

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(XXXX)XX-0001-15

论文引用格式: Rong Yu, Lin Hui, Hang Renlong. XXXX. Boundary-guided and shape-aware semi-supervised semantic segmentation for remote sensing image. Journal of Image and Graphics, XX(XX):0001-0015(戎誉, 林辉, 杭仁龙. XXXX. 边界引导与形状感知的遥感图像半监督语义分割. 中国图象图形学报, XX(XX):0001-0015)[DOI:10.11834/jig.250575]

边界引导与形状感知的遥感图像半监督语义分割

戎誉¹, 林辉², 杭仁龙²

1. 南京信息工程大学 软件学院, 南京 210044; 2. 南京信息工程大学 计算机学院, 南京 210044

摘要: 目的 针对现有遥感图像半监督分割方法存在的伪标签利用率低及物体边界处误分等问题, 提出一种边界引导与形状感知的半监督语义分割网络 (boundary-guided and shape-aware semi-supervised semantic segmentation network, BS⁵-Net), 提升模型在复杂场景下的分割性能。方法 首先, 利用像素间的相关性建模形状信息, 以此修正伪标签中由于置信度低而被忽略的正确预测, 使伪标签质量得到有效提升, 从而为模型训练提供更可靠的监督信号; 其次, 针对边界区域因下采样导致的空间细节与边缘信息丢失问题, 引入基于 Canny 边缘检测器的辅助优化模块, 通过轮廓引导特征图学习, 锐化分割边界; 最后, 设计了一个三阶段动态阈值函数, 根据置信度平均值分阶段调整阈值, 从而更精准地甄别可靠样本。结果 在 Potsdam、Vaihingen 和 LoveDA 这三个公开数据集上的实验结果表明, 仅使用 1/4 有标签训练样本时, 本研究的方法在 Potsdam 数据集上的平均交并比 (mean intersection over union, mIoU) 达到 83.43%, 性能超越现有先进方法 1.28%, 在 Vaihingen 数据集上的 mIoU 达到 75.51%, 性能超越现有先进方法 1.01%, 在 LoveDA 数据集上的 mIoU 达到 51.88%, 性能超越现有先进方法 2.72%; 当有标签训练样本减少至 1/8 时, 本研究的方法在 Potsdam 数据集上的 mIoU 为 82.95%, 性能超越现有先进方法 3.02%, 在 Vaihingen 数据集上的 mIoU 为 74.70%, 性能超越现有先进方法 2.12%, 在 LoveDA 数据集上的 mIoU 达到 51.66%, 性能超越现有先进方法 3.26%。结论 本研究提出的方法有效提升了特征判别与边界分割能力, 优化了伪标签的生成质量, 提高了伪标签的利用率, 在性能上取得了优于现有先进方法的结果, 为相关领域的探索提供了新思路。代码开源链接 <https://github.com/AquariusRY/BS5-Net>。

关键词: 语义分割; 半监督; 遥感图像; 相关性矩阵; 边界; 形状

Boundary-guided and shape-aware semi-supervised semantic segmentation for remote sensing image

Rong Yu¹, Lin Hui², Hang Renlong²

1. School of Software, Nanjing University of Information Science and Technology, Nanjing, 210044, China; 2. School of Computer Science, Nanjing University of Information Science and Technology, Nanjing, 210044, China

Abstract: Objective The success of deep learning models typically relies on supervised training with large amounts of high-quality annotated data. However, in the task of semantic segmentation for remote sensing images, obtaining pixel-level annotations is a process known to be complex, time-consuming, and demanding of professional expertise, leading to a

收稿日期: 2025-11-13; 修回日期: 2026-01-19

* 通信作者: 杭仁龙 renlong_hang@163.com

基金项目: 国家自然科学基金项目 (T2522016, 62472230); 江苏省自然科学基金资助项目 (BK20250041)

Supported by: National Natural Science Foundation of China (T2522016, 62472230); Natural Science Foundation of Jiangsu Province (BK20250041)

significant scarcity of labeled data in practical applications. This dependence on extensive annotation has become a critical bottleneck restricting further improvements in model performance and hindering practical deployment. Within this context, semi-supervised semantic segmentation methods have emerged as a promising alternative. These methods leverage specific learning mechanisms, such as consistency regularization and pseudo-labeling, to fully exploit large quantities of readily available unlabeled data. The goal is to significantly reduce reliance on expensive manual annotations while effectively enhancing the model's segmentation accuracy and generalization capability. Consequently, semi-supervised learning provides a feasible solution to the problem of annotation scarcity in remote sensing, demonstrating considerable application potential in key areas such as urban planning and precision agriculture. Despite their advantages, many existing semi-supervised methods focus on complex training strategies to extract information from unlabeled data but often fail to adequately account for the complex spatial layouts and inherent structural features of remote sensing imagery. This deficiency can result in a degree of confusion between different categorical features within the learned feature space, which ultimately limits the efficiency of unlabeled data utilization and compromises the model's ability to discriminate fine boundaries between land cover features. To overcome these limitations, a Boundary-Guided and Shape-Aware Semi-Supervised Semantic Segmentation Net (BS⁵-Net) is proposed in this study. The proposed model aims to optimize the quality of pseudo-labels and enhance the utilization of unlabeled data by explicitly integrating geometric constraints. It is anticipated that this approach will improve feature discrimination and boundary precision, thereby leading to enhanced overall performance in semi-supervised semantic segmentation of remote sensing images. **Methods** This study proposes a semi-supervised method that leverages ResNet as its backbone network. The core innovation involves using correlation information to integrate edge geometric constraints and a dynamic threshold mechanism, thereby enhancing pseudo-label quality. Firstly, considering that pixels do not exist independently, but have inherent connections. Generally speaking, pixels with similar semantics have higher values in the correlation matrix. Based on this, this study designed a Shape Extraction Module (SEM). By calculating the inter-pixel correlation matrix, the SEM module explicitly simulates local geometric relationships, enabling structured refinement of initial pseudo-labels. This process effectively corrected the problem of low confidence correct predictions being ignored due to confidence screening strategies, improved the utilization of pseudo-labels, and provided more reliable supervision signals for subsequent learning of the model. Secondly, an edge collaboration module (ECM) was designed to address the issue of spatial details and edge information loss caused by down sampling in boundary regions. This module adopts Canny edge detector as an auxiliary optimization mechanism, injecting explicit geometric priors layer by layer by extracting high confidence contours. The resulting edge response guides the spatial distribution of the feature map during the training process, significantly improving the accuracy of boundary segmentation and structural integrity. Finally, a three-stage threshold adjustment module (TAM) was developed to mitigate category bias caused by class imbalance. This function autonomously adjusts the pseudo-label threshold for each category based on its average prediction confidence, enabling a more balanced filtering strategy across categories. By integrating these strategies, the proposed network achieves precise and structurally coherent segmentation of remote sensing images. **Result** We evaluate BS⁵-Net using three publicly available labeled datasets (Potsdam, Vaihingen and LoveDA). The Experimental results demonstrate its strong performance in semantic segmentation, particularly under limited labeled data. With only 1/4 of the labeled data, BS⁵-Net achieves an mIoU of 83.43% on Potsdam, 75.51% on Vaihingen and 51.88% on LoveDA, surpassing existing advanced methods by 1.28%, 1.01% and 2.72%. When the labeled data is further reduced to 1/8, it achieves 82.95% mIoU on Potsdam, 74.70% on Vaihingen and 51.66% on LoveDA, outperforming existing methods by margins of 3.02%, 2.12% and 3.26%, respectively. **Conclusion** This article proposes a boundary-guided and shape-aware Semi-Supervised Semantic Segmentation Network (BS⁵-Net) with shape perception and boundary enhancement capabilities by structurally modeling the spatial relationships between pixels and integrating edge geometric constraints and dynamic threshold mechanisms. The experimental results on three publicly available remote sensing datasets show that BS⁵-Net can effectively utilize unlabeled data and exhibits superior segmentation performance when trained with different numbers of labeled data. The visualization results further demonstrate that this method can still maintain clear ground boundaries and consistent region segmentation effects in complex scenes. This study provides an effective solution to reduce the model's dependence on annotated data, which has positive significance for promoting intelligent interpretation of remote sensing images.

Key words: semantic segmentation; semi-supervision; remote sensing images; boundary; shape

论文引用格式: (论文引用格式: [DOI:])

0 引言

遥感图像语义分割是计算机视觉领域的关键任务之一,旨在为遥感图像中的每个像素赋予相应的语义类别标签(严毅等,2023)。该任务通过对图像进行像素级解析,能够从高分辨率遥感影像中精确提取各类地物的空间分布信息,从而为城市规划、土地利用以及灾害评估等领域提供重要的数据支撑与决策依据(董荣胜等,2022)。

自全卷积网络(fully convolutional network, FCN)(Long等,2015)问世以来,研究者们相继提出了U-Net(Ronneberger等,2015)、DeepLab(Chen等,2018)、PSPNet(pyramid scene parsing network)(Zhao等,2017)、HRNet(high-resolution net)(Wang等,2021)等一系列语义分割方法,有效提升了自然图像分割的性能,这一趋势也推动了遥感图像语义分割任务的发展(Lu等,2024; Mou等,2020; Peng等,2022)。研究者们尝试在模型中引入多尺度特征融合、注意力机制(Vaswani等,2017)以及边界信息建模等技术,从不同层面提升遥感图像语义分割的准确性。Wang等人(2024)设计了边界引导的交叉融合网络(boundary guided cross fusion network, BGCF-Net),增强模型的边界感知能力;Li等人(2024)设计了边界增强型双流网络(boundary-enhanced dual-stream network, BEDSN),通过融合边缘检测与语义分割,缓解边界模糊问题。然而,上述全监督方法在训练时高度依赖像素级标注,而遥感图像标注工作成本高昂且易受主观因素影响,这导致了标注数据的稀缺,制约了深度学习在遥感图像语义分割领域中的应用。

为了克服语义分割任务中标注数据不足的问题,学者们提出了半监督学习方法,利用少量标注数据与大量无标注数据进行协同训练,以获得接近全监督模型的性能(Ke等,2020; He等,2022)。目前主流的半监督方法包括基于对抗性的方法(李志欣等,2022)、基于一致性正则化的方法(刘腊梅等,2022)和基于伪标签的方法(项伟康等,2024)。其中基于伪标签的方法通过利用经由有标签数据预训练过的

模型来生成无标签数据的预测结果,并将预测结果中置信度高于某一阈值的部分作为“伪标签”,扩充训练集。

近年来,大量研究者对伪标签的生成与优化机制展开深入探索。FixMatch(Sohn等,2020)通过设定固定的阈值来筛选高质量伪标签,但在训练初期会因高质量样本过少,导致模型难以充分学习困难样本。FlexMatch(Zhang等,2022)意识到了以上问题并提出了针对性的解决方案,它根据模型的学习状态动态调整不同类别的阈值,直到每个类的置信度都达到固定的全局阈值。FreeMatch(Wang等,2022)则根据模型的学习状态以自适应的方式调整置信度阈值,并引入自适应类公平正则化,以鼓励模型在训练的早期进行多样化预测。Jin等人(2024)尝试性地将动态采样、分布对齐与自适应阈值相结合,确保样本较少的类被赋予更高的优先级,从而改善分割结果。

以上方法通过筛选出高置信度的伪标签,有效提升了半监督遥感语义分割的性能。然而其阈值调整策略是连续且单一的,且对低置信度像素的利用率明显不足。U2PL(Wang等,2022)结合均值教师模型(mean teacher, MT)和对比学习方法,期望尽可能多地利用不可靠伪标签的潜在信息,但却忽略了对遥感图像不同区域间的上下文信息的利用。于是, Li等人(2023)提出SegMind方法,引入掩码图像建模,以此增强不同区域间的信息交互。Sun等人(2023)提出了CorrMatch方法,该方法通过挖掘像素间的相关性信息,驱动伪标签在特征空间中传播,能有效提升低置信度区域的分割一致性。然而该方法选择从高层特征中计算相关性图,这些特征虽然富含语义信息,但大多因下采样而丢失了部分精确的、底层的几何轮廓。因此,在边界模糊、结构复杂或与背景对比度低的区域,无法精确识别物体边缘。Ni等人(2020)提出边界信息引导的网络(edge information guided network, EIGNet),引入边缘信息与多方向空间特征提取机制,提升边界处的分割精度。然而,其边缘特征仅作为权重引导特征融合,这导致模型在对比度低或语义重叠区域的边缘感知能力较弱。

尽管上述算法各有优点,但仍存在以下几个亟
© 中国图象图形学报版权所有

待解决的问题:1)使用阈值作为伪标签生成条件忽略了像素与像素间的内在联系,从而导致高置信度噪声被保留、低置信度正确预测被过滤,进而影响伪标签的完整性;2)现有方法为追求更大的感受野与高级语义特征而过度依赖下采样,导致边界处的空间细节与边缘信息严重丢失,进而使得分割结果在边界与结构复杂区域出现模糊、误分的情况。

于是,本研究提出边界引导与形状感知的半监督语义分割网络(boundary-guided and shape-aware semi-supervised semantic segmentation network, BS²-Net)。首先,在deeplabV3+(Chen等,2018)的基础上引入高置信度边缘先验,建模目标轮廓特征。接着,利用形状信息优化伪标签,同时根据各类别学习状态自适应调整伪标签筛选标准,提升伪标签质量。该方法在保持自训练框架简洁性与计算高效性的同时,能有效提升伪标签的整体质量与结构一致性。在公开数据集 Potsdam、Vaihingen 和 LoveDA 上的大量对比试验结果表明,本研究提出的方法在不同指标上均取得了较优的性能,可视化结果进一步体现了其在分割边界上的优势。

本研究主要贡献如下:1)提出了形状感知的伪标签优化策略,挖掘特征中隐含的形状信息,有效提升伪标签质量;2)引入基于Canny的边缘优化模块,

通过轮廓先验引导特征学习,从而锐化分割边界;3)设计了动态阈值调整策略,并将阈值更新分成三个阶段,确保伪标签质量持续提升。

1 研究方法

1.1 总体结构

目前主流的半监督语义分割方法通常围绕两种互补的技术路线展开:一致性正则化与自训练。本研究将上述两种方案相结合,将有标签数据 x_l 与无标签数据 x_u 分别经过弱增强与强增强处理,生成对应的弱增强图像 x_l^w, x_u^w 和强增强图像 x_l^s, x_u^s 。随后,将经过增强后的有标签图像与无标签图像同时输入编码器,并使用Canny算子提取其边缘信息。编码器提取到的深度特征与经过边界协同模块后的边缘增强特征进行拼接,再送入解码器以生成初步预测结果 P_l^w, P_u^w 和 P_u^s 。与此同时,模型进一步利用输入图像经编码器提取所得的特征,计算其语义相关性矩阵。随后将该矩阵用于优化解码器输出的预测结果,通过将语义相关性信息注入初步预测中,模型能够生成结构更为一致的预测图 \hat{P}_l^w, \hat{P}_u^w 和 \hat{P}_u^s 。最后,将相关性矩阵中每行蕴含的形状信息用来优化伪标签。整体网络结构如图1所示。

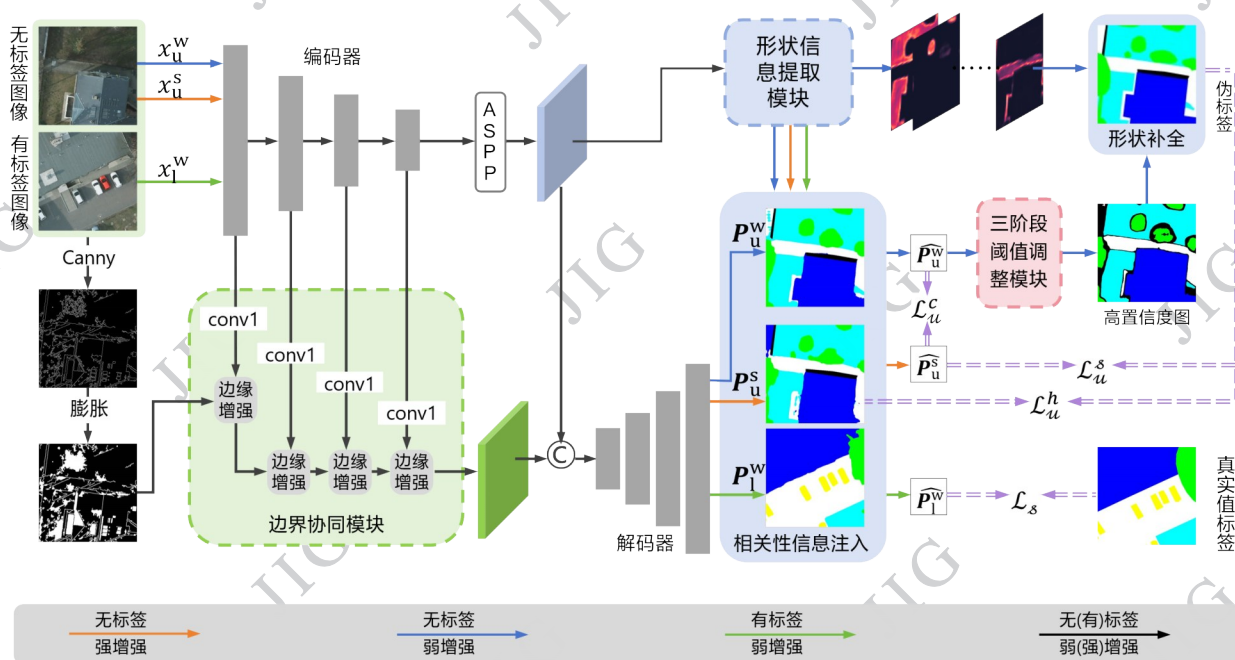


图1 本文方法整体网络结构图

Fig. 1 The overall network structure diagram of the method in this article

本研究引入了三个关键模块:一是边界协同

模块(edge collaboration module, ECM),用于显式提取遥感图像中复杂的地物边界信息,实现目标边界的准确传播;二是形状信息提取模块(shape extraction module, SEM),通过建模形状信息,对置信度低而被过滤的伪标签进行校正,提升伪标签利用率;三是三阶段阈值调整模块(three-stage threshold adjustment module, TAM),充分筛选可靠的伪标签用于模型训练。

1.2 边缘协同模块

高分辨率遥感图像往往场景复杂、地物密集,因此边界信息对实现精确的语义分割至关重要。为提升模型在复杂场景下的分割性能,本研究设计了边缘协同模块,融合 Canny 算子与编码器所得的特征,实现边缘信息的逐层引导与增强。

如图2所示,首先使用 Canny 算子提取输入图像的边缘信息作为结构先验,接着利用形态学膨

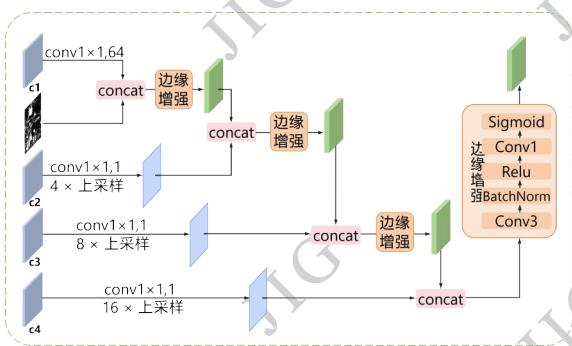


图2 边缘协同模块

Fig. 2 Edge collaboration module

胀操作扩展边缘区域,增强边缘特征。随后,将边缘信息与编码器第一层的输出特征进行拼接,并送入边缘增强模块,引导模型在训练的早期关注边界结构。该过程逐级推进,最终得到融合了多级边缘信息的深层特征。边缘增强模块作为该模块的核心,其内部采用多层特征变换与融合机制。首先用 3×3 的卷积对输入特征进行局部特征提取与变换,增强模型特征表达能力;随后引入批量归一化加速收敛、提升模型稳定性。归一化后的特征再经激活函数进行非线性变换,使网络能学习更复杂的特征模式。然后采用 1×1 的卷积对激活后的特征进行维度调整与融合,以生成紧凑且判别性强的特征表示。最后,通过激活函数将输出映射为边缘概率特征图,指导解码器在后续阶段对预测结果的边缘区域

进行结构校正与优化。

边缘协同模块的加入实现了边缘信息从像素级到语义级的协同传递,使模型在解码过程中能够有效利用结构先验,提升对细小地物与复杂场景边界的判别能力。

1.3 形状信息提取模块

在编码器-解码器结构中,编码器通过多层卷积与下采样提取高层次语义特征,能够输出具有较强的语义表示能力的特征 $F_e \in \mathbb{R}^{C_e \times H \times W}$;而解码器则通过上采样与特征融合逐步恢复空间细节,其输出特征 $F_d \in \mathbb{R}^{C_d \times H \times W}$ 在保留语义信息的同时具备更丰富的空间结构。因此,有效融合两者所得的特征,将有助于生成高质量伪标签,提升半监督模型在遥感图像语义分割中的性能。在伪标签筛选机制上,传统的基于高置信度阈值的策略虽能过滤部分不可靠预测,但也容易忽略中低置信度区域中潜在的正确标签。此类区域常涵盖小目标或样本稀少的类别,其像素间具有语义相关性,却因置信度未达阈值而被舍弃,导致模型在复杂场景下的分割性能受限。

为了解决上述问题,本研究设计了一种形状信息提取模块,通过计算像素间的语义相关性,挖掘其中的形状信息,以此优化伪标签。

如图3所示,模型首先对编码器输出的特征

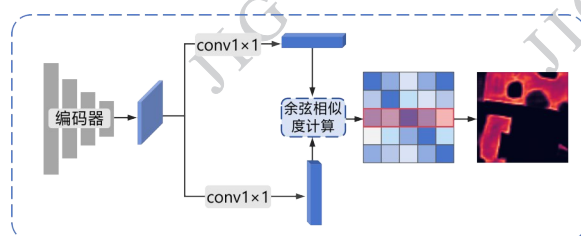


图3 形状信息提取模块

Fig. 3 Shape information extraction module

做线性变换,得到投影特征 $\omega_1, \omega_2 \in \mathbb{R}^{C_d \times HW}$,其中 C_d 是通道数, HW 是特征向量总数。随后,基于投影特征计算像素间语义相似性矩阵 S ,以捕捉图像中同类区域的语义一致性。矩阵 S 中的值越大,表示像素特征越相似,越有可能属于同一语义类别。矩阵 S 的计算方式如下:

$$S = \text{softmax} \left(\frac{F_d^T \cdot F_d}{\sqrt{C_d}} \right) \# (1)$$

在获得相似度矩阵 S 后,从中提取形状信息用以调整伪标签。具体流程如下:首先使用式(2)对相

似度矩阵 S 进行归一化处理,并通过式(3)得到二值化形状区域 \hat{S} ,用于标识具有相似视觉特征的像素对。例如, \hat{S} 中的第 i 行即代表一个未知类别的形状,该行中的值 S'_{ij} 表示第 i 个像素与其他像素之间的相似程度。

$$S' = \frac{S - \min(S)}{\max(S) - \min(S)} \#(2)$$

$$\hat{S}_{ij} = \begin{cases} 1, & \text{if } S'_{ij} > 0.5 \\ 0, & \text{if } S'_{ij} \leq 0.5 \end{cases} \#(3)$$

由于 \hat{S} 中包含的形状信息本身不具备类别属性,本研究设计通过式(5)统计高置信度形状区域 R 中每种类别所占像素数量,确定占比最大的类 k ,认为该形状属于类别 k :

$$R = \hat{P}_w^i \odot M_i \odot \hat{S} \#(4)$$

$$k = \operatorname{argmax}_{l \in L} (\sum_{HW} 1[R = l]) \#(5)$$

上式(4)中的 \hat{P}_w^i 代指伪标签, M_i 代指经过阈值函数筛选所得的高置信度区域。式(5)中的 L 代指伪标签 \hat{P}_w^i 中存在的所有唯一类的集合。为进一步提升形状信息的可靠性,还需对其进行二次筛选,通过计算 k 在高置信度形状区域 R 中的比例 r ,仅当比例 $r > 0.85$ 时,才使用形状信息 \hat{S} 对伪标签 \hat{P}_w^i 进行校正得到校正后的伪标签 \tilde{P}_w^i ,最后更新高置信度区域图 M_i ,如式(6)所示:

$$\tilde{P}_w^i = \begin{cases} k, & \hat{S} = 1 \\ P_w^i, & \hat{S} = 0 \end{cases}, M_i = M_i \cup \hat{S} \#(6)$$

该策略通过形状传播机制,有效扩展了具有明确形状信息与主导类别的高置信度区域,增强伪标签的结构一致性与语义准确性。

1.4 基于类别学习状态的动态阈值调整模块

在半监督语义分割中,由弱增强图像生成的预测结果 \hat{P}_w^i 不能直接用作伪标签,需筛选出高置信度的部分当作伪标签。然而,在类别分布不均的情况下,各类别的收敛速度存在显著差异。于是,本研究针对上述问题设计了一种阶段性动态阈值调整策略。该策略在模型训练的过程中,依据各类别的平均置信度为其分配不同的阈值,确保模型能够稳定均衡地学习每个类别。

具体而言,在训练过程中,首先根据式(7)计算每类的当前平均置信度 μ_c ,在下次迭代中更新阈值。对每一批次数据来说,训练过程中的置信度可

能受噪声或异常样本的影响,导致结果的剧烈波动,于是本研究采用平滑因子 β ,结合历史和当前数据,使每个类别平均置信度的更新过程更加平滑,进而减轻过拟合的风险:

$$\mu_c^{(t)} = \frac{C_c^{(t)}}{N_c^{(t)} + \varepsilon} \#(7)$$

$$C_c^{(t)} = \beta C_c^{(t-1)} + (1 - \beta) \sum_{i=1}^{B_s} \sum_{j=1}^{HW} \sigma_i^u(j) \cdot 1(P_i^u(j) = c) \#(8)$$

$$N_c^{(t)} = \beta N_c^{(t-1)} + (1 - \beta) \sum_{i=1}^{B_s} \sum_{j=1}^{HW} 1(P_i^u(j) = c) \#(9)$$

在上式(7)中, $C_c^{(t)}$ 表示在第 t 次迭代时所有无标注样本属于类别 c 的累积置信度。 $N_c^{(t)}$ 表示对样本中属于类别 c 的累积像素数量。

$$\Delta = (\mu_c^{(t)} - \tau_c^{(t-1)}) \#(10)$$

$$\tau_c^{(t)} = \begin{cases} \tau_c^{(t-1)} + \eta(t) \cdot \Delta, & \text{if } \mu_c^{(t)} > \tau_h \\ \mu_c^{(t)}, & \text{if } \mu_c^{(t)} < \tau_l \\ \tau_c^{(t-1)}, & \text{otherwise} \end{cases} \#(11)$$

计算出各类别的平均置信度后,依据式(11)对各类别阈值进行更新。使用高阈值 τ_h 和低阈值 τ_l 将训练过程分为三个阶段。具体来说,当 μ_c 低于 τ_l 时,表明模型处于对类 c 的探索阶段,此时,为促进模型快速学习该类的特征,应设定较低阈值,纳入更多样本并加速模型捕捉该类的特征。当 μ_c 介于 τ_l 和 τ_h 之间时,表明模型处于学习阶段。此时,模型对部分简单类别已具备较好的识别能力,但对困难类别的置信度仍较低。若此时继续提升阈值,可能加剧模型对高置信度样本的依赖,进而导致确认偏差。因此,为了平衡模型对不同难度类别的学习能力,此阶段中的阈值保持不变。当 μ_c 超过 τ_h 时,表示模型已能较好地识别该类的语义特征,往后阈值可以随着平均置信度同步提升。式(11)中为了确保阈值更新能够反映当前迭代中的置信度变化,引入差值项 Δ , 阈值变化的幅度由调整因子 $\eta(t)$ 控制,其取值范围为 $[0, 1]$, $\eta(t)$ 计算方式如式(12)所示,式中 T_{\max} 表示训练过程的最大迭代次数。随着迭代次数 t 的增加, $\eta(t)$ 逐步增大,从而平滑地调节阈值更新的步长。

$$\eta(t) = \frac{t}{T_{\max}} \#(12)$$

1.5 损失函数

模型训练使用的损失包括监督损失 L_s 和伪监督损失 L_u ,其中监督损失 L_s 如(13)所示,使用交叉熵

失函数计算有标签图像的原始预测结果 P_1^w 以及经过相关性矩阵优化后的结果 \hat{P}_1^w 与真实值标签之间的损失。交叉熵损失公式如式(14)所示,其中 $S(\cdot)$ 代指经过 softmax 函数后的输出结果, C 指总类别数。

$$L_s = \frac{1}{N} \sum_i \left[l_{ce}(P_1^w, y_1) + l_{ce}(\hat{P}_1^w, y_1) \right] \#(13)$$

$$L_{ce}(P_s^i, P_w^i) = - \sum_{c \in C} y_w^i[c] \cdot \log(S(P_s^i[c])) \#(14)$$

$$y_w^i[c] = \text{onehot}(\text{argmax}(S(P_w^i[c]))) \#(15)$$

伪监督损失如式(16)所示,首先计算伪标签 \tilde{P}_u^w 与增强图像原始预测结果 P_u^s 之间的交叉熵损失,用式(17)表示。其次计算一致性损失,用式(18)表示。最后计算伪标签 \tilde{P}_u^w 与经过优化后的增强图像预测结果 \tilde{P}_u^s 之间的交叉熵损失,用式(19)表示,为平衡不同损失在模型训练过程中所起的作用,本研究使用权重 $\lambda_1, \lambda_2, \lambda_3$ 对其进行加权,权值分别设为 $[0.5, 0.25, 0.25]$ 。

$$L_u = \lambda_1 L_u^h + \lambda_2 L_u^s + \lambda_3 L_u^c \#(16)$$

$$L_u^h = \frac{1}{N} \sum_i l_{ce}(P_u^s, \tilde{P}_u^w) \odot M_i \#(17)$$

$$L_u^c = \frac{1}{N} \sum_i \text{KL}(F(x_u^s), F(x_u^w)) \odot M_i \#(18)$$

$$L_u^s = \frac{1}{N} \sum_i l_{ce}(\tilde{P}_u^s, \tilde{P}_u^w) \odot M_i \#(19)$$

2 实验结果与分析

2.1 实验数据集

2.1.1 Vaihingen 数据集

Vaihingen 数据集采集自德国威海根,包含 33 张高分辨率航空图像,尺寸在 1996×1995 像素至 3816×2550 像素之间。涵盖六个语义类别,分别为不透水面、建筑物、低矮植被、树木、汽车和背景。实验选取其中 16 幅图像作为训练集,其余 17 幅作为测试集,以进行模型训练与性能评估。

2.1.2 Potsdam 数据集

Potsdam 数据集采集自德国波茨坦市区,包含 38 幅高分辨率航空影像,每幅影像尺寸为 6000×6000 像素。该数据集与 Vaihingen 数据集具有完全相同的语义类别体系。实验选取其中 24 幅图像作为训练集,其余 14 幅作为测试集。

2.1.3 LoveDA 数据集

LoveDA 数据集是中国城乡土地覆盖数据集,由 5,987 幅高分辨率遥感影像组成,图像尺寸为 1024×1024 像素。涵盖七个语义类别,分别为背景、建筑、道路、水、荒地、森林和农业。

2.2 实验设置

本研究在 Ubuntu 20.04 操作系统下完成,实验配置单张 NVIDIA L40 显卡,使用 Python 3.6 语言及 PyTorch 1.7 深度学习框架。

骨干网络采用 ResNet-101,训练时使用随机梯度下降算法(stochastic gradient descent, SGD),初始学习率设为 0.01,权重衰减设为 $1e-4$,采用动量为 0.9 的学习策略。训练时批大小设为 4,总训练轮数设置为 200。阶段阈值调整策略中阈值区间设为 $[0.4, 0.8]$ 。

图像均统一裁剪至 512×512 ,裁剪重叠率为 256 像素。最终, Vaihingen 数据集共得到 406 张训练图像,若将训练集前 1/8 视作有标签部分,则有标签样本为 51 张,无标签样本为 355 张; Potsdam 数据集共得到 6144 张训练图像。若将训练集前 1/8 视作有标签部分,则有标签样本为 768 张,无标签样本为 5376 张; LoveDA 数据集保持其原始大小 1024×1024 ,共 2522 张训练图像。若将训练集前 1/8 视作有标签部分,则有标签样本为 315 张,无标签样本为 2207 张。

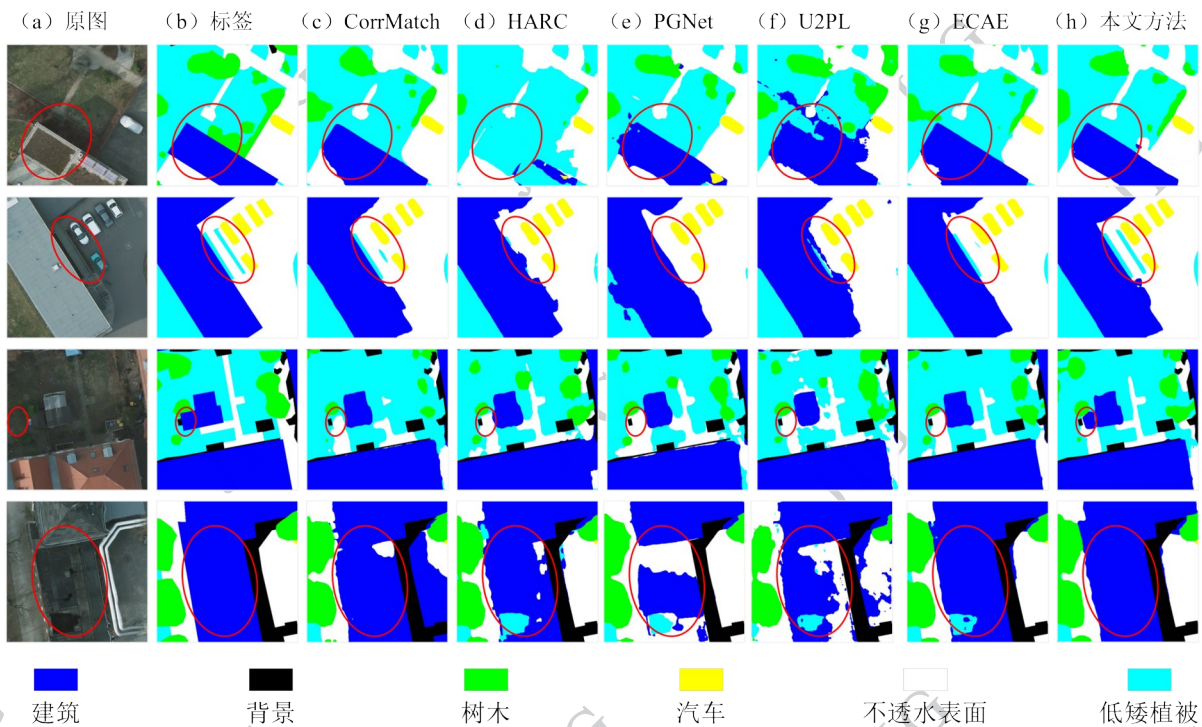
为准确衡量模型性能,本研究中涉及的实验均采用 F1 分数(F1 score, F1)、总体准确率(overall accuracy, OA)和平均交并比(mean intersection over union, mIoU)进行定量评估。

2.3 对比实验

为验证 BS⁵-Net 的有效性,本研究在 Potsdam、Vaihingen 和 LoveDA 数据集上与多个当前主流的半监督语义分割方法进行对比,其中包括 ST++(Yang 等 2022)、U2PL(Wang 等 2022)、UniMatch(Yang 等, 2023)、UNetFormer(Wang 等 2022)、ECAE(Wang 等, 2023)、CorrMatch(Sun 等, 2023)、HARC(Luo 等 2024)以及 PGNet(prior information guided net)(Gan 等, 2025)。

2.3.1 在 Potsdam 数据集上的实验结果

1)定性分析。为综合评估 BS⁵-Net 的有效性,本研究对分割结果进行了可视化,如图 4 所示。整体看来,BS⁵-Net 在不同地物类别间表现良好,误分及漏分的情况较少。在图 4 所示的建筑区域



((a) original images; (b) labels; (c) CorrMatch; (d) HARC; (e) PGNet; (f) U2PL; (g) ECAE; (h) ours)

图4 不同方法在Potsdam数据集上的可视化结果

Fig. 4 Visualization results of different methods on the Potsdam dataset

中, U2PL、PGNet 等对比方法出现了明显的误分, 而 BS⁵-Net 能完整分割建筑轮廓与整体结构, 这得益于形状提取模块能有效捕捉并强化目标的几何特征。此外 CorrMatch、PGNet 等对比方法存在物体漏分, BS⁵-Net 则能清晰界定地物边界, 准确分割细小目标并分离相邻实例, 验证了各模块协同优化的有效性。

总的来说, 相较于其他对比模型, BS⁵-Net 表现出了更强的鲁棒性与更优的分割精度。

2) 定量分析。不同方法在 Potsdam 数据集上分割结果如表 1 所示。在仅使用 1/4 有标签训练样本时, BS⁵-Net 的 mIoU 指标达到 83.43%, 相较于 ST++、U2PL 等方法, 其优势在于引入显式的几何结构先验, 比单纯依赖像素置信度提取伪标签的方式带来更多的提升。当有标签数据降至 1/8 时, BS⁵-Net 的 mIoU 指标为 82.95%, 领先次优模型 3.02%。在标签稀缺的情况下, 对比方法性能衰退明显, 而 BS⁵-Net 能从噪声更大的初始预测中提取出更稳定、更具判别力的结构信息, 证明其具有更优的噪声抑制能力和更高的伪标签利用率。

综上所述, BS⁵-Net 在各项指标上均达到较好性

能。在不同标注比例下与次优模型相比, 其 mIoU 指标分别提高了 1.06% 和 3.02%, 且在 1/8 标注比例下的性能提升比在 1/4 时更为显著。这表明在标注数据匮乏时, BS⁵-Net 具有更明显的优势, 展现了其在低标记场景下的有效性。

2.3.2 在 Vaihingen 数据集上的实验结果

1) 定性分析。从整体分割效果来看, BS⁵-Net 在不同地物类别间保持了良好的一致性, 降低了误分类和区域内部噪声。例如在图 5 所示的植被与建筑交错的复杂场景中, BS⁵-Net 仍能准确界定不同地物边界, 验证了形状提取模块与边界协同模块的有效性。在图 5 中的密集建筑群区域, 对比方法存在漏检, 而 BS⁵-Net 能够克服该问题, 并保持轮廓清晰, 这得益于模型对具有规则形状目标的敏感性。

图 5 中, “轨道”在局部纹理、颜色等视觉特征上与低矮植被较为相似, 模型易因局部特征混淆而将其误判为植被和建筑。BS⁵-Net 能将其准确识别为不透水面, 这得益于形状提取模块能建模目标的结构信息。“轨道”作为人造线性地物, 具有规则的几何形态。SEM 模块能捕获这种全局形状约束, 为网络提供超越局部像素外观的强语义引导, 使其抵抗局

表 1 不同方法在 Potsdam 数据集上的性能比较

Table 1 Performance comparison of different methods on the Potsdam dataset

| 有标签样本数 | 方法 | F1 | | | | | mF1 | OA | mIoU |
|--------|---------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | | 不透水面 | 建筑 | 低矮植被 | 树木 | 汽车 | | | |
| 1/4 | 基线 | 89.49 | 93.93 | 82.25 | 82.49 | 87.13 | 87.06 | 87.64 | 77.36 |
| | UniMatch | 90.52 | 95.44 | 84.86 | 83.39 | 86.48 | 88.14 | 89.11 | 79.07 |
| | UNetFormer | 88.96 | 93.76 | 80.42 | 81.37 | 89.32 | 86.76 | 85.97 | 77.79 |
| | ST++ | 89.35 | 93.92 | 80.98 | 81.82 | 89.52 | 87.12 | 85.65 | 77.53 |
| | U2PL | 90.57 | 95.03 | 84.21 | 83.68 | 87.81 | 88.26 | 88.96 | 79.25 |
| | ECAE | 91.58 | <u>96.32</u> | 81.93 | 82.53 | 88.39 | 88.15 | 89.84 | 78.68 |
| | CorrMatch | <u>92.43</u> | 96.30 | <u>85.38</u> | <u>85.52</u> | <u>90.49</u> | <u>90.04</u> | <u>90.59</u> | <u>82.10</u> |
| | PGNet | 89.00 | 95.80 | 83.21 | 84.32 | 90.18 | 88.50 | 87.13 | 79.22 |
| | HARC | 89.41 | 93.23 | 82.95 | 85.26 | <u>91.16</u> | 88.40 | 87.19 | 79.42 |
| | BS⁵-Net | 92.94 | 96.80 | 86.16 | 86.21 | 91.94 | 90.81 | 91.18 | 83.43 |
| 1/8 | 基线 | 88.51 | 93.27 | 82.45 | 80.30 | 87.10 | 86.33 | 86.70 | 76.23 |
| | UniMatch | 86.53 | 93.82 | 86.01 | 81.09 | 86.43 | 86.77 | 85.84 | 78.24 |
| | UNetFormer | 88.60 | 92.73 | 78.29 | 81.63 | 88.37 | 85.92 | 84.72 | 75.97 |
| | ST++ | 87.50 | 92.45 | 79.79 | 81.42 | 88.49 | 85.93 | 84.39 | 75.63 |
| | U2PL | 89.69 | 94.40 | 84.11 | 83.14 | 87.10 | 87.69 | 88.37 | 78.31 |
| | ECAE | <u>91.33</u> | <u>95.93</u> | 79.76 | 82.00 | 87.41 | 87.29 | <u>89.51</u> | 77.94 |
| | CorrMatch | 90.89 | 94.77 | 84.81 | <u>84.40</u> | 88.66 | <u>88.71</u> | 89.31 | <u>79.93</u> |
| | PGNet | 88.60 | 93.26 | 81.25 | 82.54 | 89.14 | 86.96 | 85.51 | 77.13 |
| | HARC | 89.27 | 93.05 | 81.74 | 83.55 | <u>90.71</u> | 87.66 | 85.72 | 78.30 |
| | BS⁵-Net | 92.68 | 96.75 | <u>85.90</u> | 86.11 | 91.16 | 90.52 | 91.00 | 82.95 |

注:加粗字体表示各列最优结果;“_”表示各列的次优结果。

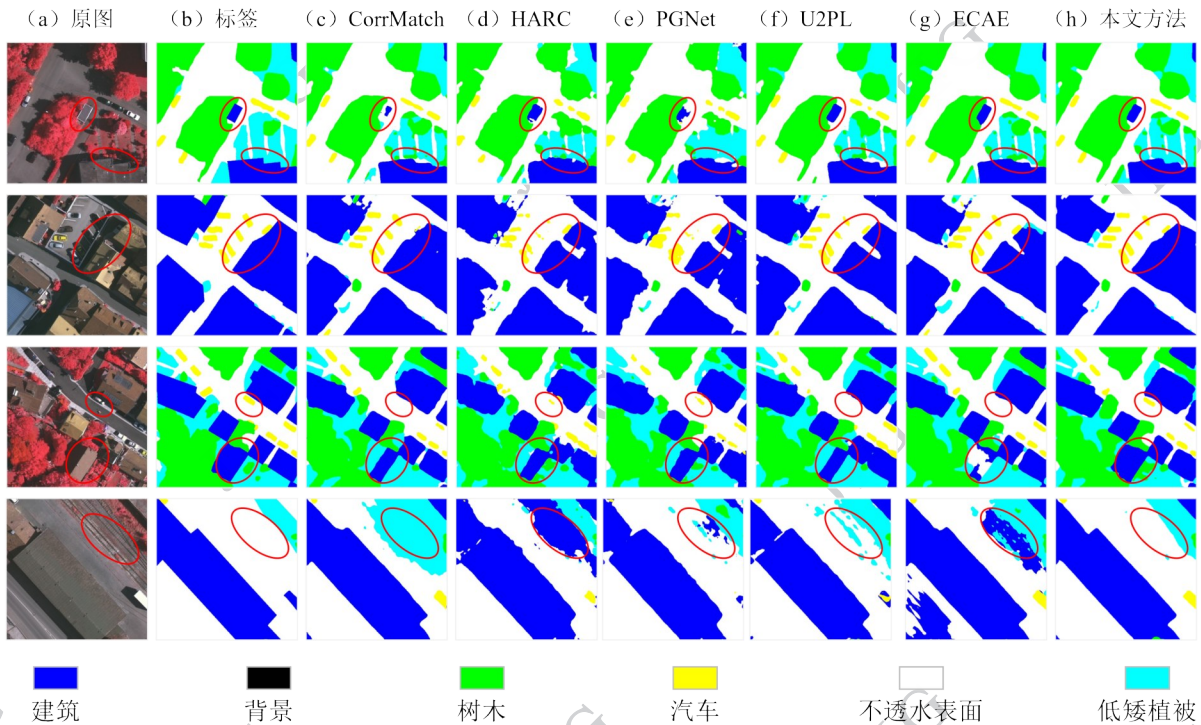
部特征相似性带来的干扰,做出符合地物真实类别的判断。

2)定量分析。不同方法在 Vaihingen 数据集上的定量结果如表 2 所示,在仅使用 1/4 有标签数据的条件下,BS⁵-Net 在 mIoU 指标上达到 75.51%,高于次优模型 0.87%。当标注比例进一步下降至 1/8 时,其 mIoU 达到 74.70%,相对于次优模型的领先优势扩大至 2.12%。这一趋势表明,在标签越稀缺、伪标签噪声越大的半监督场景下,BS⁵-Net 引入的形状与边界先验越能体现出其稳定训练、抑制错误传播的鲁棒性价值。

从整体性能看,BS⁵-Net 在 mIoU、OA 与 mF1 指标上均取得最佳或接近最佳的结果,尤其在“汽车”等小尺度、结构清晰的类别上优势突出。在 1/8 标注比

例下,其 F1 分数较次优模型提升 3.93%。这说明,形状感知模块能够从稀疏的标注中挖掘出有效的结构上下文,而边界引导模块则显著强化了细小目标的轮廓完整性,二者协同作用明显改善了小目标与模糊边界的分割质量。

值得注意的是,BS⁵-Net 在“低矮植被”和“树木”上的 F1 指标值未达最优。我们认为,边界形状先验对于有规则的几何轮廓与结构的类别极为有效。然而,植被类别的判别更依赖纹理模式及阴影关系,这些非几何的、表观层面的特征对形状先验响应较弱。相比之下,U2PL 等方法通过对比学习,能够在特征空间中对像素进行精细的特征理解,能增强模型对外观相似但类别不同的像素的区分能力。然而,这种基于特征拉近推远的优化机制,在处理目标的几



((a) original images; (b) labels; (c) CorrMatch; (d) HARC; (e) PGNet; (f) U2PL; (g) ECAE; (h) ours)

图5 不同方法在Vaihingen数据集上的可视化结果

Fig. 5 Visualization results of different methods on the Vaihingen dataset

何轮廓和锐化目标边界方面,效果不如显式的边界与形状先验有效。

2.3.3 在LoveDA数据集上的实验结果

不同对比方法在LoveDA数据集上的mF1指标和mIoU指标如表3所示。定量结果表明,在仅使用1/4有标签训练样本的条件下,BS⁵-Net在LoveDA数据集上的mIoU指标达到51.88%,高于次优模型2.72%。进一步的,当有标签训练样本数降至1/8时,mIoU指标达到了51.66%,高于次优模型3.26%。

与其他模型相比,BS⁵-Net在整体性能与关键类别分割上均展现出显著优势。从综合指标来看,BS⁵-Net在两组不同有标签样本数设置下的mIoU均位列第一,同时其mF1指标也达到了最高水平。在具体地物类别上,BS⁵-Net在建筑、道路、水体和耕地等关键类别上均取得了最佳分割效果,尤其是在水体与耕地类别上领先幅度明显,这反映了模型对规则地物轮廓与复杂纹理的建模能力。与基线方法相比,BS⁵-Net在两组实验中的mIoU分别提升了5.55%和7.58%,验证了所引入模块的有效性;即使与当前表现较好的HARC方法相比,BS⁵-Net在多数

指标上仍保持领先,进一步体现了其方法的竞争力。值得注意的是,在森林和荒地等类别上,本研究所提出的模型仍有提升空间,这指明了模型未来的优化方向。

总体而言,该实验结果充分证实了BS⁵-Net不仅在整体分割精度上表现优越,而且在一些典型地物类别上具备更精准、稳健的分割能力。BS⁵-Net的整体架构设计能有效适应多样化的遥感场景解析任务,具有一定的场景迁移潜力。

2.4 消融实验

2.4.1 有效性分析

本研究在Vaihingen数据集上采用1/4有标签数据进行消融实验。在基线模型的基础上,逐步引入形状提取模块(SEM)、边界协同模块(ECM)以及三阶段阈值调整模块(TAM),并使用mIoU作为评价指标来验证各模块的有效性。

消融实验定量结果如表4所示,其中“✓”表示添加了该模块,空白表示未添加该模块。

对表4中结果进行分析可知:1)各模块均能有效提升模型性能。具体而言,Baseline+SEM相比Baseline提升8.63%,说明形状信息对于生成可靠的

表 2 不同方法在 Vaihingen 数据集上的性能比较

Table 2 Performance comparison of different methods on the Vaihingen dataset

| 有标签样本数 | 方法 | F1 | | | | | mF1 | OA | mIoU |
|--------|---------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | | 不透水面 | 建筑 | 低矮植被 | 树木 | 汽车 | | | |
| 1/4 | 基线 | 86.55 | 90.36 | 84.95 | 74.69 | 65.99 | 80.51 | 84.54 | 68.28 |
| | UniMatch | 86.39 | 90.65 | 83.49 | 75.23 | 66.95 | 80.54 | 84.17 | 68.24 |
| | UNetFormer | 89.03 | 91.94 | 78.61 | <u>84.69</u> | 80.06 | 84.87 | 85.81 | <u>74.64</u> |
| | ST++ | 88.74 | 92.74 | 79.17 | 84.92 | <u>80.22</u> | <u>85.16</u> | 86.16 | 74.50 |
| | U2PL | 89.11 | <u>92.92</u> | 86.44 | 78.02 | 77.55 | 84.81 | <u>87.04</u> | 74.11 |
| | ECAE | 86.42 | 91.64 | 74.89 | 84.99 | 76.95 | 82.98 | 84.64 | 71.39 |
| | CorrMatch | 87.98 | 92.76 | 85.80 | 78.39 | 73.84 | 83.75 | 86.47 | 72.63 |
| | PGNet | 87.57 | 91.47 | 79.01 | 83.93 | 79.94 | 84.39 | 85.48 | 74.34 |
| | HARC | 88.08 | 91.54 | 78.20 | 84.55 | 75.82 | 83.64 | 85.21 | 72.32 |
| | BS⁵-Net | <u>89.04</u> | 93.88 | <u>86.31</u> | 79.05 | 80.61 | 85.78 | 87.44 | 75.51 |
| 1/8 | 基线 | 85.54 | 88.90 | 84.56 | 72.81 | 70.19 | 80.40 | 83.50 | 67.87 |
| | UniMatch | 85.80 | 90.44 | 84.14 | 74.03 | 66.62 | 80.20 | 83.90 | 67.80 |
| | UNetFormer | 85.82 | 89.34 | 74.17 | 82.09 | 74.17 | 81.11 | 80.79 | 70.16 |
| | ST++ | 86.83 | 90.92 | 74.85 | 81.87 | 75.62 | 82.02 | 83.55 | 70.03 |
| | U2PL | <u>88.65</u> | 92.29 | 86.02 | 76.97 | 74.65 | <u>83.72</u> | <u>86.42</u> | <u>72.58</u> |
| | ECAE | 86.71 | 90.79 | 74.53 | 83.58 | 71.31 | 81.38 | 83.49 | 69.26 |
| | CorrMatch | 88.37 | <u>92.96</u> | 85.65 | 77.10 | 73.44 | 83.50 | 86.40 | 72.34 |
| | PGNet | 87.30 | 91.19 | 77.57 | 81.87 | <u>78.29</u> | 83.24 | 84.45 | 72.14 |
| | HARC | 86.13 | 91.79 | 74.09 | <u>82.84</u> | 71.77 | 81.32 | 80.87 | 68.07 |
| | BS⁵-Net | 89.31 | 93.55 | <u>85.91</u> | 77.71 | 79.55 | 85.21 | 86.94 | 74.70 |

注:加粗字体表示各列最优结果;“_”表示各列的次优结果。

伪标签具有十分重要的指导意义,它能有效纠正伪标签的结构性错误,增强对建筑物、道路等大型规则地物整体轮廓的识别能力,从而改善分割结果的结构完整性与空间连续性。Baseline + ECM 较 Baseline 提升 8.07%,表明边缘先验能够显著增强模型对目标边界的判别能力,从而提高伪标签在物体轮廓处的质量,提升边界分割的清晰度与准确性。Baseline + TAM 相比 Baseline 取得了 5.31% 的性能提升,表明动态阈值调整策略能自适应筛选可靠像素,有效减少噪声标签干扰,从而提升伪标签的质量。综合来看,各模块单独引入均可提升模型的分割性能,验证了其设计的有效性。

2) 模块间存在协同增强效应。在 Baseline + SEM 的基础上,引入 ECM 或 TAM 均能带来额外性能

提升,且效果优于各模块单独使用。其中,Baseline + SEM + ECM 组合的提升较为明显。这表明 SEM 与 ECM 在半监督训练框架中具有较好的互补性。具体来说,SEM 通过挖掘像素间的相关性建模形状信息,维持区域内语义一致性;而 ECM 则利用 Canny 算子提取显式边缘先验,通过多级边缘增强单元注入到特征中,有一定的细化边界的能力。前者提供了“应是何形状”的语义指导,后者提供了“边界在哪”的几何约束,二者形成互补。在训练过程中,ECM 得到的边界特征有助于生成更准确的高置信度区域,SEM 的形状一致性信息则能校正伪标签的结构错误。二者协同优化伪标签,再经阈值调整模块筛选后用于监督训练,推动模型在迭代中不断学习更准确的形状与边缘表征,从而在复杂边界和结构一

表 3 不同方法在 LoveDA 数据集上的性能比较

Table 3 Performance comparison of different methods on the LoveDA dataset

| 有标签样本数 | 方法 | F1 | | | | | | | mF1 | mIoU |
|--------|---------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | | 背景 | 建筑 | 道路 | 水体 | 荒地 | 森林 | 耕地 | | |
| 1/4 | 基线 | 67.23 | 69.55 | 65.46 | 79.04 | 38.19 | 54.49 | 62.35 | 62.33 | 46.33 |
| | UniMatch | 65.90 | <u>75.15</u> | 65.65 | 72.36 | 39.45 | 52.84 | 64.54 | 62.13 | 46.17 |
| | U2PL | <u>68.92</u> | 69.39 | 67.52 | 79.34 | 28.30 | <u>55.24</u> | 64.77 | 61.93 | 47.42 |
| | CorrMatch | 67.97 | 62.70 | 68.99 | <u>80.69</u> | 50.21 | 52.72 | <u>65.41</u> | 64.10 | 47.91 |
| | HARC | 69.52 | 74.49 | <u>69.95</u> | 77.72 | 40.95 | 59.72 | 65.38 | <u>65.39</u> | <u>49.16</u> |
| | BS^s-Net | 67.65 | 76.45 | 70.69 | 82.63 | <u>42.87</u> | 54.15 | 75.53 | 67.14 | 51.88 |
| 1/8 | 基线 | 65.68 | 66.44 | 62.14 | 76.11 | 39.79 | 53.97 | 59.02 | 60.48 | 44.08 |
| | UniMatch | 66.61 | 72.97 | 64.71 | 74.21 | 40.03 | 51.48 | 65.58 | 62.23 | 45.98 |
| | U2PL | <u>67.94</u> | 69.66 | 69.65 | 72.46 | <u>45.49</u> | 56.68 | 68.33 | <u>64.32</u> | 47.00 |
| | CorrMatch | 67.28 | 66.22 | 68.89 | <u>82.25</u> | 42.83 | 53.21 | <u>68.47</u> | 64.16 | 48.32 |
| | HARC | 65.28 | 74.28 | 67.49 | 76.85 | 40.83 | <u>55.53</u> | 63.59 | 63.41 | <u>48.40</u> |
| | BS^s-Net | 71.91 | <u>73.82</u> | <u>68.90</u> | 84.41 | 50.58 | 50.77 | 74.00 | 67.77 | 51.66 |

注:加粗字体表示各列最优结果;“_”表示各列的次优结果。

表 4 不同模块及组合带来的性能增益 (Vaihingen_1/4)

Table 4 Performance gains brought by different modules and their combinations

| SEM | ECM | TAM | mIoU (%) |
|-----|-----|-----|---------------|
| | | | 65.87(基线) |
| ✓ | | | 74.50(8.63 ↑) |
| | ✓ | | 73.94(8.07 ↑) |
| | | ✓ | 71.18(5.31 ↑) |
| ✓ | ✓ | | 75.28(9.41 ↑) |
| ✓ | | ✓ | 75.01(9.14 ↑) |
| | ✓ | ✓ | 74.07(8.20 ↑) |
| ✓ | ✓ | ✓ | 75.51(9.64 ↑) |

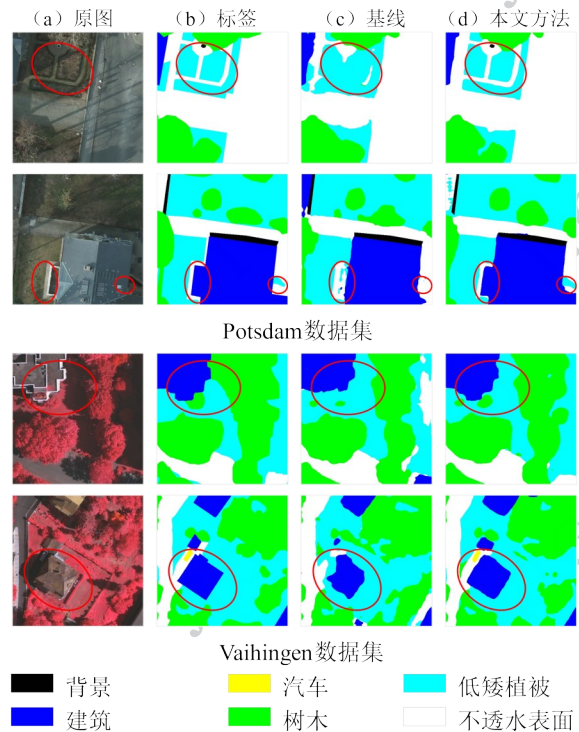
注:“↑”表示 mIoU 指标相较于基线有所提升。

致性上实现有效提升。

2.4.2 消融实验可视化结果

为进一步验证提出的 BS^s-Net 在半监督遥感语义分割任务中的有效性,本研究在 Potsdam 和 Vaihingen 数据集中各选择了部分具有代表性的图像进行可视化分析,结果如图 6 所示。

尽管 DeepLabv3+ 通过其编码器-解码器架构与空洞空间金字塔池化 (atrous spatial pyramid pooling, ASPP) 模块,有效融合了多尺度上下文信息并在自



((a) original image; (b) label; (c) baseline; (d) ours)

图 6 在 Potsdam 和 Vaihingen 上的消融结果

Fig. 6 The ablation results on the Potsdam dataset and Vaihingen dataset

然图像语义分割中取得了卓越性能,但其设计初衷更侧重于处理通用场景中的尺度变化问题。当直接

应用于具有独特挑战的高分辨率遥感图像时,该模型仍存在一定的局限性,其解码器通过简单的特征拼接与上采样恢复细节,未能显式地建模遥感地物中至关重要的几何结构信息(如建筑物的规整边缘、道路的线性轮廓)。这导致在复杂边界区域容易产生平滑效应,难以实现像素级精准定位,而本研究中的边缘形状协同方法对该问题进行了针对性优化。边界协同模块通过引入高置信度边缘先验,为特征学习注入强几何约束,一定程度上锐化了分割边界;形状提取模块(SEM)则通过挖掘像素间的相关性,隐式地建模地物的形状先验,增强了区域内部的一致性。这种针对性的设计带来了性能上的提升。在ISPRS Potsdam与Vaihingen数据集上,可视化结果明显地体现出本方法的优势,相较于DeepLabv3+在建筑物边缘出现的模糊与错误分割现象,本方法的分割结果轮廓更为清晰完整,更贴合地物的真实几何形状。

实验结果表明,针对遥感图像地物结构特性设计的边缘与形状协同优化机制,能有效地解决遥感语义分割中的边界精细化难题,为实现像素级精准分割提供了新的技术途径。

2.4.3 不同阈值区间对实验结果的影响

在三阶段阈值调整模块(TAM)中,低阈值 τ_l 与高阈值 τ_h 是伪标签筛选策略中的两个关键参数。本研究采用步长0.1的方式,在区间内系统测试不同阈值组合, τ_l 的测试区间为[0.3, 0.5], τ_h 的测试区间为[0.7, 0.9],并以mIoU作为评价指标,用以评估不同阈值区间对模型性能的影响。

由表5可知,当 $\tau_l = 0.4$, $\tau_h = 0.8$ 时,模型达到最优性能。我们认为这一组合能有效抑制噪声并提高学习效能。

若 τ_l 设置过低(如<0.3),训练初期将引入过多低置信度噪声样本,导致模型在早期累积难以修正

的误差;若 τ_h 设置过高(如>0.9),则会丢失大量潜在有效样本,延缓模型的学习进程。 $\tau_l = 0.4$ 为训练初期提供了适中的置信度,使模型能够基于形状提取模块(SEM)与边界协同模块(ECM)提供的初始结果,筛选出具有可靠语义信息的伪标签。随着训练推进,SEM与ECM的协同作用能逐步提升模型表征能力,在后期选择 $\tau_h = 0.8$ 作为高阈值,能及时识别出已达到稳定识别状态的类别,为其分配更严格的置信度,从而引导模型将资源集中分配至置信度介于0.4与0.8之间的关键学习区间。该区间涵盖了大量边界模糊或类间相似度高的困难样本,恰好是形状信息与边缘约束能够产生协同作用的核心区域。这能避免训练早期因噪声积累导致的偏差和中后期因样本利用不足造成的学习停滞,不断提升模型在复杂场景下的分割鲁棒性与精度。

3 结论

本研究提出了一种边界引导与形状感知的半监督遥感图像分割网络(BS⁵-Net),旨在解决现有方法中伪标签利用率低及物体边界分割模糊的问题。通过引入形状感知模块建模像素相关性以修正低置信度伪标签,结合边界引导模块利用Canny边缘检测器锐化分割轮廓,并设计三阶段动态阈值策略优化样本筛选,有效提升了伪标签的质量与目标边界分割精度。

在Potsdam、Vaihingen和LoveDA这三个公开数据集上进行的大量实验验证了本研究所提出方法及各组成模块的有效性。可视化分析进一步表明,BS⁵-Net在建筑物、车辆等具有明确结构的地物类别上表现出更高的区域内部一致性和更精细的边界分割结果。

尽管BS⁵-Net在整体性能上取得了较为显著的提升,但在复杂场景中,模型仍会产生局部误分现象,其根本原因在于当前形状建模方式对高度不规则且分布密集的目标表征能力仍然有限。后续研究将尝试通过融合多源遥感数据并引入注意力机制以增强模型的特征表达能力,从而进一步提升模型在复杂场景下的分割鲁棒性。

表5 不同阈值组合的影响

Table 5 The impact of different threshold combinations

| mIoU(%) | High threshold | | | |
|---------------|----------------|-------|--------------|-------|
| | 0.7 | 0.8 | 0.9 | |
| | 0.3 | 74.31 | 65.40 | 60.32 |
| Low threshold | 0.4 | 68.72 | 75.51 | 67.88 |
| | 0.5 | 65.58 | 55.99 | 60.36 |

注:加粗字体表示最优结果

参考文献(References)

- Chen L C, Zhu Y, Papandreou G, Schroff F and Adam H. 2018. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 833 - 851 [DOI: 10.1007/978-3-030-01234-2_49]
- Chen L C, Papandreou G, Kokkinos I, Murphy K and Yuille A L. 2018. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4): 834-848 [DOI: 10.1109/TPAMI.2017.2699184]
- Dong R S, Ma Y Q, Liu Y and Li F Y. 2022. CRNet: class relation network for crop remote sensing image semantic segmentation. *Journal of Image and Graphics*, 27(11): 3382-3394 (董荣胜, 马雨琪, 刘意, 李凤英. 2022. 加强类别关系的农作物遥感图像语义分割. *中国图象图形学报*, 27(11): 3382-3394) [DOI: 10.11834/jig.210760]
- Gan X, Li W, Zhang Y, Long W, Lu Y and Chen Z. 2025. Prior Information-Guided Semi-Supervised Semantic Segmentation of Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, 63: 1-16 [DOI: 10.1109/TGRS.2025.3571066]
- He Y, Wang J, Liao C, Shan B and Zhou X. 2022. ClassHyPer: ClassMix-based hybrid perturbations for deep semi-supervised semantic segmentation of remote sensing imagery. *Remote Sensing*, 14(4): 879 [DOI: 10.3390/rs14040879]
- Jin J, Lu W, Yu H, Rong X, Sun X and Wu Y. 2024. Dynamic and Adaptive Self-Training for Semi-Supervised Remote Sensing Image Semantic Segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 62: #5639814 [DOI: 10.1109/TGRS. 2024. 3407142]
- Ke Z, Qiu D, Li K, Yan Q and Lau R W. 2020. Guided collaborative training for pixel-wise semi-supervised learning// Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 429 - 445 [DOI: 10.1007/978-3-030-58601-0_26]
- Li Z X, Zhang J, Wu J L and Ma H F. 2022. Semi-supervised adversarial learning based semantic image segmentation [J]. *Journal of Image and Graphics*, 2022, 27(7): 2157-2170. (李志欣, 张佳, 吴璟莉, 马慧芳. 基于半监督对抗学习的图像语义分割[J]. *中国图象图形学报*, 2022, 27(7): 2157-2170). [DOI: 10.11834/jig.200600]
- Li Z, Chen H, Wu J, Li J and Jing N. 2023. SegMind: Semi supervised Remote Sensing Image Semantic Segmentation With Masked Image Modeling and Contrastive Learning Method. *IEEE Transactions on Geoscience and Remote Sensing*, 61: #4408917 [DOI: 10.1109/TGRS.2023.3321041]
- Li X H, Xie L L, Wang C F, Miao J H, Shen H F and Zhang L P. 2024. Boundary-enhanced dual-stream network for semantic segmentation of high-resolution remote sensing images. *GIScience & Remote Sensing*, 61(1): 1548-1603. [DOI: 10.1080/15481603.2024. 2356355]
- Liu L M, Zong J X, Xiao Z J, Lan H and Qu H C. 2022. Cross-consistent semantic segmentation algorithm based on manifold regularization[J]. *Journal of Image and Graphics*, 27(12): 3542-3552. (刘腊梅, 宗佳旭, 肖振久, 兰海, 曲海成. 2022. 流形正则化的交叉一致性语义分割算法[J]. *中国图象图形学报*, 27(12): 3542-3552). [DOI: 10.11834/jig.210571]
- Luo Y, Sun B, Li S and Hu Y. 2025. Hierarchical Augmentation and Region-Aware Contrastive Learning for Semi-Supervised Semantic Segmentation of Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, 63: 1-11 [DOI: 10.1109/TGRS. 2024.3521586]
- Long J, Shelhamer E and Darrell T. 2015. Fully convolutional networks for semantic segmentation// Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE: 3431-3440 [DOI: 10.1109/CVPR.2015.7298965]
- Lu C, Zhang X, Du K, Xu H and Liu G. 2024. CTCFNet: CNN-Transformer Complementary and Fusion Network for High-Resolution Remote Sensing Image Semantic Segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 62: #5408717 [DOI: 10.1109/TGRS.2024.3458446]
- Mou L, Hua Y and Zhu X X. 2020. Relation matters: Relational context aware fully convolutional network for semantic segmentation of high resolution aerial images. *IEEE Transactions on Geoscience and Remote Sensing*, 58(11): 7557-7569 [DOI: 10.1109/TGRS.2020. 2979552]
- Ni Y, Liu J, Cui J, Yang Y and Wang X. 2023. Edge Guidance Network for Semantic Segmentation of High-Resolution Remote Sensing Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16: 9382-9395 [DOI: 10.1109/JSTARS.2023.3316307]
- Peng C, Zhang K, Ma Y and Ma J. 2022. Cross fusion net: A fast semantic segmentation network for small-scale semantic information capturing in aerial scenes. *IEEE Transactions on Geoscience and Remote Sensing*, 60: # 5601313 [DOI: 10.1109/TGRS.2021. 3053062]
- 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, Germany: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- Sun B, Yang Y, Zhang L, Cheng M M and Hou Q. 2024. CorrMatch: Label Propagation via Correlation Matching for Semi-Supervised Semantic Segmentation.// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: 3097-3107 [DOI: 10.1109/CVPR52733.2024.00299]

- Sohn K, Berthelot D, Carlini N, Zhang Z, Zhang Het al. 2020. Fix-match: Simplifying semi supervised learning with consistency and confidence [EB/OL].[2025-11-17].
<https://arxiv.org/pdf/2001.07685.pdf>
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones Let al. 2017. Attention is all you need [EB/OL].[2025-11-17].
<https://arxiv.org/pdf/1706.03762.pdf>
- Wang Y, Chen H, Heng Q, Hou W, Savvides Met al. 2023. Freematch: Self-adaptive thresholding for semi supervised learning [EB/OL].[2025-11-17].
<https://arxiv.org/pdf/2205.07246.pdf>
- Wang Y C, Wang H C, Shen Y J, Fei J J, Li W, Jin G Q, Wu L W, Zhao R and Le X Y. 2022. Semi-Supervised Semantic Segmentation Using Unreliable Pseudo-Labels// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, LA, USA: 4238-4247 [DOI: 10.1109/CVPR52688.2022.00421]
- Wang J D, Sun K, Cheng T H, Jiang B R, Deng C R, Zhao Y, Liu D, Mu Y D, Tan M K, Wang X G, Liu W Y and Xiao B. 2021. Deep High-Resolution Representation Learning for Visual Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(10): 3349-3364 [DOI: 10.1109/TPAMI.2020.2983686]
- Wang W, Zhang Y F, Wang X and Li J. 2024. A Boundary Guided Cross Fusion Approach for Remote Sensing Image Segmentation. IEEE Geoscience and Remote Sensing Letters, 21: 1-5 [DOI: 10.1109/LGRS.2024.3350690]
- Wang L B, Li R, Zhang C, Fang S H, Duan C X, Meng X L and Atkinson P M. 2022. UNetFormer: A UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery. ISPRS Journal of Photogrammetry and Remote Sensing, 190: 196-214 [DOI: /10.1016/j.isprsjprs.2022.06.008]
- Miao W, Xu Z, Geng J and Jiang W. 2023. ECAE: Edge-Aware Class Activation Enhancement for Semi-supervised Remote Sensing Image Semantic Segmentation. IEEE Transactions on Geoscience and Remote Sensing, 61: 1-14 [DOI: 10.1109/TGRS. 2023.3330490]
- Xiang W K, Zhou Q, Cui J C, Mo Z Y, Wu X F, Ou W H, Wang J D and Liu W Y. 2024. Weakly supervised semantic segmentation based on deep learning. Journal of Image and Graphics, 29(05): 1146-1168. (项伟康, 周全, 崔景程, 莫智懿, 吴晓富, 欧卫华, 王井东, 刘文予. 2024. 基于深度学习的弱监督语义分割方法综述. 中国图象图形学报, 29(05): 1146-1168). [DOI: 10.11834/jig.230628]
- Yan Y, Deng C, Li, Zhu L K and Ye B. 2023. Survey of image semantic segmentation methods in the deep learning era. Journal of Image and Graphics, 28(11): 3342-3362 (严毅, 邓超, 李琳, 朱凌坤, 叶彪. 2023. 深度学习背景下的图像语义分割方法综述. 中国图象图形学报, 28(11): 3342-3362) [DOI: 10.11834/jig.220292]
- Yang L H, Zhuo W, Qi L, Shi Y H and Gao Y. 2022. ST++: Make Self-training Work Better for Semi-supervised Semantic Segmentation [EB/OL].[2025-11-17].
<https://arxiv.org/pdf/2106.05095.pdf>
- Yang L H, Qi L, Feng L T, Zhang W and Shi Y H. 2023. Revisiting Weak-to-Strong Consistency in Semi-Supervised Semantic Segmentation [EB/OL].[2025-11-17].
<https://arxiv.org/pdf/2208.09910.pdf>
- Zhao H S, Shi J P, Qi X J, Wang X G and Jia J Y. 2017. Pyramid scene parsing network//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 6230-6239 [DOI: 10.1109/CVPR.2017.660]
- Zhang B W, Wu H, Hou W X, HAO, Wang Y Det al. 2022. Flex-Match: Boosting Semi-Supervised Learning with Curriculum Pseudo Labeling [EB/OL].[2025-11-17].
<https://arxiv.org/pdf/2110.08263.pdf>

作者简介

戎誉,女,硕士研究生,主要研究方向为计算机视觉和深度学习及其在遥感图像中的应用。E-mail: ry10121@163.com
杭仁龙,通信作者,男,教授,主要研究方向为模式识别、遥感图像分析和气象大数据智能解译。E-mail: renlong_hang@163.com
林辉,男,博士研究生,主要研究方向为计算机视觉和深度学习及其在遥感图像和气象学中的应用。E-mail: huilin@nuist.edu.cn