.2 实体级连通约束

图像原始数据 I 输入深度语义分割网络，输出分类置信度图 $F \in \mathbb{R}^{C \times H \times W}$ ，对 F 按通道取最大值的序号即得到分割结果。实体级连通约束损失 ℓ_r 以标签中连通域为单元进行计算，首先在分类置信度图 F 上对 $S_i (1 \leq i \leq N)$ 实体区域内所有像素 $P_j (P_j \in S_i)$ 的分类置信度向量 D 按通道计算均值，得到具有通道维数 C 的向量，这个向量作为该实体的分类置信度向量 \hat{Y}_i ；然后根据实体的分类置信度向量 \hat{Y}_i 和真实类别 Y_i 计算损失；最后取所有实体损失项的均值，该均值即为实体级连通约束损失 ℓ_r 。具体计算公式如下：

$$\ell_r = \frac{1}{N} \sum_i \text{loss}(\hat{Y}_i, Y_i) \quad (4)$$

式中， $\hat{Y}_i = \frac{1}{N_p} \sum_{p_j=1}^{N_p} D_j$ ， N_p 为实体 S_i 范围内的总像素数， $1 \leq i < N$ 。

2.2.3 实体间共生知识约束

与实体级连通约束损失构建一样，实体间共生约束损失 ℓ_k 以标签中连通域为实体单元进行计算。首先确定每个实体的分类类别，取实体 S_i ($1 \leq i \leq N$) 的分类置信度向量 \hat{Y}_i ，将向量 \hat{Y}_i 中最大值及其最大值序号分别作为该实体的分类置信度 \hat{y}_i 和分类类别 k ($1 \leq k \leq C$)；其次给实体 S_i 打分，取中心实体 S_i 邻域内的所有实体 $\{S_j | S_j \text{ Adjacent to } S_i\}$ 对 S_i 打分（邻域内共 m_i 个实体）， S_i 的分值由邻域内实体的分类置信度向量 \hat{Y}_j 的最大值 \hat{y}_j 和共生条件概率 P 按类别加权求和而来，得到具有通道维数的分值向量 $H_i \in \mathbb{R}^C$ ，该向量代表中心实体在当前邻域空间分布下所属各类的分类置信度；最后根据实体的基于空间分布的分类置信度向量 H_i 和真实类别 Y_i 计算损失，取所有实体损失项的均值作为实体间共生约束损失 ℓ_k 。具体计算公式如下：

$$\ell_k = \frac{1}{N} \sum_i \text{loss}(H_i, Y_i) \quad (5)$$

式中， $H_i = \frac{1}{m_i} \sum_q \sum_j \hat{y}_j \cdot P(C_q | C_k)$ ， $\hat{y}_j = \max(\hat{Y}_j)$ ， $k = \text{argmax}(\hat{Y}_i)$ ， m_i 是指中心实体 S_i 的相邻实体个数， q 代指类别， $1 \leq q \leq C$ ， $P(C_q | C_k)$ 表示条件概率。

实体间共生知识约束和实体级连通约束的关系如图 1 所示。二者都以连通域单元作为基本计算单位且损失计算都依赖实体的分类置信度向量，但损失计算方式、约束角度有明显区别。其中，连通域单元是标签中各类别实体相连通的像素区域，实体的分类置信度向量是通过计算每个连通域单元的预测结果的逐通道均值得到。如图 1 第 3 行所示，实体级连通约束计算实体的分类置信度向量和真实标签之间的损失，从实体层级对网络进行约束，使网络自主学习实体级别的特征表示。如图 1 第四行所示，实体间共生知识约束基于多个邻域实体的分类置信度向量和地学知识图谱中的共生条件概率得到空间分布的分类置信度向量，再计算空间分布的分类置信度向量和真实标签间损失，从领域知识的角度约束深度网络训练。实体间共生知识约束将非结构化的知识嵌入到数据驱动的深度语义分割网络中，从而自动优化分割结果的空间分布。

2.3 深度语义分割网络的优化方法

深度语义分割网络输出的分类置信度图 F 及其标签影像 $Y \in \mathbb{R}^{h \times w}$ 计算总体损失，通过后向传播算法降低总体损失 \mathcal{L} ，从而优化深度语义分割网络。基于实体级连通约束的损失 ℓ_r 引导深度语义分割网络在训练的过程中自主学习实体级别的特征表示，使得网络输出的分割结果更具整体性，边界模糊和随机噪声现象得到抑制。基于实体间共生约束的损失 ℓ_k 根据实体间的空间共生知识调整目标的空间分布，从而实现分割实体的空间分布的自动优化。

3 数据处理与结果分析

3.1 实验数据集

本文基于 UCM 遥感数据集 (Shao 等, 2018) 和 DeepGlobe 遥感数据集 (Demir 等, 2018)。UCM 数据集包含 21 个地物类别，每个类别有 100 张遥感图像，每张图像尺寸为 256×256 像素，地面分辨率为 0.3 m。样本集采用密集标注的 DLRSD 数据集 (Shao 等, 2018)，共包含 17 类。参照 (Alirezaie 等, 2019) 的做法，为了缩小类别之间的相似性，本文将样本集中的 17 类合并成了 8 类，分别是植被 (Tree, Grass)、裸地 (Bare soil, Sand, Chaparral)、道路 (Pavement, Dock)、建筑 (Building, Mobile home, Tank)、水体 (Water, Sea)、飞机 (Airplane)、车辆 (Cars) 和船只 (Ship)。每个类别由括号里面的原类别合并而成。为了减低类别之间的相似性，移除了包含 Field 或 Tennis court 类的图像。将这些筛选出来的图像按 8 : 1 : 1 的比例随机划分出训练集、验证集和测试集，分别包含 1513、189 和 190 张图像。

DeepGlobe 土地覆盖分类数据集提供了 1146 幅亚米分辨率的遥感影像，图像尺寸为 2448×2448 像素。人工标注了 7 类，分别是城镇、耕地、牧场、森林、水体、裸地和未知类。整个数据集分为训练集、验证集和测试集，分别包含 803、171 和 172 幅图像。从每个原始图像中均匀地裁剪大小为 256×256 像素的图像。将裁剪后的图像按 8 : 1 : 1 的比例随机划分出训练集、验证集和测试集，分别包含 10272、1280 和 1296 张图像。

UCM 遥感数据集中，训练所使用的地学知识图谱各类别共生条件概率如表 1 所示；DeepGlobe

数据集中各类别共生条件概率如表2所示。表中每一行表示地学知识图谱中的类别 C_i ，每一列表示与 C_i 不同的类别 C_j ；表中的具体数值表示共生条件概率，即本体类 C_i 实体出现的条件下邻域内出现本体类 C_j 实体的概率 $P(C_j|C_i)$ 。

表1 UCM数据集各类别共生条件概率

Table 1 Symbiosis conditional probability of each category in UCM dataset

类别	植被	裸地	道路	建筑	水体	飞机	汽车	船只
植被		0.4840	0.7274	0.5369	0.0783	0.0100	0.5132	0.0027
裸地	0.5986		0.5332	0.3866	0.1747	0.0000	0.3235	0.0022
道路	0.7150	0.4238		0.5241	0.0905	0.0716	0.6129	0.0707
建筑	0.8511	0.4956	0.8453		0.0158	0.0346	0.6184	0.0043
水体	0.3385	0.6102	0.3976	0.0433		0.0000	0.0354	0.3110
飞机	0.1375	0.0000	1.0000	0.3000	0.0000		0.6625	0.0000
汽车	0.7929	0.4042	0.9633	0.6028	0.0126	0.0746		0.0014
船只	0.0375	0.0250	0.9875	0.0375	0.9875	0.0000	0.0125	

表2 DeepGlobe数据集各类别共生条件概率

Table 2 Symbiosis conditional probability of each category in DeepGlobe dataset

类别	城镇	耕地	牧场	森林	水体	裸地
城镇		0.8921	0.6625	0.2049	0.6040	0.5161
耕地	0.8019		0.6551	0.2285	0.5706	0.5027
牧场	0.8237	0.9061		0.2911	0.6570	0.5498
森林	0.6855	0.8505	0.7835		0.4896	0.5463
水体	0.8200	0.8619	0.7175	0.1987		0.5774
裸地	0.7827	0.8481	0.6705	0.2476	0.6448	

3.2 实验设置及评估指标

在实验测试的过程中，本文采用了经典的U-Net (Ronneberger等, 2015) 和先进的DeepLab V3+ (Chen等, 2018) 深度语义分割网络。U-Net是一种用于图像分割任务的卷积神经网络，以其编码—解码结构和跳跃连接的设计而闻名，能够有效地处理医学图像和自然图像等领域的分割问题。其采用了一种类似于“U”形的结构，将输入图像分成两个部分，分别进行卷积和池化操作，然后将它们合并起来得到最终的输出结果。这种结构可以有效地捕捉图像中的细节信息，从而提高了分割的精度。DeepLabv3+是一种用于图像语义分割的深度神经网络模型。它结合了全卷积网络和

空洞卷积，以实现高效准确的图像分割。模型采用编码器—解码器结构，使用预训练的卷积神经网络提取特征，并通过转置卷积将特征映射上采样到原始图像尺寸。空洞卷积和自适应空洞扩张模块允许模型在不增加参数的情况下捕捉多尺度上下文信息。多尺度推理和空间金字塔池化进一步提高分割性能。DeepLabv3+在图像分割任务中取得了优秀的性能。对于网络的训练，分别采用随机梯度下降法 (SGD) 和交叉熵 (Cross Entropy) 作为优化器和损失函数。在深度学习领域中，当数据规模等条件差距较大时，为了保证网络的收敛，通常使用不同的初始学习率进行网络训练 (He等, 2016)。本文中2个数据集的数据规模和分辨率等条件差异较大，因此，在UCM数据集和DeepGlobe数据集上的学习率设置不同，分别为 3×10^{-4} 和 2×10^{-4} 。另外，超像素分割方法采用了简单线性迭代聚类SLIC (Simple Linear Iterative Cluster) (Achanta等, 2012)。所有的实验均在PyTorch框架下使用一块NVIDIA 1080Ti GPU进行的。

语义分割结果的评价指标采用总体精度OA (Overall Accuracy)、交并比IoU (Intersection over Union)、均交并比 (Mean Intersection over Union, MIoU) 和频权交并比FWIoU (Frequency Weighted Intersection over Union)。上述指标计算公式如下：

$$OA = \frac{TP + TN}{TP + FP + TN + FN} \quad (6)$$

$$IoU_i = \frac{TP_i}{TP_i + FP_i + FN_i}, \quad i = 1, 2, \dots, n \quad (7)$$

$$MIoU = \frac{1}{n} \sum_1^n IoU_i \quad (8)$$

$$FWIoU = \left(IoU_i \cdot \frac{TP_i + FN_i}{TP_i + FP_i + TN_i + FN_i} \right) \quad (9)$$

式中， n 为类别的数量，TP、TN、FP和FN分别为正类被正确判别的像素数、正类被错误判别的像素数、负类被正确判别的像素数和负类被错误判别的像素数。

3.3 超参数的敏感性分析

超参数主要包括式(2)中的 α 和 β ， α 用于调节实体级连通约束损失项 ℓ_r 在总体损失 \mathcal{L} 中的占比， β 用于调节实体间共生约束损失项 ℓ_k 在总体损失 \mathcal{L} 中的占比。由于 ℓ_r 和 ℓ_k 对于总体损失 \mathcal{L} 贡献比例并无法直接通过理论得到，因此我们将其设

置为超参数，通过实验来探究合适的比例。在 UCM 数据集和 DeepGlobe 数据集上，本文使用 DeepLab V3+ (Chen 等, 2018) 作为基础网络进行超参数的敏感性分析实验。

3.3.1 UCM 数据集上超参数敏感性分析

UCM 验证集上 α 的敏感性分析结果见表 3。可见：当 $\beta = 0.5$ 时，随着 α 增加，基础网络在验证集上的分割精度先上升后下降，当 $\alpha = 0.5$ 时分割精度到达顶点。在最佳 $\alpha = 0.5$ 取值条件下， β 的敏感性分析结果见表 4。可见基础网络在验证集上的分割精度随着 α 变化，当 $\beta = 0.5$ ，得到最佳精度。

表 3 在 UCM 验证集上 α 的敏感性分析 ($\beta = 0.5$)

Table 3 The sensitivity analysis of α on the validation set of the UCM dataset ($\beta = 0.5$)

指标	$\alpha = 0$	$\alpha = 0.25$	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 1.0$
OA	88.36	88.47	88.50	88.30	88.37
MIoU	76.45	76.66	76.83	76.85	76.60
FWIoU	79.30	79.47	79.49	79.22	79.30

注：粗体数字表示几组设置中精度最优设置。

表 4 在 UCM 验证集上 β 的敏感性分析 ($\alpha = 0.5$)

Table 4 The sensitivity analysis of β on the validation set of the UCM dataset ($\alpha = 0.5$)

指标	$\beta = 0$	$\beta = 0.25$	$\beta = 0.5$	$\beta = 0.75$	$\beta = 1.0$
OA	88.41	88.34	88.50	88.39	88.41
MIoU	76.61	76.72	76.83	76.66	76.71
FWIoU	79.33	79.24	79.49	79.32	79.34

注：粗体数字表示几组设置中精度最优设置。

3.3.2 DeepGlobe 数据集上超参数敏感性分析

表 5 为 DeepGlobe 验证集上 α 的敏感性分析结果。可见：当 $\beta = 0.5$ 时，基础网络在验证集上的分割精度随着 α 变化，当 $\alpha = 0.5$ ，得到最佳精度。在最佳 $\alpha = 0.5$ 取值条件下， β 的敏感性结果见表 6，可见随着 β 增加，基础网络在验证集上的分割精度先上升后下降，当 $\beta = 0.5$ 时分割精度到达最佳。

3.4 消融实验

在 UCM 数据集和 DeepGlobe 数据集上，本文使用 U-Net (Ronneberger 等, 2015) 和 DeepLab V3+ (Chen 等, 2018) 作为基础网络进行消融实验。

表 5 在 DeepGlobe 验证集上 α 的敏感性分析 ($\beta = 0.5$)

Table 5 The sensitivity analysis of α on the validation set of the DeepGlobe dataset ($\beta = 0.5$)

指标	$\alpha = 0.25$	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 1.0$	$\alpha = 1.25$
OA	87.37	87.50	87.18	87.24	87.24
MIoU	59.42	59.24	58.95	58.97	58.71
FWIoU	78.90	78.96	78.58	78.67	78.56

注：粗体数字表示几组设置中精度最优设置。

表 6 在 DeepGlobe 验证集上 β 的敏感性分析 ($\alpha = 0.5$)

Table 6 Sensitivity analysis of β on the validation set of the DeepGlobe dataset ($\alpha = 0.5$)

指标	$\beta = 0$	$\beta = 0.25$	$\beta = 0.5$	$\beta = 0.75$	$\beta = 1.0$
OA	87.20	87.34	87.50	87.26	87.40
MIoU	58.96	58.74	59.24	59.08	58.93
FWIoU	78.00	78.50	78.96	78.79	78.82

注：粗体数字表示几组设置中精度最优设置。

3.4.1 UCM 数据集上的结果

本文方法在 UCM 测试集上的语义分割精度如表 7 所示。可见：相比于基于像素级稠密约束的常规损失 (ℓ_p) 训练的语义分割网络，加入实体级连通约束 ($\ell_p + \ell_r$) 训练的网络的分割精度 OA、MIoU 和 FWIoU 均更高，尤其是在 MIoU 上提升明显，提升幅度达 9%，这说明实体级学习能够有效提高深度语义分割网络的性能；综合实体级连通约束与实体间共生约束 ($\ell_p + \ell_r + \ell_k$) 的深度语义分割网络的精度优于只加入实体级连通约束 ($\ell_p + \ell_r$) 的网络，这体现了空间共生知识对于遥感影像语义分割的重要性。

表 7 在 UCM 测试集上的语义分割精度

Table 7 The accuracy of semantic segmentation on the test set of the UCM dataset

基础网络框架	损失约束项	OA	MIoU	FWIoU
U-Net	$\mathcal{L} = \ell_p$	81.45	61.26	68.76
	$\mathcal{L} = \ell_p + \ell_r$	83.36	70.40	71.56
	$\mathcal{L} = \ell_p + \ell_r + \ell_k$	83.47	70.72	71.76
DeepLab v3+	$\mathcal{L} = \ell_p$	83.67	66.68	72.33
	$\mathcal{L} = \ell_p + \ell_r$	85.80	75.00	75.26
	$\mathcal{L} = \ell_p + \ell_r + \ell_k$	85.95	75.13	75.49

注：粗体数字表示几组设置中精度最优设置。

深度语义分割网络在 UCM 测试集上的分割结果如图 3 所示。可见：实体级连通约束 ($\ell_p + \ell_r$)

分割结果和综合实体级连通约束与实体间共生约束 ($\ell_p + \ell_R + \ell_K$) 的分割结果均明显优于常规的像素级稠密约束 (ℓ_p) 的分割结果; 对于飞机、车辆等小目标场景, 施加实体级连通约束能获取轮廓更加清晰的分割结果, 对于建筑物和油罐等规则地物, 施加实体级连通约束的分割结果整体性

更佳, 噪声较少同时轮廓更加清晰, 这表明约束项使得网络学习到了实体级的特征表示。另外, 综合实体级连通约束与实体间共生约束 ($\ell_p + \ell_R + \ell_K$) 的分割结果中地物实体空间分布更加接近真实世界中的空间布局, 如建筑物邻近道路与植被。

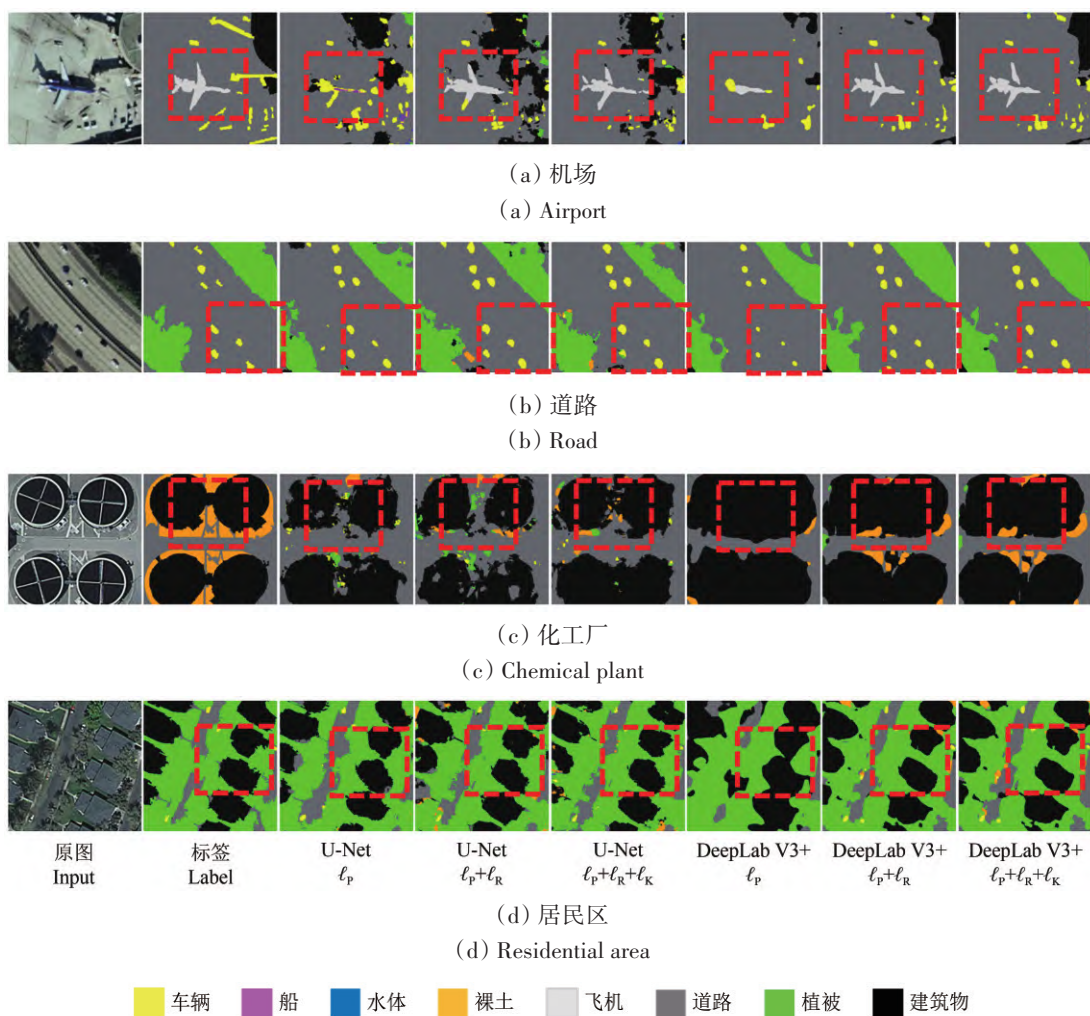


图3 UCM测试集上的可视化分割结果图

Fig. 3 The visible semantic segmentation of the test set of the UCM dataset

3.4.2 DeepGlobe数据集上的结果

本文方法在DeepGlobe测试集上的语义分割精度如表8所示。可见相比于基于像素级稠密约束的常规损失 (ℓ_p) 训练的语义分割网络, 加入实体级连通约束 ($\ell_p + \ell_R$) 训练的网络的分割精度OA、MIoU和FWIoU均更高, 这说明了实体级学习能够有效提高深度语义分割网络的性能; 综合实体级连通约束与实体间共生约束 ($\ell_p + \ell_R + \ell_K$) 的深度语义分割网络精度优于常规的像素级稠密约束 (ℓ_p) 的网络。

表8 在DeepGlobe测试集上的语义分割精度

Table 8 The accuracy of semantic segmentation on the test set of the DeepGlobe dataset

基础网络框架	损失约束项	OA	MIoU	FWIoU
U-Net	$\mathcal{L} = \ell_p$	81.08	50.74	70.62
	$\mathcal{L} = \ell_p + \ell_R$	81.53	51.91	71.15
	$\mathcal{L} = \ell_p + \ell_R + \ell_K$	82.30	52.20	71.50
DeepLab v3+	$\mathcal{L} = \ell_p$	84.26	55.90	74.15
	$\mathcal{L} = \ell_p + \ell_R$	85.32	57.45	75.83
	$\mathcal{L} = \ell_p + \ell_R + \ell_K$	85.75	57.84	76.29

注: 粗体数字表示几组设置中精度最优设置。

图4为深度语义分割网络在DeepGlobe测试集上的分割结果。可见：实体级连通约束 ($\ell_p + \ell_r$) 分割结果和综合实体级连通约束与实体间共生约束 ($\ell_p + \ell_r + \ell_k$) 的分割结果均明显优于基于像素级稠密约束的常规损失 (ℓ_p) 的分割结果；第一行和第二行的水体和牧场区域施加实体级连通约束与实体间共生约束 ($\ell_p + \ell_r + \ell_k$) 的分割结果

整体性更佳；第三行和第四行的水体和农业用地的分割结果施加约束后轮廓清晰，噪声更少，这表明约束项使得网络学习到了目标级的特征表示。综合以上结果可以看出综合实体级连通约束与实体间共生约束的分割结果中地物目标空间分布更加简单，一些错分目标类别被修正为邻域目标类别，从而优化了地物目标的空间布局。

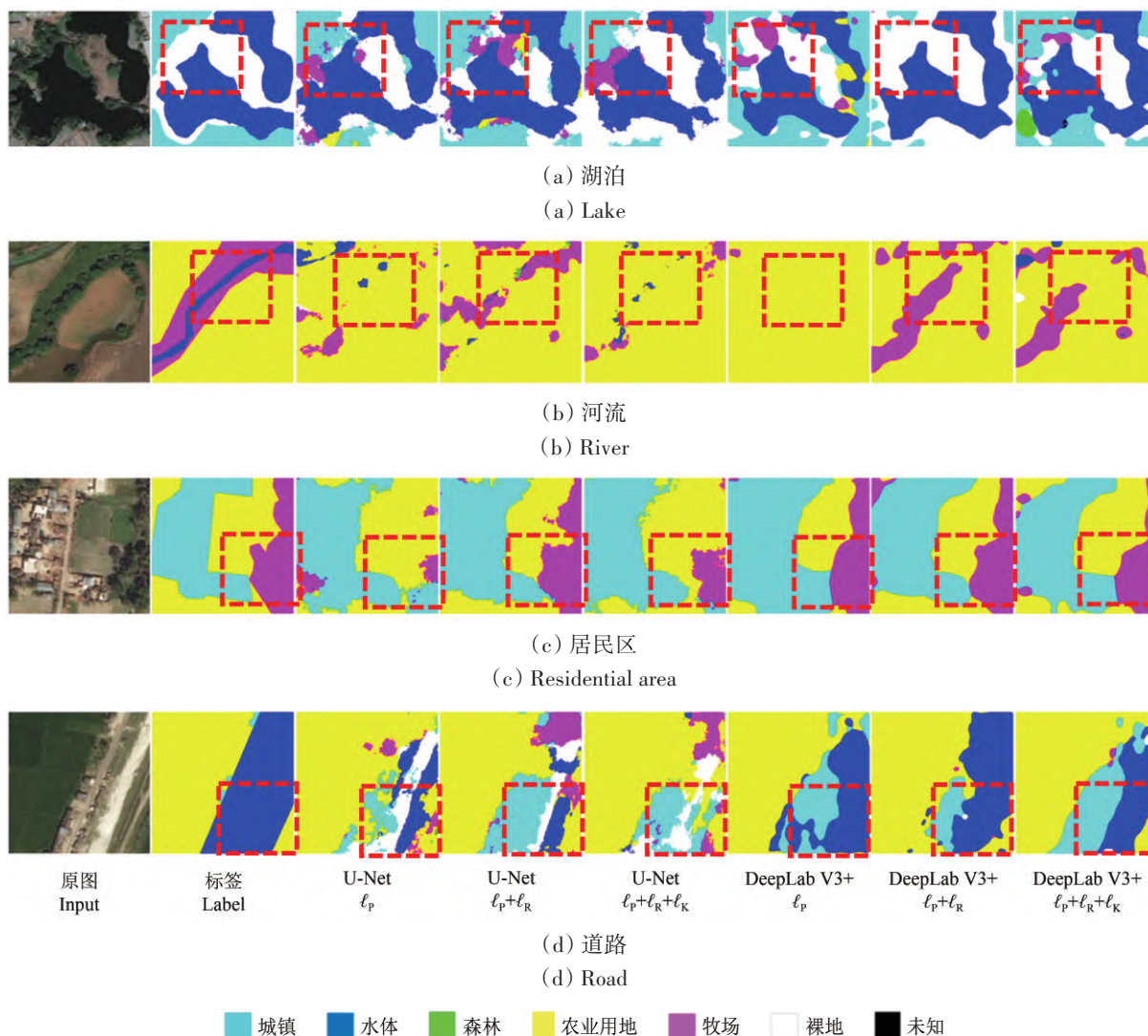


图4 DeepGlobe测试集上的可视化分割结果图

Fig. 4 The visible semantic segmentation of the test set of the DeepGlobe dataset

3.5 与已有方法的对比分析

为进一步验证本文方法的有效性，本文选择了经典的U-Net网络 (Ronneberger等, 2015)、先进的DeepLab V3+网络 (Chen等, 2018)、深度语义分割网络和图卷积网络相结合的DSSN-GCN方法 (Ouyang和Li, 2021) 作为对比方法，其中

DSSN-GCN方法以DeepLab V3+作为基础网络。本文方法以DeepLab V3+作为基础网络，超参数 α 和 β 取各数据集下的最佳值。各方法在UCM测试集和DeepGlobe测试集上的分割精度见表9和表10所示。可见本文提出的地学知识图谱引导的遥感影像深度语义分割方法综合了实体级连通约束与实

体间共生约束 ($l_p + l_r + l_k$), 均取得最佳的分割精度, 这充分验证了本文方法对于遥感影像语义分割的有效性, 同时说明了在深度语义分割网络中嵌入地学先验知识的重要性。相比于地学先验知识预嵌入的DSSN-GCN方法, 本文方法能够自主运用和学习地学先验知识, 从而获得精度高和鲁棒性强的分割结果。各方法在UCM测试集和DeepGlobe测试集上的语义分割结果如图5和图6所示。与表9和表10结果一致, 本文方法取得比其余方法更佳的分割效果。各方法对建筑物的分割结果表明在实体级连通损失的约束下, 本文方法的分割结果更具整体性, 轮廓清晰(第一行至第三行)且对于干扰较大的场景时噪声更少(第四行至第六行); 实空间共生知识约束修正了分割结果中地物目标的空间分布, 使其更接近现实世界的空间布局, 如第一行所示的车辆通常近邻道路(图5)。图6中各方法分割结果也表现出类似的结论, 本文方法对城镇和水体的分割结果更具整体性, 对于水体轮廓的分割结果更加清晰, 且对于复杂城镇场景的分割噪声更少。

表9 各方法在UCM测试数据集上的语义分割精度

Table 7 The accuracy of semantic segmentation of various methods on the test set of UCM dataset

方法	指标		
	OA	MIoU	FWIoU
U-Net(Ronneberger等,2015)	81.45	61.26	68.76
DeepLab V3+(Chen等,2018)	83.67	66.37	72.11
DSSN-GCN(Ouyang和Li,2021)	84.29	63.03	72.96
本文方法	85.95	75.13	75.49

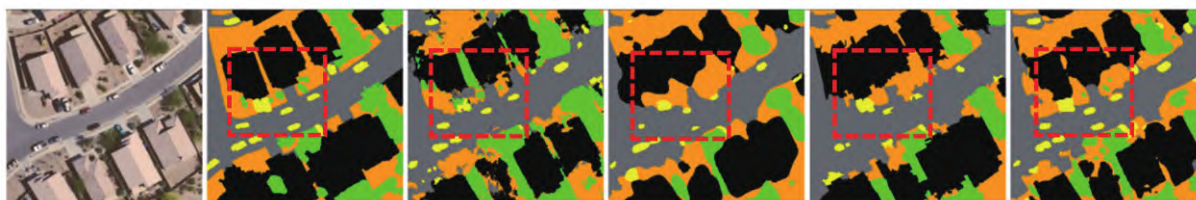
注: 粗体数字表示几组设置中精度最优设置。

表10 各方法在DeepGlobe数据集上的语义分割精度

Table 10 The accuracy of semantic segmentation of various methods on the test set of DeepGlobe dataset

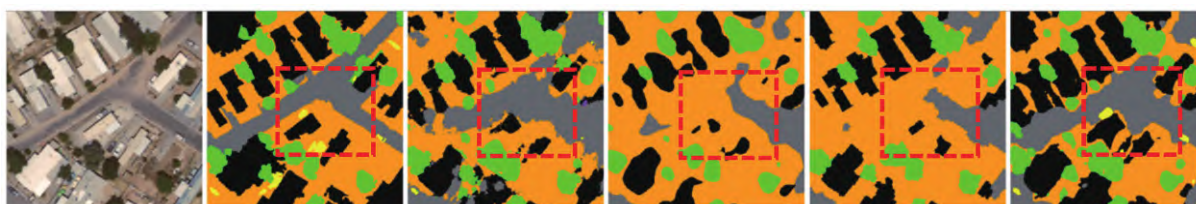
方法	指标		
	OA	MIoU	FWIoU
U-Net(Ronneberger等,2015)	81.08	50.74	70.62
DeepLab V3+(Chen等,2018)	84.26	55.90	74.15
DSSN-GCN(Ouyang和Li,2021)	84.57	56.31	74.41
本文方法	85.74	57.84	76.29

注: 粗体数字表示几组设置中精度最优设置。



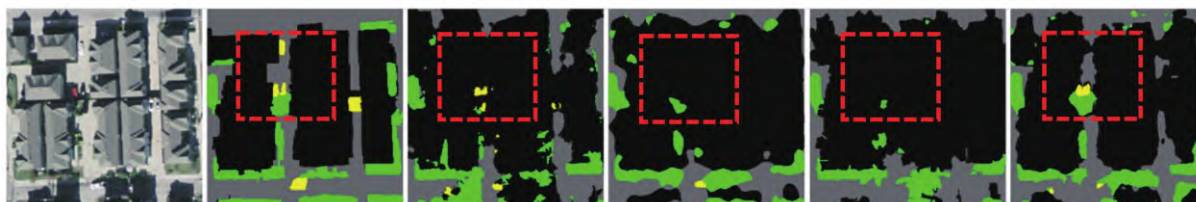
(a) 街道1

(a) Street 1



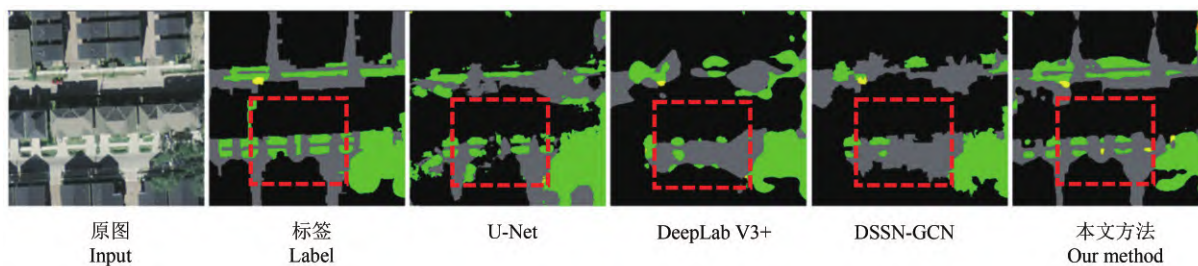
(b) 街道2

(b) Street 2



(c) 居民区1

(c) Residential area 1



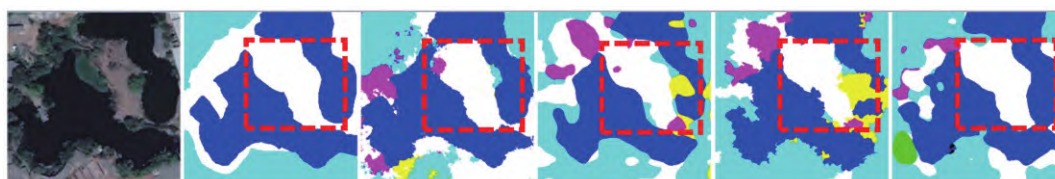
(d) 居民区 2

(d) Residential area 2



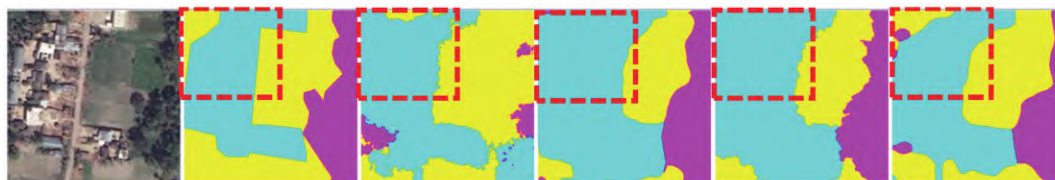
图 5 各方法在 UCM 测试集上的语义分割结果

Fig. 5 The semantic segmentation results of the test set of the UCM dataset



(a) 湖泊

(a) Lake



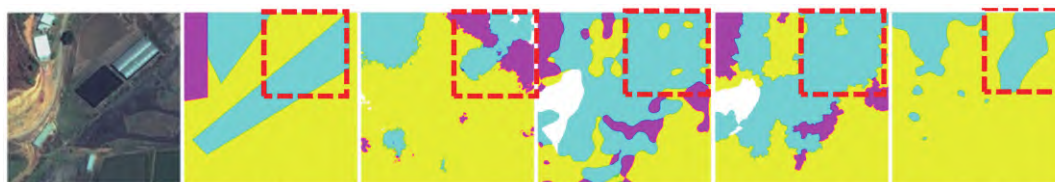
(b) 居民区 1

(b) Residential area one



(c) 居民区 2

(c) Residential area two



(d) 农田

(d) Farmland



图 6 各方法在 DeepGlobe 测试集上的语义分割结果

Fig. 6 The semantic segmentation results of the test set of the DeepGlobe dataset

4 结 论

针对深度语义分割网络缺乏目标级学习以及难以利用地学先验知识和空间语义信息的问题, 本文提出了地学知识图谱引导的遥感影像深度语义分割方法。地物目标的语义信息以地学知识图谱的形式进行表达, 地学先验知识从地学知识图谱中抽取, 地学先验知识和目标空间语义信息用于构建实体级连通约束损失和实体间共生约束损失。实体级连通约束损失引导深度语义分割网络自主学习目标级特征, 实现了对实体的整体约束, 使得分割结果更具整性并减少了边界模糊和随机噪声; 实体间共生约束损失成功地将非结构化的地学先验知识嵌入到数据驱动的深度语义分割网络中, 完成了对分割目标的空间分布的自动优化。本文方法有效地提高了深度语义分割网络的性能和鲁棒性, 但同时也存在所采用的地学先验知识单一的问题。后续研究工作将引入更全面的地学先验知识, 包括地物目标空间拓扑信息、形状结构信息等知识。

志 谢 本研究的数值计算部分得到了武汉大学超级计算中心的帮助。

参考文献 (References)

- Achanta R, Shaji A, Smith K, Lucchi A, Fua P and Süsstrunk S. 2012. SLIC Superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11): 2274-2282 [DOI: 10.1109/TPAMI.2012.120]
- Alirezaie M, Långkvist M, Sioutis M and Loutfi A. 2019. Semantic referee: a neural-symbolic framework for enhancing geospatial semantic segmentation. *Semantic Web*, 10(5): 863-880 [DOI: 10.3233/SW-190362]
- Arvor D, Belgiu M, Falomir Z, Mougenot I and Durieux L. 2019. Ontologies to interpret remote sensing images: why do we need them? *GIScience and Remote Sensing*, 56(6): 911-939 [DOI: 10.1080/15481603.2019.1587890]
- Badrinarayanan V, Kendall A and Cipolla R. 2017. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12): 2481-2495 [DOI: 10.1109/TPAMI.2016.2644615]
- Camps-Valls G, Tuia D, Bruzzone L and Benediktsson J A. 2014. Advances in hyperspectral image classification: earth monitoring with statistical learning methods. *IEEE Signal Processing Magazine*, 31(1): 45-54 [DOI: 10.1109/MSP.2013.2279179]
- Chen L C, Zhu Y K, Papandreou G, Schroff F and Adam H. 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation//*Proceedings of the 15th European Conference on Computer Vision*. Munich, Germany: Springer: 833-851 [DOI: 10.1007/978-3-030-01234-2_49]
- Dash T, Chitlangia S, Ahuja A and Srinivasan A. 2022. A review of some techniques for inclusion of domain-knowledge into deep neural networks. *Scientific Reports*, 12(1): 1040 [DOI: 10.1038/s41598-021-04590-0]
- Demir I, Koperski K, Lindenbaum D, Pang G, Huang J, Basu S, Hughes F, Tuia D and Raska R. 2018. DeepGlobe 2018: a challenge to parse the earth through satellite images//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Salt Lake City: IEEE: 172-17209 [DOI: 10.1109/CVPRW.2018.00031]
- Fan J, Yu W Z, Wu W and Shen Y. 2017. Knowledge-guided fitting method for sparse time series remote sensing data. *Journal of Remote Sensing*, 21(5): 749-756 (范菁, 余维泽, 吴炜, 沈瑛. 2017. 知识引导的稀疏时间序列遥感数据拟合. *遥感学报*, 21(5): 749-756) [DOI: 10.11834/jrs.20176434]
- Goodenough D G, Goldberg M, Plunkett G and Zelek J. 1987. An expert system for remote-sensing. *IEEE Transactions on Geoscience and Remote Sensing*, GE-25(3): 349-359 [DOI: 10.1109/TGRS.1987.289805]
- Gu H Y, Li H T, Yan L, Liu Z J, Blaschke T and Soergel U. 2017. An object-based semantic classification method for high resolution remote sensing imagery using ontology. *Remote Sensing*, 9(4): 329 [DOI: 10.3390/rs9040329]
- Gui R, Xu X, Dong H, Song C and Pu F L. 2016. Individual building extraction from TerraSAR-X images based on ontological semantic analysis. *Remote Sensing*, 8(9): 708 [DOI: 10.3390/rs8090708]
- He K M, Gkioxari G, Dollár P and Girshick R. 2018. Mask R-CNN//*Proceedings of 2017 IEEE International Conference on Computer Vision*. Venice, Italy: IEEE: 2980-2988 [DOI: 10.1109/ICCV.2017.322]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//*Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- Hogan A, Blomqvist E, Cochez M, D'Amato C, de Melo G, Gutierrez C, Kirrane S, Gayo J E L, Navigli R, Neumaier S, Ngomo A C N, Polleres A, Rashid S M, Rula A, Schmelzeisen L, Sequeda J, Staab S and Zimmermann A. 2022. Knowledge graphs. *ACM Computing Surveys*, 54(4): 71 [DOI: 10.1145/3447772]
- Li F S, Li X J, Chen W T, Dong Y S, Li Y K and Wang L Z. 2022. Automatic lithology classification based on deep features using dual polarization SAR images. *Earth Science*, 47(11): 4267-4279 (李发森, 李显巨, 陈伟涛, 董玉森, 李雨柯, 王力哲. 2022. 基于深度特征的双极化 SAR 遥感图像岩性自动分类. *地球科学*, 47(11): 4267-4279) [DOI: 10.3799/dqkx.2022.129]
- Li J, Wu Z C, Sheng Q H, Wang B, Hu Z W, Zheng S B, Camps-Valls G and Molinier M. 2022. A hybrid generative adversarial network for weakly-supervised cloud detection in multispectral images.

- Remote Sensing of Environment, 280: 113197 [DOI: 10.1016/j.rse.2022.113197]
- Li N, Zhu X F, Pan Y Z and Zhan P. 2018. Optimized SVM based on artificial bee colony algorithm for remote sensing image classification. *Journal of Remote Sensing*, 22(4): 559-569 (李楠, 朱秀芳, 潘耀忠, 詹培. 2018. 人工蜂群算法优化的SVM遥感影像分类. *遥感学报*, 22(4): 559-569) [DOI: 10.11834/jrs.20187176]
- Li Y S, Chen W, Huang X, Gao Z, Li S W, He T and Zhang Y J. 2023. MFVNet: a deep adaptive fusion network with multiple field-of-views for remote sensing image semantic segmentation. *Science China Information Sciences*, 66(4): 140305 [DOI: 10.1007/s11432-022-3599-y]
- Li Y S, Chen W, Zhang Y J, Tao C, Xiao R and Tan Y H. 2020. Accurate cloud detection in high-resolution remote sensing imagery by weakly supervised deep learning. *Remote Sensing of Environment*, 250: 112045 [DOI: 10.1016/j.rse.2020.112045]
- Li Y S, Kong D Y, Zhang Y J, Tan Y H and Chen L. 2021a. Robust deep alignment network with remote sensing knowledge graph for zero-shot and generalized zero-shot remote sensing image scene classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 179: 145-158 [DOI: 10.1016/j.isprsjprs.2021.08.001]
- Li Y S, Shi T, Zhang Y J, Chen W, Wang Z B and Li H. 2021b. Learning deep semantic segmentation network under multiple weakly-supervised constraints for cross-domain remote sensing image semantic segmentation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 175: 20-33 [DOI: 10.1016/j.isprsjprs.2021.02.009]
- Li Y S and Zhang Y J. 2022. A new paradigm of remote sensing image interpretation by coupling knowledge graph and deep learning. *Geomatics and Information Science of Wuhan University*, 47(8): 1176-1190 (李彦胜, 张永军. 2022. 耦合知识图谱和深度学习的新一代遥感影像解译范式. *武汉大学学报(信息科学版)*, 47(8): 1176-1190) [DOI: 10.13203/j.whugis20210652]
- Li Y S, Zhang Y J, Huang X, Zhu H and Ma J Y. 2018. Large-scale remote sensing image retrieval by deep hashing neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 56(2): 950-965 [DOI: 10.1109/TGRS.2017.2756911]
- Li Y S, Zhou Y H, Zhang Y J, Zhong L H, Wang J and Chen J D. 2022. DKDFN: domain knowledge-guided deep collaborative fusion network for multimodal unitemporal remote sensing land cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 186: 170-189 [DOI: 10.1016/j.isprsjprs.2022.02.013]
- Liu B, Du S H, Du S J and Zhang X Y. 2020. Incorporating deep features into GEOBIA paradigm for remote sensing imagery classification: a patch-based approach. *Remote Sensing*, 12(18): 3007 [DOI: 10.3390/rs12183007]
- Long J, Shelhamer E and Darrell T. 2015. Fully convolutional networks for semantic segmentation//*Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition*. Boston: IEEE: 3431-3440 [DOI: 10.1109/CVPR.2015.7298965]
- Ma L, Liu Y, Zhang X L, Ye Y X, Yin G F and Johnson B A. 2019. Deep learning in remote sensing applications: a meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152: 166-177 [DOI: 10.1016/j.isprsjprs.2019.04.015]
- Ouyang S and Li Y S. 2021. Combining deep semantic segmentation network and graph convolutional neural network for semantic segmentation of remote sensing imagery. *Remote Sensing*, 13(1): 119 [DOI: 10.3390/rs13010119]
- Ouyang S B, Chen W T, Li X J, Dong Y S and Wang L Z. 2022. Geomorphological scene classification dataset of high-resolution remote sensing imagery in vegetation-covered areas. *National Remote Sensing Bulletin*, 26(4): 606-619 (欧阳淑冰, 陈伟涛, 李显巨, 董玉森, 王力哲. 2022. 植被覆盖区高精度遥感地貌场景分类数据集. *遥感学报*, 26(4): 606-619) [DOI: 10.11834/jrs.20221385]
- Ronneberger O, Fischer P and Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation//*Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention*. Munich: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- Shao Z F, Yang K and Zhou W X. 2018. Performance evaluation of single-label and multi-label remote sensing image retrieval using a dense labeling dataset. *Remote Sensing*, 10(6): 964 [DOI: 10.3390/rs10060964]
- Wang Z H, Yang X M and Zhou C H. 2021. Geographic knowledge graph for remote sensing big data. *Journal of Geo-information Science*, 23(1): 16-28 (王志华, 杨晓梅, 周成虎. 2021. 面向遥感大数据的地理知识图谱构想. *地球信息科学学报*, 23(1): 16-28) [DOI: 10.12082/dqxxkx.2021.200632]
- Wu R, Luo W F and Chen J H. 2022. Knowledge graph embedding spectral unmixing. *National Remote Sensing Bulletin*, 26(12): 1-17 (吴瑞, 罗文斐, 陈江浩. 2022. 知识图谱嵌入的光谱解混算法. *遥感学报*, 26(12): 1-17) [DOI: 10.11834/jrs.20222253]
- Wu X, Shi Z W and Zou Z X. 2021. A geographic information-driven method and a new large scale dataset for remote sensing cloud/snow detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, 174: 87-104 [DOI: 10.1016/j.isprsjprs.2021.01.023]
- Xiao C J, Li Y, Zhang H Q and Chen J. 2020. Semantic segmentation of remote sensing image based on deep fusion networks and conditional random field. *Journal of Remote Sensing*, 24(3): 254-264 (肖春皎, 李宇, 张洪群, 陈俊. 2020. 深度融合网络结合条件随机场的遥感图像语义分割. *遥感学报*, 24(3): 254-264) [DOI: 10.11834/jrs.20208298]
- Xie E Z, Wang W H, Yu Z D, Anandkumar A, Álvarez J M and Luo P. 2021. SegFormer: simple and efficient design for semantic segmentation with transformers. *arXiv*: 2105.15203. [DOI: 10.48550/arXiv.2105.15203]
- Xu J C, Xiong Z X and Bhattacharyya S P. 2023. PIDNet: a real-time semantic segmentation network inspired by PID controllers//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vancouver: IEEE: 19529-19539 [DOI: 10.1109/CVPR52729.2023.01871]
- Xu Z X, Wu K, Huang L, Wang Q M and Ren P. 2022. Cloudy image arithmetic: a cloudy scene synthesis paradigm with an application to deep-learning-based thin cloud removal. *IEEE Transactions on*

- Geoscience and Remote Sensing, 60: 5612616 [DOI: 10.1109/TGRS.2021.3122253]
- Zhang H Y, Zhou C H, Lv G N, Wu Z F, Lu F, Wang J F, Yue T X, Luo J C, Ge Y and Qin C Z. 2020. The connotation and inheritance of Geo-information Tupu. *Journal of Geo-information Science*, 22 (4): 653-661 (张洪岩, 周成虎, 闫国年, 吴志峰, 陆锋, 王劲峰, 岳天祥, 骆剑承, 葛咏, 秦承志. 2020. 试论地学信息图谱思想的内涵与传承. *地球信息科学学报*, 22(4): 653-661) [DOI: 10.12082/dqxxkx.2020.200167]
- Zhang Y J, Wang F, Li Y S, Ouyang S, Wei D, Liu X J, Kong D Y, Chen R X and Zhang B. 2023. Remote sensing knowledge graph construction and its application in typical scenarios. *National Remote Sensing Bulletin*, 27(2): 249-266 (张永军, 王飞, 李彦胜, 欧阳松, 魏东, 刘晓建, 孔德宇, 陈瑞贤, 张斌. 2023. 遥感知识图谱创建及其典型场景应用技术. *遥感学报*, 27(2): 249-266) [DOI: 10.11834/jrs.20210469]
- Zhao H S, Shi J P, Qi X J, Wang X G and Jia J Y. 2017. Pyramid scene parsing network//*Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE: 6230-6239 [DOI: 10.1109/CVPR.2017.660]
- Zhu X X, Tuia D, Mou L C, Xia G S, Zhang L P, Xu F and Fraundorfer F. 2017. Deep learning in remote sensing: a comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4): 8-36 [DOI: 10.1109/MGRS.2017.2762307]

Geographic knowledge graph-guided remote sensing image semantic segmentation

LI Yansheng¹, WU Kang¹, OUYANG Song¹, YANG Kun², LI Heping³, ZHANG Yongjun¹

1. School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China;

2. Basic Geographic Information Center of Guizhou Province, Guiyang 550004, China;

3. Guizhou First Surveying and Mapping Institute, Guiyang 550025, China

Abstract: Although the Deep Semantic Segmentation Network (DSSN) has notably enhanced remote-sensing image semantic segmentation, it still falls short of human experts' visual interpretation. Unlike DSSN's data-driven, pixel-level optimization, human experts rely on visual features, semantic insight, and prior knowledge for remote-sensing image interpretation. DSSN's pixel-level approach is constrained by spatial scale, lacking comprehensive target inference and struggling to bridge structured data and unstructured knowledge. In response to the two issues above, this paper proposes a geographic knowledge graph-guided deep semantic segmentation network for remote-sensing imagery. We use the ground-object semantic information and geoscience prior knowledge extracted from the geographic knowledge graph to construct loss constraints, thereby autonomously guiding the training process of DSSN.

The essence of our approach lies in the intricately crafted design of loss constraints. These loss constraints include the entity-level connectivity constraint and the inter-entity symbiosis constraint. The former calculates the loss in the unit of connected domain entities instead of pixels to achieve overall constraints on the entity. The latter embeds the spatial symbiosis knowledge quantified by the symbiosis conditional probability into the data-driven DSSN to constrain the spatial distribution of segmented entities. The entity-level connectivity constraint guides DSSN to autonomously learn entity-level feature representations during training. Accordingly, the segmentation results become more holistic and suppresses blurry boundaries and random noise. The inter-entity symbiosis constraint adjusts the spatial distribution of entities according to the spatial semantic information and the prior geoscience knowledge. This adjustment realizes the automatic optimization of the spatial distribution of segmented entities.

Extensive experiments show that under the guidance of the entity-level connectivity constraint and the inter-entity symbiosis constraint, DSSN can complete the learning of entity-level features. It can also automatically optimize the spatial distribution of ground objects based on spatial symbiosis knowledge, thereby effectively improving the performance of remote-sensing image semantic segmentation.

Our novel geographic knowledge graph-guided approach to deep semantic segmentation in remote-sensing imagery has successfully addressed the challenges posed by DSSN's pixel-level optimization. By incorporating entity-level connectivity and inter-entity symbiosis constraints, we have enabled DSSN to autonomously learn comprehensive feature representations and optimize spatial distribution. The resulting improvements in semantic segmentation performance showcase the potential of merging domain-specific knowledge with data-driven techniques, bridging the gap between automated methods and human interpretation in remote-sensing image analysis.

Key words: geographic knowledge graph, deep semantic segmentation network, entity-level connectivity constraint, spatial symbiosis knowledge constraint, geographic knowledge embedding optimization

Supported by National Natural Science Foundation of China (No. 42030102, 41971284, 42371321)