基于跨视图上下文感知的高分辨率遥感图像半监督语义分割方法

吕 亮[1],兰 杰[1],兰 猛[2],卢宪凯[3],张乐飞[1*]
(1. 武汉大学计算机学院,湖北武汉 430072;2. 香港科技大学电子与计算机工程系,香港 999077;
3. 山东大学软件学院,山东济南 250101)

摘 要: 高分辨率遥感图像的半监督语义分割旨在利用少量标注样本与大量未标注样本联合训练,从而提升语义分割模型的性能. 此种方法在显著降低人工标注成本的同时,能够充分挖掘未标注数据的潜在价值. 现有方法通常采用将高分辨率遥感图像裁剪为多个子视图的方式进行训练,主要聚焦于同一视图在不同扰动条件下预测结果的一致性. 然而,这类策略往往忽略了不同视图之间的语义与空间关联,限制了模型在标注数据不足时对遥感图像更广泛上下文信息的学习能力. 为此,本文提出了一种基于跨视图上下文感知的高分辨率遥感图像半监督语义分割方法,该方法通过显式建模跨视图之间的上下文交互关系,有效提升伪标签的质量,并引入多重跨视图一致性约束机制,以在更广泛的上下文环境中保持预测结果的一致性. 具体而言,本方法在训练过程中从原始高分辨率遥感图像中采样多个具有重叠区域的视图,包括一个主视图和若干上下文视图,并将这些视图同时输入模型. 进一步设计了空间感知交互融合模块(Spatial-aware Interaction Fusion,SIF),该模块通过交叉注意力与自注意力机制,对不同视图的特征进行交互与融合,生成空间注意力激活图,从而自适应地融合各视图的预测结果,提升伪标签的准确性. 同时,本文提出了多重跨视图上下文一致性约束(Cross-View Context Consistency,CVCC),通过匹配重叠区域的空间位置关系,约束多个视图在重叠区域中的预测结果趋于一致,增强模型对跨视图上下文信息的感知与建模能力,避免因视角变化引发的语义歧义. 为全面评估所提方法的性能,本文基于国际摄影测量与遥感学会提供的 Vaihingen 与 Potsdam 遥感图像语义分割数据集,设置了多种标注比例并进行系统性实验. 实验结果表明,所提出的方法在多种标注比例下均显著优于现有主流半监督语义分割方法. 特别是在仅使用一张标注图像的低标注设定下,相较于监督训练的基线模型,本文方法在 Vaihingen 和 Potsdam 数据集上分别实现了 6.84% 和 12.73% 的 mIoU 提升,充分验证了其在低标注条件下的卓越性能与强泛化能力.

关键词: 遥感;语义分割;半监督学习;跨视图上下文一致性;空间感知交互融合;伪标签
基金项目: 国家自然科学基金(No.62431020)
中图分类号: TP751.1 **文献标识码:** A **文章编号:** 0372-2112(2025)10-3744-15
电子学报 URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20250483

Cross-View Context-Aware Semi-Supervised Semantic Segmentation for High-Resolution Remote Sensing Images

LÜ Liang[1], LAN Jie[1], LAN Meng[2], LU Xian-kai[3], ZHANG Le-fei[1*]
(1. *School of Computer Science*, *Wuhan University*, *Wuhan*, *Hubei* 430072, *China*;
2. *Department of Electronic and Computer Engineering*, *The Hong Kong University of Science and Technology*, *Hong Kong* 999077, *China*;
3. *School of Software*, *Shandong University*, *Jinan*, *Shandong* 250101, *China*)

Abstract: Semi-supervised semantic segmentation of high-resolution remote sensing images aims to leverage a small number of labeled samples together with a large amount of unlabeled data for joint training, thereby enhancing the performance of semantic segmentation models, as this approach not only significantly reduces the cost of manual annotation but also fully exploits the potential value of unlabeled data. Existing methods typically divide high-resolution remote sensing images into multiple sub-views for training, focusing primarily on enforcing prediction consistency under different perturba-

收稿日期:2025-06-06;录用日期:2025-10-09;责任编辑:覃怀银
*通讯作者:张乐飞

tions of the same view. However, such strategies often overlook the semantic and spatial relationships between different views, limiting the model's ability to learn broader contextual information when labeled data are scarce. To address this issue, this paper proposes a cross-view context-aware semi-supervised semantic segmentation method for high-resolution remote sensing images. The proposed approach explicitly models the contextual interactions among multiple views to improve the quality of pseudo labels and introduces a multi-level cross-view consistency constraint to maintain prediction consistency within a broader contextual scope. Specifically, during training, multiple overlapping views—including a primary view and several contextual views—are sampled from the original high-resolution image and jointly fed into the model. A spatial-aware interaction fusion (SIF) module is designed to perform cross-view feature interaction and fusion via cross-attention and self-attention mechanisms. This module generates spatial attention activation maps that adaptively fuse the predictions from different views, thereby improving pseudo label accuracy. In addition, a multiple cross-view context consistency (CVCC) mechanism is introduced to enforce consistent predictions in overlapping regions by aligning their spatial correspondences. This constraint enhances the model's ability to perceive and model cross-view contextual information, mitigating semantic ambiguity caused by view variations. To comprehensively evaluate the proposed method, extensive experiments are conducted on the Vaihingen and Potsdam datasets provided by the International Society for Photogrammetry and Remote Sensing, under various labeling annotation ratios. Results show that the proposed method consistently outperforms state-of-the-art semi-supervised segmentation approaches. In particular, under an extremely low-label setting using only one labeled image, it achieves 6.84% and 12.73% mIoU improvements over the supervised baseline on Vaihingen and Potsdam, respectively, validating its superior performance and strong generalization under limited annotation.

Key words: remote sensing; semantic segmentation; semi-supervised learning; cross-view context consistency; spatial-aware interaction fusion; pseudo-label

Foundation Item(s): National Natural Science Foundation of China (No.62431020)

1　引言

遥感图像语义分割作为计算机视觉与遥感技术交叉领域的核心任务,旨在为每个像素分配语义类别标签[1](如建筑物、道路、植被等). 该任务在城市建设、土地规划、生态环境监测、灾害评估等关键应用场景中具有广泛的应用价值. 随着卫星遥感技术的不断发展,获取高分辨率遥感图像变得日益容易,而基于数据驱动的深度学习技术在遥感图像语义分割领域也取得了显著进展[2]. 然而,在实际应用中,仍然面临许多挑战,尤其是高分辨率遥感图像的像素级标注过程需要耗费大量专业人力,这在很大程度上制约了监督学习方法的可扩展性.

为应对标注数据不足的挑战,半监督学习[3]作为一种能够有效利用未标注数据的技术,结合了少量标注数据与大量未标注数据,显著提升了模型性能,并减少了对标注数据的依赖. 在计算机视觉领域,半监督学习最早被应用于图像分类任务,诸如协同训练[4]、半监督字典学习[5]等方法均属于早期的研究方向. 随着深度学习技术的快速发展,新的半监督学习方法不断涌现并不断完善,特别是基于自训练[6]和一致性学习[7]的方法. 其中,一致性学习驱动的半监督方法逐渐成为研究的热点. 通过数据增强和一致性约束来优化训练过程,这些方法能够从未标注数据中有效挖掘潜在信息,并在图像分类、语义分割等领域取得显著进展.

在此背景下,遥感图像解译领域日益重视利用未标注数据的潜力[8,9],积极推动了高分辨率遥感图像半监督语义分割的发展,以期在有限的人工标注资源下提升模型的泛化能力与表达能力. 已有方法如基于双教师-学生结构的一致性方法[10]、结合熵图的自适应重加权策略[11]以及基于高质量伪标签的加权学习法[12]等,都在一定程度上缓解了因标注不足而带来的性能瓶颈. 然而,它们大多依赖于将高分辨率遥感图像裁剪为多个独立的子视图,并关注同一子视图在不同扰动下的预测一致性,忽略了图像不同区域之间的长距离上下文依赖关系,这种局限性使得模型在标注数据稀缺的情况下,难以有效建模和理解全局语义信息.

为了解决上述问题,本文提出了一种基于跨视图上下文感知的半监督语义分割方法,旨在更有效地挖掘和利用高分辨率遥感图像中的丰富语义上下文信息,提升模型在有限标注条件下的分割性能. 具体而言,该方法在每次训练迭代过程中,从原始高分辨率遥感图像中随机采样多个具有重叠区域的视图,并引入空间感知交互融合模块 SIF(Spatial-aware Interaction Fusion),通过交叉注意力和自注意力机制建模各视图间的上下文依赖关系,生成空间注意力激活图并自适应融合各个视图的伪标签,从而提升伪标签的质量,增强训练监督信号. 此外,本文设计多重跨视图上下文一致性约束 CVCC(Cross-View Context Consistency),通过对具有空间重叠区域的多视图预测结果进行语义对齐,鼓励模型在不同视角下对相同区域保持预测一致

性,这种机制有效缓解了标注数据不足带来的训练困难,促进了模型对更广泛的上下文信息的学习.本文的主要贡献如下:

(1)提出了一种基于跨视图上下文感知的高分辨率遥感图像半监督语义分割方法,旨在解决标注信息不足导致模型对上下文信息学习不充分的问题.

(2)设计空间感知交互融合模块,用于自适应融合不同视图的伪标签,提升伪标签质量;同时引入多重跨视图一致性约束,有效提升伪标签质量和上下文建模能力.

(3)在 Vaihingen 和 ISPRS Potsdam 两个遥感语义分割数据集上进行了广泛的实验验证,实验结果表明,所提出方法在多种标注比例设定下均显著优于现有方法.

2 相关工作

2.1 半监督语义分割

半监督学习是一种旨在结合少量标注数据与大量未标注数据来提升模型性能的机器学习范式.该方法在深度学习兴起之前就已受到广泛关注和研究,而随着深度神经网络的发展,近年来更是涌现出许多创新性的半监督学习策略,并广泛应用于图像分类和语义分割等任务中.基于现有研究,当前的半监督语义分割方法主要基于伪标签和一致性正则化两种方法展开.

伪标签[6]法先用标注数据训练一个模型,然后对未标注样本生成伪标签,再将这些伪标签与标注数据一起用于后续训练,提高模型的泛化能力.一致性正则化是通过约束模型对输入数据的微小扰动保持预测一致性来优化模型的方法.这类方法鼓励模型在相同输入经过不同的数据增强处理后产生一致的输出,从而提高模型的分割性能.近期,结合伪标签与一致性正则化的策略在半监督语义分割中取得显著进展.典型方法如 FixMatch[13]采用弱到强的一致性策略,从弱增强的未标注图像生成伪标签,用于监督强增强对应的未标注图像.UniMatch[14]采用基于 FixMatch[13]的框架,通过增加特征扰动分支和强增强分支,进一步提升了模型性能,这种简洁而强大的方法在近期的半监督学习研究中得到了广泛应用.与上面的方法不同,CorrMatch[15]重视并利用了相关图在建模位置对之间的关系上的作用,并通过设计两种标签传播策略提高了未标注数据的使用效率和模型性能.AllSpark[16]采用通道交叉注意力机制,能够从无标签数据中增强有标签特征,并引入了语义记忆和通道语义分组策略,有效避免了无关信息的干扰,提升了模型的准确性和效率.ScaleMatch[17]通过引入混合双尺度伪标签和尺度一致性学习,有效提升了半监督语义分割中对多尺度信息的建模能力.

2.2 高分辨率遥感图像半监督语义分割

针对高分辨率遥感图像标注成本高的问题,研究人员结合遥感数据特性,提出了多种半监督分割方法.现有方法可根据技术特点分为以下几类.

基于数据增强的方法:RanPaste[18]提出将标注图像中的区域粘贴至未标注图像并进行混合,以提升样本多样性并缓解伪标签噪声.然而,此类方法主要依赖低层图像操作,未能有效利用高层次语义信息.

基于自训练的方法:DAST[19]提出了动态自适应自训练框架,结合动态伪标签采样、分布对齐和自适应阈值机制,以应对遥感图像中常见的数据分布差异与类别不平衡问题.但此类方法易受初始预测偏差影响,错误累积会误导模型的训练.

基于一致性学习的方法:ICNet[10]构建了一种双教师-学生的迭代训练结构,通过在不同网络之间施加预测一致性约束以提升模型鲁棒性.此类方法强调在多视图或多扰动下保持输出一致,但通常只在单视图上建立约束,对长程空间上下文信息学习不足.

基于掩码图像建模的方法:MIMSeg[20]将掩码重建任务引入半监督语义分割,通过重构被遮蔽的图像区域促使模型学习更具判别性的表征.该方法有助于局部上下文理解,难以建模高分辨率遥感图像的跨区域空间关联.

基于对比学习的方法:Xin 等人[21]通过在特征空间中施加对比约束,增强模型对类内差异和类间相似性的鲁棒性.对比学习在密集预测任务中能提升表征质量,但通常计算代价较高且像素级对比难以高效建模全局上下文.

混合方法:SegMind[22]将一致性、掩码重建与对比学习融合,以期同时强化局部与整体表征.尽管能综合各类优势,但会增加方法的计算复杂度,同时仍缺乏对高分辨率遥感图像的长程上下文的学习.

尽管上述方法从不同角度推进了高分辨率遥感图像半监督语义分割的发展,它们大多基于一个共同的训练范式:将高分辨率遥感图像裁剪为独立子图像进行训练,仅侧重于单视图层面的一致性或伪标签生成,未能显式地建模不同视图或裁剪区域之间的空间语义交互与对齐关系.这种训练方式导致模型难以在标注稀缺的情况下对复杂遥感场景的整体语义理解与推理能力.因而亟需一种能够充分利用更广泛的上下文依赖的新方法.基于此,本文提出一种跨视图上下文感知的半监督语义分割方法,通过显式建模具有空间重叠的多视图之间的上下文依赖,利用上下文交互模块与多重跨视图上下文一致性约束,充分挖掘高分辨率遥感图像的长程上下文,并在有限标注条件下提升模型的泛化能力与分割性能.

2.3 跨视图学习

跨视图学习通过利用来自不同视角或图像区域的信息,提高模型的表示与泛化能力,这在监督、自监督和半监督任务中都表现出重要价值. 在半监督语义分割中,核心思路是建模不同视图之间的语义一致性,以实现对未标注数据的更高效利用.

已有工作主要可分为两类策略:一类基于同一图像的不同增强视图的特征约束(如CCVC[23]、AGCV[24]、CrossMatch[25]等),它们通常通过拉开或聚拢分支间的特征距离来强化表征,未能考虑不同空间位置视图之间的语义对齐;另一类侧重于利用具有重叠区域的局部裁剪视图来建立局部语义一致性约束,例如CAC[26]提出方向性对比损失以拉近重叠区域的特征表达,CWC[27]提出了基于跨窗口一致性的渐进学习框架和有偏跨窗口一致性损失,明确约束来自同一图像不同裁剪窗口在重叠区域内的语义一致性. 尽管这些方法在利用重叠区域的上下文信息方面取得进展,但仍存在不足:CAC[26]仅在两个重叠视图之间构建对比损失,未能充分覆盖多视图的互信息;CWC[27]虽然约束了多个重叠视图的重叠区域一致性,但常将视图独立处理,缺乏对具有重叠关系的视图之间的显式交互机制来实现信息互补与深度融合.

受视频处理中的帧间交互[28]启发,本文从高分辨率遥感影像中采样主视图(当前帧)与上下文视图(参考帧),并设计空间感知交互融合模块对不同视图特征进行交互与融合,使视图间语义依赖得到显式建模,从而有效提升伪标签的质量. 与此同时,我们引入多重跨视图上下文一致性约束,通过主视图与上下文视图的裁剪方式,系统性地构建并约束多个重叠区域间的语义一致性,增强模型对上下文信息的感知与利用. 相比已有方法,本文的方法在视图构建上更贴合遥感影像的空间特性,并通过视图间的交互与融合机制实现从"视图独立"向"视图交互",从"局部约束"向"全局感知"的跨越,为高分辨率遥感图像的半监督语义分割提供了新的研究视角与技术路径.

3 基于跨视图上下文感知的高分辨率遥感图像半监督语义分割

3.1 方法概述

在半监督语义分割中,训练数据涵盖了标注数据与未标注数据. 标注数据集可表示为$\{\boldsymbol{x}_i^{\mathrm{l}},\boldsymbol{y}_i^{\mathrm{l}}\}_{i=1}^{B_{\mathrm{l}}}$,其中$\boldsymbol{x}_i^{\mathrm{l}}\in\mathbb{R}^{H\times W\times 3}$表示标注遥感图像,以及$\boldsymbol{y}_i^{\mathrm{l}}\in\mathbb{R}^{H\times W\times K}$是与之对应的真实标签,标签具有$K$个不同类别. 未标记数据集表示为$\{\boldsymbol{x}_j^{\mathrm{u}}\}_{j=1}^{B_{\mathrm{u}}}$,其中$\boldsymbol{x}_j^{\mathrm{u}}\in\mathbb{R}^{H\times W\times 3}$表示为未标注的图像,$B_{\mathrm{l}}$和$B_{\mathrm{u}}$分别表示标注图像和未标注图像的数量. 根据从弱到强的一致性学习框架,每个未标记的遥感图像$\boldsymbol{x}_j^{\mathrm{u}}$首先会经过弱数据增强$A_{\mathrm{w}}(\boldsymbol{x}_j^{\mathrm{u}})$,从而获得弱增强图像$\boldsymbol{x}_j^{\mathrm{u_w}}$. 弱增强通常指一些轻量级的图像变换操作,包括随机缩放、裁剪、旋转和翻转等. 在此基础上,$\boldsymbol{x}_j^{\mathrm{u_w}}$进一步经过强数据增强$A_{\mathrm{s}}(\boldsymbol{x}_j^{\mathrm{u_w}})$,得到强增强图像$\boldsymbol{x}_j^{\mathrm{u_s}}$. 强增强会引入更大幅度的扰动,例如随机遮挡、颜色空间变换、灰度化、高斯模糊等. 整体的优化目标函数可表示为

$$L=L_{\mathrm{s}}+\lambda_{\mathrm{u_s}}L_{\mathrm{u_s}} \tag{1}$$

其中,有监督损失和无监督损失各自表示为L_{s}和$L_{\mathrm{u_s}}$,$\lambda_{\mathrm{u_s}}$是调整$L_{\mathrm{u_s}}$权重的超参数. 有监督损失L_{s}和无监督损失$L_{\mathrm{u_s}}$分别表示为

$$L_{\mathrm{s}}=\frac{1}{B_{\mathrm{l}}}\sum_{i=1}^{B_{\mathrm{l}}}L_{\mathrm{ce}}(\boldsymbol{y}_i^{\mathrm{l}},\boldsymbol{p}_i^{\mathrm{l}}) \tag{2}$$

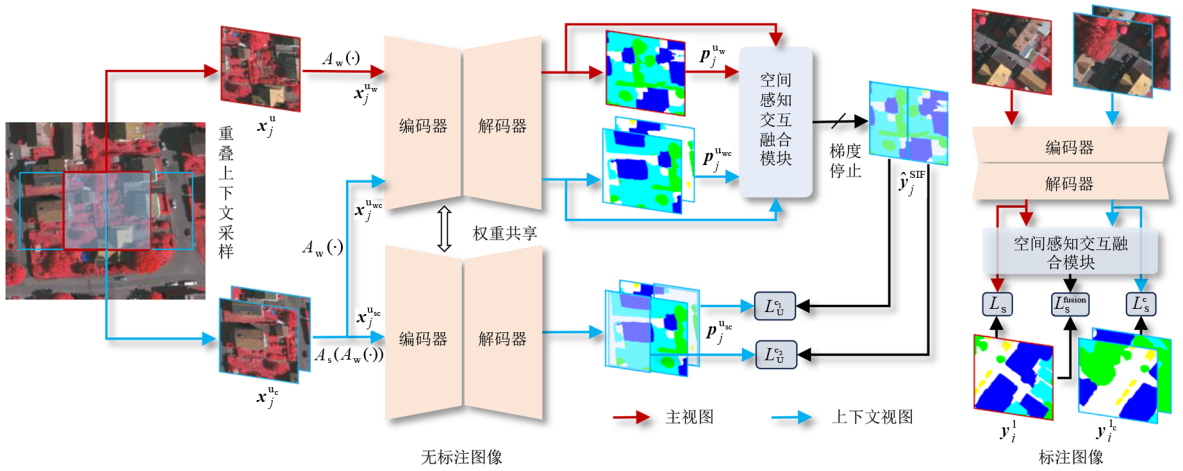
$$L_{\mathrm{u_s}}=\frac{1}{B_{\mathrm{u}}}\mathbb{I}\left(\max(\boldsymbol{p}_j^{\mathrm{u_w}})\geqslant\tau\right)L_{\mathrm{ce}}(\hat{\boldsymbol{y}}_j,\boldsymbol{p}_j^{\mathrm{u_s}}) \tag{3}$$

其中,τ是置信度阈值;$\boldsymbol{p}_j^{\mathrm{u_w}}=f(\boldsymbol{x}_j^{\mathrm{u_w}})$和$\boldsymbol{p}_j^{\mathrm{u_s}}=f(\boldsymbol{x}_j^{\mathrm{u_s}})$分别表示预测概率图;$\hat{\boldsymbol{y}}_j=\mathrm{argmax}(\boldsymbol{p}_j^{\mathrm{u_w}})$是生成的伪标签. 这里的$f(\cdot)$代表分割模型.

本文基于弱到强一致性学习框架,深入探讨了如何通过引入上下文一致性约束,有效挖掘遥感图像在不同区域之间潜在的上下文关系. 所提出方法的整体框架如图1所示,该方法创新地引入了重叠上下文采样和跨视图上下文分支. 对于无标注图像,首先通过重叠上下文采样策略,从高分辨率遥感图像$\boldsymbol{x}_{\mathrm{origin}}^{\mathrm{u}}$中采样获得主视图和多个上下文视图. 采样过程中,依据预设的上下文视图数量及与主视图的重叠比例,从$\boldsymbol{x}_{\mathrm{origin}}^{\mathrm{u}}$裁剪得到主视图$\boldsymbol{x}_j^{\mathrm{u}}$和上下文视图$\boldsymbol{x}_j^{\mathrm{u_c}}$;接着,主视图和上下文视图均经过弱增强,分别得到弱增强主视图$\boldsymbol{x}_j^{\mathrm{u_w}}$和弱增强上下文视图$\boldsymbol{x}_j^{\mathrm{u_{wc}}}$,被分别送入编码器和解码器进行特征提取得到特征$\boldsymbol{e}_j^{\mathrm{u_w}}$、$\boldsymbol{e}_j^{\mathrm{u_{wc}}}$,并且从分割头中获取对应主视图的预测结果$\boldsymbol{p}_j^{\mathrm{u_w}}$、上下文视图的预测结果$\boldsymbol{p}_j^{\mathrm{u_{wc}}}$. 随后,将特征$\boldsymbol{e}_j^{\mathrm{u_w}}$、$\boldsymbol{e}_j^{\mathrm{u_{wc}}}$连同预测结果$\boldsymbol{p}_j^{\mathrm{u_w}}$、$\boldsymbol{p}_j^{\mathrm{u_{wc}}}$输入到空间感知交互融合模块得到融合的预测结果$\hat{\boldsymbol{y}}_j^{\mathrm{SIF}}$;最后,对弱增强上下文视图$\boldsymbol{x}_j^{\mathrm{u_{wc}}}$实施强增强,进一步得到$\boldsymbol{x}_j^{\mathrm{u_{sc}}}$,由融合后的伪标签$\hat{\boldsymbol{y}}_j^{\mathrm{SIF}}$监督模型对强增强上下文视图$\boldsymbol{x}_j^{\mathrm{u_{sc}}}$的预测,从而确保模型在不同上下文输入下的一致性. 对于有标注图像,处理流程与无标注图像类似,同样采用重叠上下文采样获取主视图$\boldsymbol{x}_i$和上下文视图$\boldsymbol{x}_i^{\mathrm{l_c}}$,再依次经过编码器、解码器和空间感知交互融合模块分别得到主视图的预测结果$\boldsymbol{p}_i^{\mathrm{l}}$,融合后的预测结果$\boldsymbol{p}_i^{\mathrm{l_{SIF}}}$以及上下文视图的预测结果$\boldsymbol{p}_i^{\mathrm{l_c}}$.

3.2 重叠上下文采样

重叠上下文采样旨在生成与主视图存在一定重叠



注:该方法首先通过重叠上下文采样生成主视图(红色)和上下文视图(蓝色),并对这些视图进行弱增强后输入网络以获得初步预测结果;随后,利用空间感知交互融合模块对伪标签进行优化,并将优化后的伪标签作为监督信号,用于监督强增强上下文视图的预测结果,从而实现跨视图的上下文一致性学习.

图1 基于跨视图上下文感知半监督语义分割方法的整体结构图

区域的上下文视图,这些视图不仅与主视图有部分区域重合,还覆盖图像中的不同局部区域,从而增强模型对空间上下文关系的建模能力. 具体而言,执行重叠上下文采样时,首先需要确定两个关键参数:上下文视图的数量N,上下文视图与主视图的重叠区域最小比例R. 其中,N决定了从原始图像中获取的上下文视图的数量,增加N能够获取更加全面的上下文信息,但同时也会增加计算成本. R则控制着上下文视图和主视图之间的重叠程度,合适的重叠比例有助于模型捕捉相邻区域间的关联信息,避免信息丢失.

在参数设定完成后,在高分辨率遥感图像$\boldsymbol{x}^{\mathrm{u}}_{\mathrm{origin}}$随机选取一个位置裁剪得到主视图$\boldsymbol{x}^{\mathrm{u}}_{j}\in\mathbb{R}^{H\times W\times 3}$,这里记为第$k$次裁剪,并记录该主视图在原始图像中的左上角坐标$X_k,Y_k$表示为

$$\boldsymbol{x}^{\mathrm{u}}_{j},X_k,Y_k=\mathrm{Crop}_k\left(\boldsymbol{x}^{\mathrm{u}}_{\mathrm{origin}},[H,W]\right) \tag{4}$$

通过预先设定的超参数上下文视图的数量、重叠区域比例,以及式(4)的主视图坐标,对$\boldsymbol{x}^{\mathrm{u}}_{\mathrm{origin}}$进行随机上下文裁剪,得到第$k$个上下文视图$\boldsymbol{x}^{\mathrm{u_c}}_{jk}$,上下文视图相对于主视图的重叠区域的掩码$\mathbf{Overlap}^{\mathrm{contex}}_{jk}$,以及主视图相对于上下文视图的重叠区域的掩码$\mathbf{Overlap}_{jk}$,其中$k=1,2,\cdots,N$,共进行$N$次裁剪,$\boldsymbol{x}^{\mathrm{u_c}}_{j}$表示将裁剪的$N$个上下文视图在批次维度拼接的结果,计算过程如下所示:

$$\begin{aligned}&\boldsymbol{x}^{\mathrm{u_c}}_{jk},\mathbf{Overlap}^{\mathrm{contex}}_{jk},\mathbf{Overlap}_{jk}\\&=\mathrm{NCrop}_k\left(\boldsymbol{x}^{\mathrm{u}}_{\mathrm{origin}},[H,W],[X_k,Y_k],R\right)\end{aligned} \tag{5}$$

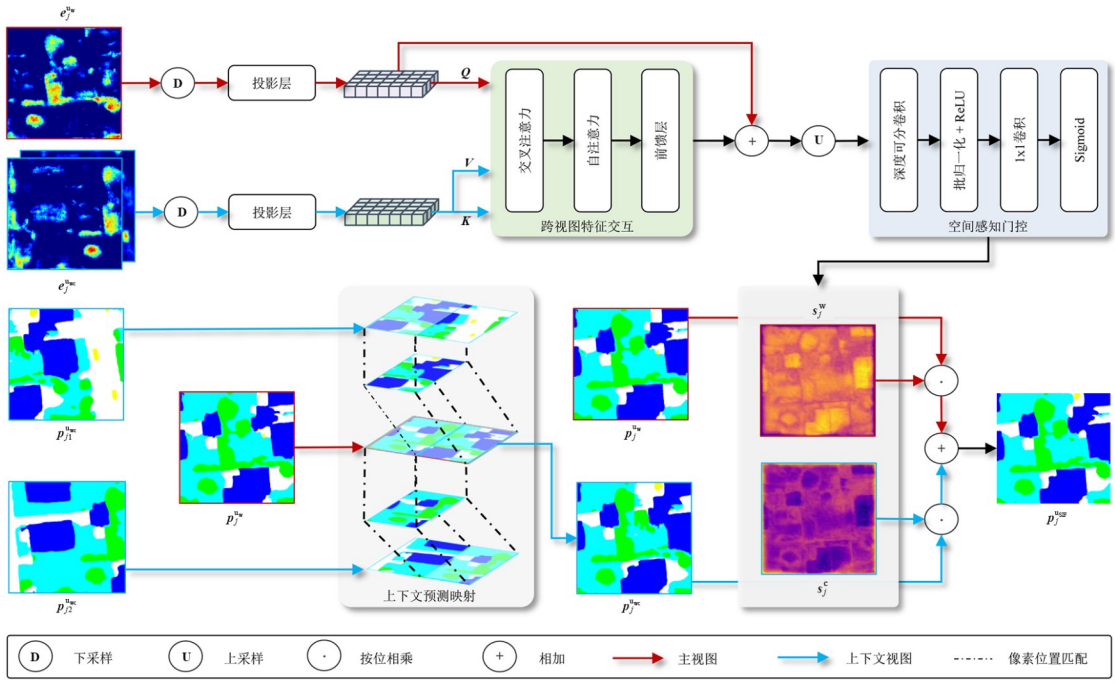
其中,近邻随机裁剪NCrop_k(Near Random Crop)指在主视图位置邻近区域进行第k次随机裁剪,保证与主视图$\boldsymbol{x}^{\mathrm{u}}_{j}$重叠区域比例至少为$R$,并且每次裁剪确保$N$个上下文视图能够完全覆盖主视图. $\boldsymbol{x}^{\mathrm{u}}_{j}$和$\boldsymbol{x}^{\mathrm{u_c}}_{j}$均经过弱增强,分别得到弱增强主视图$\boldsymbol{x}^{\mathrm{u_w}}_{j}$和弱增强上下文视图$\boldsymbol{x}^{\mathrm{u_{wc}}}_{j}$,对弱增强上下文视图$\boldsymbol{x}^{\mathrm{u_{wc}}}_{j}$实施强增强,进一步得到$\boldsymbol{x}^{\mathrm{u_{sc}}}_{j}$. 对于有标注图像,同样按照上述流程来得到上下文视图. 本文仅使用真实标签监督和更新空间感知交互融合模块,提升伪标签融合效果.

3.3 空间感知交互融合模块

在语义分割任务中,结合上下文信息进行推理能够显著提升分割结果的质量. 通过引入跨视图上下文分支并从不同上下文获得预测结果进行融合,能够生成高质量的伪标签. 不同于传统方法通过单个视图的预测来生成结果,本文旨在通过为不同上下文视图的预测生成空间注意力图,依据注意力图对不同上下文的预测进行加权融合,使更准确的预测获得更高权重,生成更为精确的伪标签.

为此,本文提出了空间感知交互融合模块,该模块主要包含跨视图特征交互和空间感知门控. 在训练过程中,该模块使用来自不同上下文的特征图来学习每个预测的空间激活图. 空间感知交互融合模块如图2所示,其计算流程如下文所示.

弱增强主视图$\boldsymbol{x}^{\mathrm{u_w}}_{j}$和弱增强上下文视图$\boldsymbol{x}^{\mathrm{u_c}}_{j}$输入模型$f(\cdot)$提取各自的特征图$\boldsymbol{e}^{\mathrm{u_w}}_{j}$和$\boldsymbol{e}^{\mathrm{u_{wc}}}_{j}$,其中$\boldsymbol{e}^{\mathrm{u_w}}_{j}$为弱增强图像特征,$\boldsymbol{e}^{\mathrm{u_{wc}}}_{j}$为弱增强上下文图像特征,$H_{\mathrm{f}}$和$W_{\mathrm{f}}$是特征图的高度和宽度. $\boldsymbol{e}^{\mathrm{u_w}}_{j}$和$\boldsymbol{e}^{\mathrm{u_{wc}}}_{j}$首先使用下采样操作以压缩特征,从而降低计算开销. 压缩后的特征通过投影层进行变换,得到弱增强图像的特征表示,作为查询(Query,Q),同时,将上下文图像的特征分别作为键(Key,K)和值(Value,V),实现弱增强视图特征与弱增强



注:该模块首先将主视图特征与上下文视图特征输入跨视图特征交互模块,实现特征间的信息交互;随后经过空间感知门控模块生成空间激活图;最后利用该空间激活图自适应地融合不同视图的预测结果,从而获得更精细的伪标签.

图2　空间感知交互融合模块的结构图

上下文视图特征之间的信息交互:

$$\begin{aligned}\boldsymbol{Q}&=\mathrm{proj}\left(\mathrm{Down}(\boldsymbol{e}_j^{\mathrm{u_w}})\right),\\ \boldsymbol{K}=\boldsymbol{V}&=\mathrm{proj}\left(\mathrm{Down}(\boldsymbol{e}_j^{\mathrm{u_{wc}}})\right)\end{aligned}\tag{6}$$

该交互过程依次经过交叉注意力机制、自注意力机制以及前馈神经网络层,以增强特征表达能力. 最后对融合后的特征进行上采样,并将特征恢复至原始尺寸,得到最终的上下文交互特征$\boldsymbol{e}_j^{\mathrm{u_l}}$:

$$\boldsymbol{e}_j^{\mathrm{u_l}}=\mathrm{Up}\left(\mathbf{Q}+\mathrm{FFN}\left(\mathrm{SA}\left(\mathrm{CA}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})\right)\right)\right)\tag{7}$$

完成特征融合后,通过上下文预测映射对主视图与上下文视图在重叠区域的像素位置进行对齐,将二者的预测结果进行空间匹配,基于不同视角在重叠区域的预测结果,得到与主视图像素位置一致的上下文预测结果$\boldsymbol{p}_j^{\mathrm{u_{wc}}}$.

接下来,空间感知门控利用上一阶段融合后的特征生成空间激活图:

$$\boldsymbol{s}_j^{\mathrm{w}}=\sigma\left(\mathrm{Conv}_{1\times1}\left(\mathrm{ReLU}\left(\mathrm{BN}\left(\mathrm{DWConv}(\boldsymbol{e}_j^{\mathrm{u_l}})\right)\right)\right)\right)\tag{8}$$

$\boldsymbol{s}_j^{\mathrm{w}}$对应于$\boldsymbol{p}_j^{\mathrm{u_w}}$的空间激活图,使用深度可分离卷积DWConv和$1\times1$卷积$\mathrm{Conv}_{1\times1}$生成. 该图使用sigmoid函数$\sigma$作为门控函数,将其映射到范围(0,1). 对于$\boldsymbol{p}_j^{\mathrm{u_{wc}}}$的空间激活图,则为$\boldsymbol{s}_j^{\mathrm{c}}=1-\boldsymbol{s}_j^{\mathrm{w}}$. 以空间激活图为引导,将弱增强图像的概率图$\boldsymbol{p}_j^{\mathrm{u_w}}$与上下文图像的概率图$\boldsymbol{p}_j^{\mathrm{u_{wc}}}$进行加权融合,可以获得融合的预测概率图$\boldsymbol{p}_j^{\mathrm{u_{SIF}}}$:

$$\boldsymbol{p}_j^{\mathrm{u_{SIF}}}=\boldsymbol{s}_j^{\mathrm{w}}\odot\boldsymbol{p}_j^{\mathrm{u_w}}+\boldsymbol{s}_j^{\mathrm{c}}\odot\boldsymbol{p}_j^{\mathrm{u_{wc}}}\tag{9}$$

$$\hat{\boldsymbol{y}}_j^{\mathrm{SIF}}=\mathrm{argmax}(\boldsymbol{p}_j^{\mathrm{u_{SIF}}})\tag{10}$$

其中,$\boldsymbol{p}_j^{\mathrm{u_w}}\in\mathbb{R}^{H\times W\times K}$是主视图的预测概率;$\odot$表示按位置相乘;最后,融合后的伪标签$\hat{\boldsymbol{y}}_j^{\mathrm{SIF}}$用作指导模型在未标记数据上学习的监督信号. 值得注意的是,在有监督训练中同样使用空间感知交互融合模块,并通过真实标签监督其参数更新,计算过程与上述步骤一致.

3.4　多重跨视图上下文一致性约束

为了进一步挖掘遥感图像中不同区域间的上下文信息,本文引入了多重跨视图上下文一致性约束,旨在确保训练过程中对具有重叠区域的不同视图实现语义输出的一致性. 具体而言,该约束要求在每次训练迭代中,上下文图像的预测概率图应在重叠区域内与主视图生成的伪标签保持一致. 在实现过程中,首先通过重叠采样策略获取上下文图像的重叠区域掩码,并据此从预测概率图中提取重叠区域. 同样地,通过主视图的重叠掩码从伪标签中去除非重叠部分,仅保留与上下文图像对应的区域. 这样,便完成了上下文预测与主视图伪标签在重叠区域的空间对齐. 最后,计算重叠区域内预测概率图与伪标签之间的交叉熵损失,以此约束模型在不同视图下对相同空间区域的预测保持一致性:

$$\boldsymbol{p}_{jk}^{\mathrm{u_{wc}}}=\mathrm{Mask}(\boldsymbol{p}_j^{\mathrm{u_{wc}}},\mathbf{Overlap}_{jk}^{\mathrm{context}})\tag{11}$$

$$\hat{\boldsymbol{y}}_{jk}^{\mathrm{SIF}}=\mathrm{Mask}(\hat{\boldsymbol{y}}_j^{\mathrm{SIF}},\mathbf{Overlap}_{jk})\tag{12}$$

$$L_{\mathrm{U}}^{\mathrm{c_k}}=\frac{1}{B_{\mathrm{u}}}\mathbb{I}\left(\max(\boldsymbol{p}_{jk}^{\mathrm{u_{sc}}})\geqslant\tau\right)L_{\mathrm{ce}}(\hat{\boldsymbol{y}}_{jk}^{\mathrm{SIF}},\boldsymbol{p}_{jk}^{\mathrm{u_{sc}}}) \tag{13}$$

其中，Mask 表示提取重叠区域；$\hat{\boldsymbol{y}}_{jk}^{\mathrm{SIF}}$ 表示通过空间感知交互融合模块生成的伪标签；$\boldsymbol{p}_{jk}^{\mathrm{u_{sc}}}$ 表示通过掩码提取的第 k 个强增强上下文视图 $\boldsymbol{x}_{jk}^{\mathrm{u_{sc}}}$ 的预测概率图；$\hat{\boldsymbol{y}}_{jk}^{\mathrm{SIF}}$ 表示通过掩码提取 $\boldsymbol{p}_{jk}^{\mathrm{u_{sc}}}$ 对应的伪标签，因此，上下文一致性损失可表示为

$$L_{\mathrm{U}}^{\mathrm{context}}=\sum_{k=1}^{N}L_{\mathrm{U}}^{\mathrm{c_k}} \tag{14}$$

对于标注数据，训练过程中的监督损失除了 L_{s}，还包含融合后主视图的交叉熵损失 $L_{\mathrm{S}}^{\mathrm{fusion}}$，以及上下文视图的交叉熵损失 $L_{\mathrm{S}}^{\mathrm{C}}$：

$$L_{\mathrm{S}}^{\mathrm{fusion}}=\frac{1}{B_{\mathrm{l}}}\sum_{i=1}^{B_{\mathrm{l}}}L_{\mathrm{ce}}(\boldsymbol{y}_i^{\mathrm{l}},\boldsymbol{p}_i^{\mathrm{l_{SIF}}}) \tag{15}$$

$$L_{\mathrm{S}}^{\mathrm{C}}=\frac{1}{B_{\mathrm{l}}}\sum_{i=1}^{B_{\mathrm{l}}}\sum_{m=1}^{N}L_{\mathrm{ce}}(\boldsymbol{y}_{im}^{\mathrm{l_c}},\boldsymbol{p}_{im}^{\mathrm{l_c}}) \tag{16}$$

其中，$\boldsymbol{p}_i^{\mathrm{l_{SIF}}}$ 表示经过空间感知交互融合模块后的主视图预测概率图；$\boldsymbol{y}_{im}^{\mathrm{l_c}}$ 表示第 i 次采样的第 m 个上下文视图的标签，表示对应视图的模型预测概率图，因此，总的目标函数可表示为

$$L_{\mathrm{total}}=L_{\mathrm{S}}^{\mathrm{fusion}}+L_{\mathrm{S}}+L_{\mathrm{S}}^{\mathrm{C}}+\lambda_{\mathrm{con}}L_{\mathrm{U}}^{\mathrm{context}}+\lambda_{\mathrm{us}}L_{\mathrm{u_s}} \tag{17}$$

其中，λ_{con} 和 λ_{us} 分别为 $L_{\mathrm{U}}^{\mathrm{context}}$ 和 $L_{\mathrm{u_s}}$ 的权重.

4 实验结果与分析

4.1 数据集

本文选用遥感领域广泛应用的两个标准数据集，由国际摄影测量与遥感学会工作组 WG II/4 提供的 Vaihingen 数据集和 Potsdam 数据集[29]，作为实验评估的基础，确保实验结果的权威性和对比的可参考性. Vaihingen 数据集共包含 33 张遥感图像，图像尺寸不一，平均大小约为 2 494 × 2 064 像素，空间分辨率为 9 cm. 该数据集覆盖典型的城市环境，划分为 5 个具有代表性的地物类别：不透水地面、建筑物、低矮植被、树木和汽车，能够较为全面地反映复杂的地表覆盖特征. Potsdam 数据集包含 38 张遥感图像，每张图像尺寸为 6 000 × 6 000 像素，空间分辨率为 5 cm，具有更高的精度和细节表达能力. 该数据集定义的地物类别与 Vaihingen 数据集保持一致，包括不透水地面、建筑物、低矮植被、树木和汽车. 为了确保所提出方法在半监督学习框架下的实验设计具有科学性和严谨性，并实现与现有方法的公平对比，本文分别在 Vaihingen 和 Potsdam 数据集的训练集中选取 1、2、4 张与 1、3、6 张图像作为有标签样本，构建标注数据集，其余图像则作为无标签样本用于模拟训练过程中的无标注样本. 测试阶段采用官方提供的测试集进行评估. 这种数据划分方式可以有效评估模型在标注样本极为有限的现实场景下的性能表现，体现所提方法在标注不足条件下的适应性与有效性.

4.2 实验设置

本文采用以 Swin-Transformer-Tiny（Swin-T）为主干网络的 UperNet 语义分割模型. 数据增强方式如下：弱数据增强包括图像在 0.5~2.0 之间的随机缩放、随机裁剪，以及以 0.5 的概率进行的水平翻转；强数据增强包括：随机亮度增强（0.5~1.5 倍）、随机灰度化、随机饱和度调整（0.5~0.25）、高斯模糊以及 CutMix 等策略. 所有实验均基于 PyTorch 框架，在两块 NVIDIA RTX 4090 GPU 上进行，输入图像分辨率为 512 × 512. 训练过程中使用 AdamW 优化器进行训练，其中权重衰减系数设为 0.01，基础学习率为 0.000 1，使用单周期学习率调度（One Cycle Learning Rate，OneCycleLR）策略，动态调整学习率以加速模型收敛. 模型在 Vaihingen 和 Potsdam 两个遥感图像语义分割标准数据集上均训练了 75 个训练轮次，批次大小设置为 12，λ_{con} 和 λ_{us} 均设置为 0.5. 所有实验均采用交并比（Intersection over Union，IoU）、平均交并比（mean Intersection over Union，mIoU）、平均 F1 值（mean F1 score，mF1）、总体精度（Overall Accuracy，OA）以及 Kappa 系数作为评价指标.

4.3 与现有方法对比

在本节中，本文在 Vaihingen 和 Potsdam 两个标准遥感图像语义分割数据集上，基于不同的标注划分比例，系统地对比评估了所提出方法与当前 8 种主流的先进半监督语义分割算法的性能. 所选对比方法包括：FixMatch[13]、RanPaste[18]、UniMatch[14]、CorrMatch[15]、SegMind[22]、DWL[12]、CAC[26] 和 CCVC[23]，其中 FixMatch 被用作本研究的基准方法.

（1）表 1 展示了在 Vaihingen 数据集上，当仅使用 1、2 和 4 张图像作为标注数据时，不同方法的语义分割性能. 可以观察到，所有采用半监督策略的方法在不同标注比例下均显著优于 SupOnly（Supervised-Only，仅监督训练），证明引入未标注数据对模型性能具有积极促进作用. 当标注图像数量为 1 张时，本文方法达到 75.27% 的 mIoU、85.62% 的 mF1、86.25% 的 OA 和 81.84% 的 Kappa，较 SupOnly 在各项指标上均有显著提升，分别提高了 6.84、4.61、4.49 和 5.91 个百分点. 相较于基准方法 FixMatch，本文方法在 mIoU、mF1、OA 和 Kappa 上分别领先 5.19、3.76、3.72 和 4.93 个百分点，在所有对比方法中表现最优. 尤其在不透水地面、建筑、树木和车辆等类别上均获得领先表现，体现了其在极少标注条件下对上下文信息建模的优势. 在使用 2 张标注图像的情况下，本文方法依旧保持领先地位，取得 76.18% 的

表1　在Vaihingen数据集上与6种先进的方法的比较　　单位:%

方法	不透水地面	建筑	低矮植被	树木	车辆	mIoU	mF1	OA	Kappa
1张标注图像									
SupOnly	73.98	77.94	55.51	68.51	66.24	68.43	81.01	81.76	75.93
FixMatch[13]	77.94	84.49	51.34	64.88	71.75	70.08	81.86	82.53	76.91
RanPaste[18]	79.41	86.93	**63.58**	72.15	66.55	73.72	84.38	85.22	80.78
UniMatch[14]	74.36	82.52	55.92	70.01	59.32	68.43	80.85	83.00	77.47
CorrMatch[15]	74.84	81.06	49.36	66.39	66.85	67.70	80.23	81.29	79.17
SegMind[22]	80.18	87.16	55.31	67.53	72.55	72.55	83.63	84.47	79.46
DWL[12]	79.24	86.86	60.33	71.44	72.60	74.09	84.58	85.57	81.08
CAC[26]	78.52	82.91	58.15	70.88	67.59	71.61	83.10	84.25	79.12
CCVC[23]	79.35	83.48	59.30	71.50	68.37	72.40	83.70	84.73	79.83
本文方法	**80.20**	**87.56**	61.98	**72.27**	**74.35**	**75.27**	**85.62**	**86.25**	**81.84**
2张标注图像									
SupOnly	76.59	82.15	59.51	71.03	70.19	71.89	83.42	84.05	78.97
FixMatch[13]	79.75	87.91	62.34	72.86	74.71	75.51	85.79	86.37	81.99
RanPaste[18]	79.45	87.57	60.51	70.94	72.12	74.12	84.83	85.62	81.00
UniMatch[14]	79.41	87.79	**63.53**	**74.93**	73.25	75.78	85.99	86.43	82.08
CorrMatch[15]	78.56	86.97	60.09	71.91	73.15	74.14	84.85	85.45	80.78
SegMind[22]	79.58	87.18	58.99	71.13	73.69	74.11	84.79	85.54	80.90
DWL[12]	78.35	85.42	62.22	72.92	75.09	74.80	85.36	85.66	81.08
CAC[26]	78.95	85.13	58.71	70.24	72.88	73.18	84.07	84.55	79.69
CCVC[23]	79.41	85.60	59.03	70.91	73.38	73.67	84.53	85.00	80.21
本文方法	**80.12**	**88.59**	62.39	73.53	**76.27**	**76.18**	**86.21**	**86.72**	**82.45**
4张标注图像									
SupOnly	77.82	83.78	64.83	75.19	70.3	74.38	85.15	86.07	81.60
FixMatch[13]	79.07	86.90	65.18	75.54	73.33	76.00	86.18	86.92	82.42
RanPaste[18]	79.14	86.96	65.61	**76.12**	72.17	76.00	86.18	87.05	82.59
UniMatch[14]	79.69	**88.42**	65.06	75.28	75.54	76.80	86.67	87.19	83.07
CorrMatch[15]	78.64	86.95	63.49	75.14	73.56	75.56	85.86	86.40	82.01
SegMind[22]	79.60	87.52	65.05	75.34	74.23	76.35	86.39	87.09	82.97
DWL[12]	79.07	86.62	65.08	75.87	74.66	76.26	86.35	86.95	82.66
CAC[26]	78.01	84.55	64.10	75.25	72.50	74.88	85.30	86.25	81.75
CCVC[23]	78.86	86.98	64.35	75.16	73.03	75.68	85.91	86.88	82.06
本文方法	**79.82**	87.67	**65.79**	75.66	**75.66**	**76.92**	**86.77**	**87.26**	**83.19**

注:SupOnly表示只使用标注数据进行训练,加粗数据表示最优结果.

mIoU、86.21% 的 mF1、86.72% 的 OA 和 82.45% 的 Kappa,全面超过 FixMatch(mIoU:75.51%,mF1:85.79%,OA:86.37%,Kappa:81.99%)及其他对比方法如 UniMatch(mIoU:75.78%,mF1:85.99%,OA:86.73%,Kappa:82.48%)和 DWL(mIoU:74.80%,mF1:85.36%,OA:85.66%,Kappa:81.08%). 当标注图像增加到4张时,本文方法达到76.92%的mIoU、86.77%的mF1、87.26%的OA和83.19%的Kappa,继续领先于所有对比方法,如 FixMatch(mIoU:76.00%,mF1:86.18%,OA:86.92%,Kappa:82.42%)和 SegMind(mIoU:76.35%,mF1:86.39%,OA:87.09%,Kappa:82.97%). 验证了该方法具有良好的可扩展性和稳定性,在数据量持续增长的情况下,依然能保持强劲的性能.

(2)表2展示了在Potsdam数据集上,使用1、3和6张标注图像时,各方法的语义分割性能情况. 整体来看,本文方法在所有标注划分下均取得了最优的性能表现. 仅使用1张标注图像时,本文方法就达到了79.92%的mIoU、88.49%的mF1、88.07%的OA和84.42%的Kappa,相比SupOnly基线四项指标分别显著提升12.73、8.03、8.38和11.11个百分点,也明显优于DWL、CAC、CCVC等对比方法,显示出在极低标注资源下仍能保持良好的分割性能与稳定性. 随着标注数量

增至3张和6张，本文方法继续全面领先，在mIoU、mF1、OA和Kappa上一致超过包括SegMind、UniMatch在内的主流方法，进一步验证了本文所提出方法适应不同标注规模的优势与可靠性.

表2 在Potsdam数据集上与6种先进的方法的比较 单位：%

方法	不透水地面	建筑	低矮植被	树木	车辆	mIoU	mF1	OA	Kappa
1张标注图像									
SupOnly	67.05	70.51	59.45	61.25	77.67	67.19	80.46	79.69	73.31
FixMatch[13]	67.71	74.79	65.80	63.01	80.67	70.40	82.60	81.81	76.43
RanPaste[18]	77.24	85.27	66.51	64.96	91.68	77.13	86.62	85.92	81.67
UniMatch[14]	62.84	57.44	62.25	61.47	89.17	66.63	79.36	77.31	70.23
CorrMatch[15]	62.36	65.28	58.41	59.63	80.71	65.28	79.09	77.07	70.16
SegMind[22]	79.46	87.38	68.35	63.98	81.21	76.08	86.14	85.61	81.46
DWL[12]	74.35	77.89	66.80	**68.12**	90.72	75.58	85.65	84.34	79.72
CAC[26]	75.34	83.12	68.91	65.45	90.55	76.67	86.23	85.78	81.52
CCVC[23]	80.83	87.56	67.83	65.58	91.02	78.56	87.35	86.97	83.39
本文方法	**82.16**	**88.68**	**70.01**	66.78	**91.95**	**79.92**	**88.49**	**88.07**	**84.42**
3张标注图像									
SupOnly	78.34	82.91	70.34	71.83	90.52	78.79	87.47	87.06	83.53
FixMatch[13]	83.87	90.47	73.93	73.27	93.53	83.01	90.62	90.14	87.14
RanPaste[18]	83.38	89.71	74.02	73.74	93.17	82.80	90.31	89.73	86.69
UniMatch[14]	84.17	91.31	73.53	72.77	**94.30**	83.22	90.71	89.98	86.71
CorrMatch[15]	83.96	90.64	72.93	72.75	94.09	82.87	90.33	89.83	86.79
SegMind[22]	**85.30**	**92.07**	74.87	73.27	91.37	83.38	90.75	90.24	87.12
DWL[12]	83.71	89.22	73.49	73.38	93.71	82.70	90.27	89.71	86.63
CAC[26]	84.55	91.05	73.15	73.61	93.92	83.26	90.68	90.21	87.05
CCVC[23]	83.82	90.93	73.37	74.39	93.85	83.27	90.65	90.10	87.11
本文方法	84.82	91.48	**74.94**	**74.50**	94.12	**83.97**	**91.07**	**90.63**	**87.78**
6张标注图像									
SupOnly	82.74	88.77	72.87	72.34	92.21	81.79	89.87	89.37	86.13
FixMatch[13]	84.86	91.55	74.37	73.43	93.12	83.47	90.65	90.28	87.38
RanPaste[18]	84.20	91.07	74.09	72.97	92.61	82.98	90.43	90.06	87.07
UniMatch[14]	84.26	91.97	73.42	72.10	93.14	82.98	90.35	90.02	87.01
CorrMatch[15]	84.83	92.75	73.70	72.79	93.50	83.51	90.69	90.33	87.48
SegMind[22]	85.58	92.73	75.66	74.64	92.46	84.21	91.03	90.86	88.02
DWL[12]	84.69	91.03	74.02	73.92	93.73	83.47	90.70	90.31	87.31
CAC[26]	84.91	92.15	74.12	72.58	93.43	83.44	90.72	90.38	87.41
CCVC[23]	85.04	92.66	74.48	74.64	93.62	84.09	90.97	90.74	87.93
本文方法	**86.16**	**93.26**	**75.97**	**74.87**	**93.94**	**84.84**	**91.59**	**91.30**	**88.64**

注：SupOnly表示只使用标注数据进行训练，加粗数据表示最优结果.

综上所述，本文方法在3种不同标注划分设置下均取得最优结果，展示出其在低标注环境中对上下文信息建模的强大能力与泛化优势. 无论在极低还是相对较高标注资源条件下，本文方法都实现了对现有方法的超越.

(3)可视化结果对比. 本文在只有1张标注图像的情况下，对Vaihingen数据集和Potsdam数据集上不同方法的分割结果进行了可视化，如图3所示. 可以观察到，所提出的方法获得了更精确的分割结果. 这进一步验证了通过跨视图上下文增强和一致性学习，本文方法在遥感图像上展现了更佳的性能.

(4)各方法的训练时间和计算复杂度对比. 如表3所示，由于引入了上下文视图的交互和一致性约束，所提出的方法在计算复杂度上有所提升，但整体资源消耗仍在可控范围内. 具体而言，本文方法的训练时间为每轮1.50 min，接近DWL的1.42 min，远低于SegMind的6.58 min；同时，所需模型参数量为37.57 M，低于RanPaste、CorrMatch、CCVC与SegMind的参数量. 更为重要

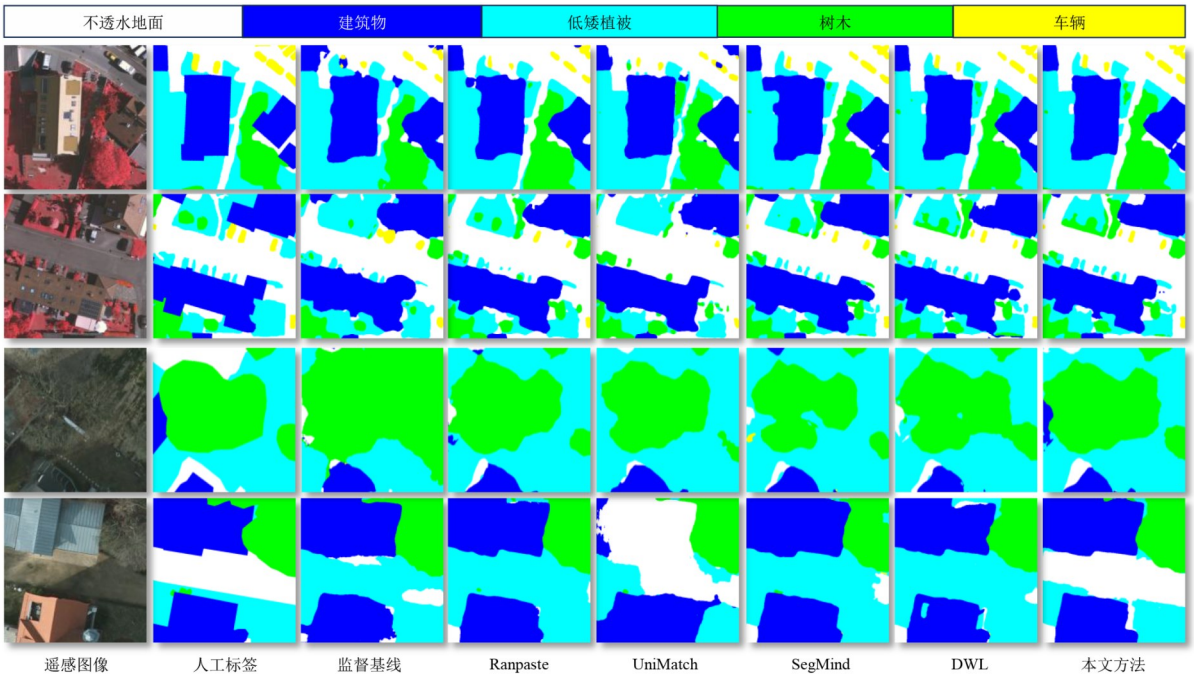


图3　各方法在Vaihingen数据集(第1行至第2行)和Potsdam数据集(第3行至第4行)上分割结果的对比

表3　在Potsdam数据集各方法的训练时间、计算复杂度和参数量对比

方法	每轮训练时间/min	GFLOPs	参数量/M	mIoU/%	mF1/%	OA/%	Kappa/%
FixMatch[13]	1.36	233.71	36.71	70.40	82.60	81.81	76.43
RanPaste[18]	1.53	311.61	73.42	77.13	86.62	85.92	81.67
UniMatch[14]	2.02	545.32	36.71	66.63	79.36	77.31	70.23
CorrMatch[15]	2.12	498.02	38.50	65.28	79.09	77.07	70.16
SegMind[22]	6.58	311.62	76.62	76.08	86.14	85.61	81.46
DWL[12]	1.42	233.71	36.71	75.58	85.65	84.34	79.72
CAC[26]	1.52	277.90	36.71	76.67	86.23	85.78	81.52
CCVC[23]	4.10	355.81	73.42	78.56	87.35	87.87	83.39
本文方法	1.50	488.26	37.57	79.92	88.49	88.07	84.42

注:GFLOPs表示单次前向传播所需的浮点运算量.

的是,本文方法在仅使用1张标注图像时,在Potsdam测试集上取得79.92%的mIoU、88.49%的mF1、88.07%的OA和84.42%的Kappa,整体表现均优于RanPaste(mIoU:77.13%,mF1:86.62%,OA:85.92%,Kappa:81.67%)和DWL(mIoU:75.58%,mF1:85.65%,OA:84.34%,Kappa:79.72%)等方法.综上所述,本文方法在保持训练效率和轻微增加模型参数量的同时,显著提高了分割精度,展示了在兼顾性能与效率方面的显著优势与实用价值.

4.4　消融实验

(1)不同组件的消融实验分析.为深入评估所提出方法中各关键模块对模型性能的贡献,本文在Potsdam数据集上进行了系统的消融实验.实验以FixMatch为半监督学习基线,逐步引入跨视图上下文一致性CVCC和空间感知交互融合SIF两个核心组件.具体实验结果如表4所示.监督基线模型在1、3和6张标注图像下的mIoU分别为67.19%、78.79%和81.79%,mF1为80.46%、87.47%和89.87%,OA为79.69%、87.06%和89.37%,Kappa系数分别为73.31%、83.53%和86.13%,作为对比的基础性能.引入半监督基线FixMatch后,各项指标均有显著提升,mIoU分别达到70.40%、83.01%和83.47%,mF1为82.60%、90.62%和90.65%,OA为81.81%、90.14%和90.28%,Kappa为76.43%、87.14%和87.38%,验证了半监督框架在低标注数据情境下的有效性.

进一步引入CVCC约束机制后,模型性能获得全面提升.在1张标注图像下,mIoU、mF1、OA及Kappa分别达到78.33%、87.23%、86.73%和83.03%,相比半监督基线分别显著提升了7.93、4.63、4.92和6.60个百分点;在3张和6张标注图像下,mIoU也分别增至83.65%和

83.85%，mF1、OA与Kappa指标也一致提高．这一结果表明，CVCC有效强化了模型对多视图之间语义一致性的建模能力，帮助其在具有不同上下文信息的同一区域仍能保持相同的预测．

表4 在Potsdam数据集对本文方法各组成部分的消融实验 单位：%

监督基线	半监督基线	CVCC	SIF	1张标注图像				3张标注图像				6张标注图像			
				mIoU	mF1	OA	Kappa	mIoU	mF1	OA	Kappa	mIoU	mF1	OA	Kappa
✓	—	—	—	67.19	80.46	79.69	73.31	78.79	87.47	87.06	83.53	81.79	89.87	89.37	86.13
✓	✓	—	—	70.40	82.60	81.81	76.43	83.01	90.62	90.14	87.14	83.47	90.65	90.28	87.38
✓	✓	✓	—	78.33	87.23	86.73	83.03	83.65	90.75	90.35	87.45	83.85	90.91	90.56	87.73
✓	✓	✓	✓	**79.92**	**88.49**	**88.07**	**84.42**	**83.97**	**91.07**	**90.63**	**87.78**	**84.84**	**91.59**	**91.30**	**88.64**

注：加粗数据表示最优结果．

最后，在CVCC基础上集成SIF模块，模型在所有标注设置下均达到最佳性能．1张标注图像时，mIoU、mF1、OA和Kappa进一步提升至79.92%、88.49%、88.07%和84.42%；3张和6张标注图像下，mIoU分别达到83.97%和84.84%，mF1超过91%，OA和Kappa也呈持续增长趋势．这说明SIF能够通过学习不同上下文伪标签之间的空间依赖关系，自适应调整伪标签融合权重，从而进一步提升伪标签的准确性与鲁棒性．

综上，CVCC和SIF两个模块在提升模型性能方面均发挥了关键作用，尤其是在标注样本极少的情况下，其对模型泛化能力的增强效果更加显著，充分证明了所提方法的设计合理性与有效性．

(2)上下文视图数量N的消融实验．为进一步探究上下文视图数量对模型性能的影响，本文在Potsdam数据集上进行了相关的消融实验．实验设置在使用1张、3张和6张标注图像的3种监督强度下，比较了不同上下文视图数量$N\in\{0,1,2,3\}$对分割精度和显存占用的影响，结果如表5所示．随着上下文视图数量N的增加，除mIoU之外，mF1、OA和Kappa均表现出同步提升的趋势．当$N=0$时，模型仅为半监督基线，各项指标处于最低水平(例如1张标注图像下mIoU：70.40%，mF1：82.60%、OA：81.81%，Kappa：76.43)．引入1个上下文视图后，mIoU显著提升至78.01%，mF1、OA和Kappa分别提升至86.93%、86.45%和83.06%，说明上下文一致性机制不仅提升了类别区分能力，也增强了整体分类正确率与一致性．进一步增加视图数量至$N=2$和$N=3$时，mIoU、mF1、OA和Kappa均稳步提高，在$N=3$时达到最佳(1张标注图像下mIoU=78.57%，mF1 = 87.45%、OA=86.95%、Kappa= 83.25)；3张标注图像和6张标注图像的条件下各项指标进一步提升．这表明，上下文视图数量的增加能够有效提升模型在不同标注划分下的整体性能，尤其是Kappa指标的持续提升，验证了模型预测结果与真实标注在一致性上的增强．然而，随着N增大，显存占用也由5 020 MB增加至8 400 MB，提升约为原来的1.7倍．综合性能与效率，$N=2$在保证较优分割精度的同时，显存占用(约6 810 MB)在可接受范围内，是最佳折中方案，确保了方法在实际应用中的可行性与效率．

表5 在Potsdam数据集对上下文视图数量N的消融实验

N/个	1张标注图像/%				3张标注图像/%				6张标注图像/%				显存占用/MB
	mIoU	mF1	OA	Kappa	mIoU	mF1	OA	Kappa	mIoU	mF1	OA	Kappa	
0	70.40	82.60	81.81	76.43	83.01	90.62	90.14	87.14	83.47	90.65	90.28	87.38	5 020
1	78.01	86.93	86.45	83.06	83.17	90.42	89.87	86.80	83.52	90.68	90.33	86.73	6 078
2	78.33	87.23	86.73	83.03	83.65	90.75	90.35	87.45	83.85	90.91	90.56	87.73	6 810
3	**78.57**	**87.45**	**86.95**	**83.25**	**83.97**	**91.03**	**90.65**	**87.81**	**84.18**	**91.15**	**90.80**	**87.95**	8 400

注：加粗数据表示最优结果．

(3)重叠区域占比R的消融实验．为探究上下文视图与主视图之间的最小重叠区域占比R对模型性能的影响，本文在N=2的情况下，在Potsdam数据集上进行了系统的消融实验，结果如表6所示．实验在不同标注数量(1、3、6张)下，对比了最小重叠比例$R\in\{0,25\%,50\%,75\%,90\%\}$时模型的分割表现．$R=0$表示半监督基线FixMatch．随着引入上下文视图，模型性能显著提升．当$R=25\%$时，在1张标注图像条件下，mIoU、mF1、OA和Kappa分别达到77.95%、86.75%、86.25%和82.80%；当$R=50\%$时，这些指标进一步提升至78.33%、87.23%、86.73%和83.03%，说明适度的重叠能够增强不同视图间的信息交互，为SIF模块提供更丰富的上下文线索，从而全面提升伪标签质量并促进模型学习语义一致性．当重叠比例进一步增加至75%和90%时，多项指标的性能提升开始逐渐下降或趋于平缓．在1张标注图像条件下，$R=75\%$时，mIoU为78.32%，mF1为87.20%，OA为86.70%，Kappa为83.00%；而$R=90\%$时，各项指标分别为78.15%、87.00%、86.50%和82.90%，均

略低于R = 50%时的结果,在3张和6张标注图像条件下也呈现出类似规律. 原因在于,过高的重叠比例虽然带来了更多共享区域,但也引入了大量冗余信息,限制了上下文多样性,削弱了模型对上下文信息的学习.

表6　在Potsdam数据集对上下文视图与主视图重叠区域占比R的消融实验　　单位:%

R	1张标注图像				3张标注图像				6张标注图像			
	mIoU	mF1	OA	Kappa	mIoU	mF1	OA	Kappa	mIoU	mF1	OA	Kappa
0	70.40	82.60	81.81	76.43	83.01	90.62	90.14	87.14	83.47	90.65	90.28	87.38
25	77.95	86.75	86.25	82.80	83.16	90.58	89.87	86.78	83.57	90.72	90.37	87.40
50	**78.33**	**87.23**	**86.73**	**83.03**	**83.65**	**90.75**	**90.35**	**87.45**	83.85	90.91	**90.56**	87.73
75	78.32	87.20	86.70	83.00	83.52	90.68	90.32	87.31	**83.87**	**90.95**	90.50	**87.76**
90	78.15	87.00	86.50	82.90	83.18	90.39	89.92	86.95	83.75	90.85	90.50	87.58

注:加粗数据表示最优结果.

综上所述,重叠区域比例对模型性能有显著影响,既不能过小也不能过大. 综合对比之下,R = 50%是当前实验中表现最优的设置,能够在保证信息交互的同时避免冗余信息带来的负面效应,为上下文一致性建模提供了良好的平衡点.

(4)SIF模块的消融实验. 本文对所提出的SIF模块融合伪标签进行了详尽分析,包括基于SIF的上下文推理、SIF伪标签的分析. 首先我们进行了基于SIF的上下文推理的消融实验. 在推理阶段,我们采用滑动窗口对高分辨率遥感图像进行采样,每次采样获得一张主视图,以及来自左右相邻区域的两张上下文视图(若超出边界,则通过镜像填充),并结合SIF模块进行上下文推理. 表7中带有"*"标记的结果即表示结合SIF推理后的性能. 从表7可以看出,在不同标注图像数量下,引入SIF上下文推理后模型的各项性能指标均获得一致且稳定的提升. 在使用1张标注图像时,mIoU从79.92%提升至80.41%,mF1从88.49%提升至88.80%,OA从88.07%提升至88.39%,Kappa系数从84.42%提升至84.84%. 在3张和6张标注图像条件下也观察到类似趋势,mIoU分别提升0.74个百分点和0.68个百分点,mF1分别提升0.45个百分点和0.40个百分点,OA分别提升0.47个百分点和0.43个百分点,Kappa分别提升0.60个百分点和0.56个百分点. 这些结果表明,SIF模块能够有效整合多视图上下文信息,在推理阶段显著增强模型的分割性能. 进一步地,图4(a)展示了Potsdam数据集(1张标注图像)下,不同训练轮次中初始伪标签与SIF融合伪标签的平均mIoU(仅用真实标签评估,未参与训练). 结果显示,在训练的各个阶段,基于SIF融合的伪标签均显著优于初始伪标签,证明其能够在不同训练轮次中持续提供更高质量的监督信号. 本文在图4(b)选取了训练过程中1 000次迭代中每个批次生成的伪标签,并计算其mIoU,以观察伪标签精度的变化趋势. 左图展示了初始伪标签与SIF融合伪标签在每个批次中的精度曲线. 可以看出,在大多数迭代中,SIF融合伪标签(红色曲线)均优于初始伪标签(黄色曲线),表明基于上下文融合的策略能够显著提升伪标签的质量与可靠性. 图4(c)绘制了1 000次迭代中融合伪标签相对于初始伪标签精度提升的累计差值曲线,清晰地展现了伪标签质量随训练进展不断提升的趋势. 这一增长趋势说明,SIF融合机制在训练过程中能够稳定地优化伪标签质量,进而推动模型更有效地学习,提升整体性能与收敛效果. 同时在图5中,我们展示了由模型预测主视图得到的伪标签、上下文视图的伪标签,以及SIF生成的空间激活图. 其中,空间激活图中像素的亮度表示融合权重,亮度越高权重越大. 对比蓝色(Vaihingen)和红色框(Potsdam)区域可见,SIF能有效结合不同视图的上下文信息,为正确分割的像素分配更高权重,从而生成更精确、更鲁棒的伪标签.

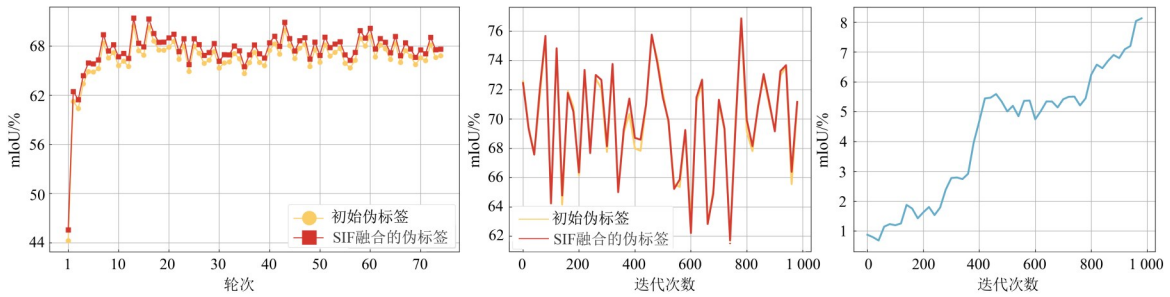
表7　Potsdam数据集上基于SIF模块的上下文推理消融实验

标注图像数量	方法	mIoU/%	mF1/%	OA/%	Kappa/%
1张标注图像	本文方法	79.92	88.49	88.07	84.42
	本文方法*	**80.41**	**88.80**	**88.39**	**84.84**
3张标注图像	本文方法	83.97	91.07	90.63	87.78
	本文方法*	**84.71**	**91.52**	**91.10**	**88.38**
6张标注图像	本文方法	84.84	91.59	91.30	88.64
	本文方法*	**85.52**	**91.99**	**91.73**	**89.20**

注:*表示结合上下文视图进行推理,加粗数据为相对较优结果.

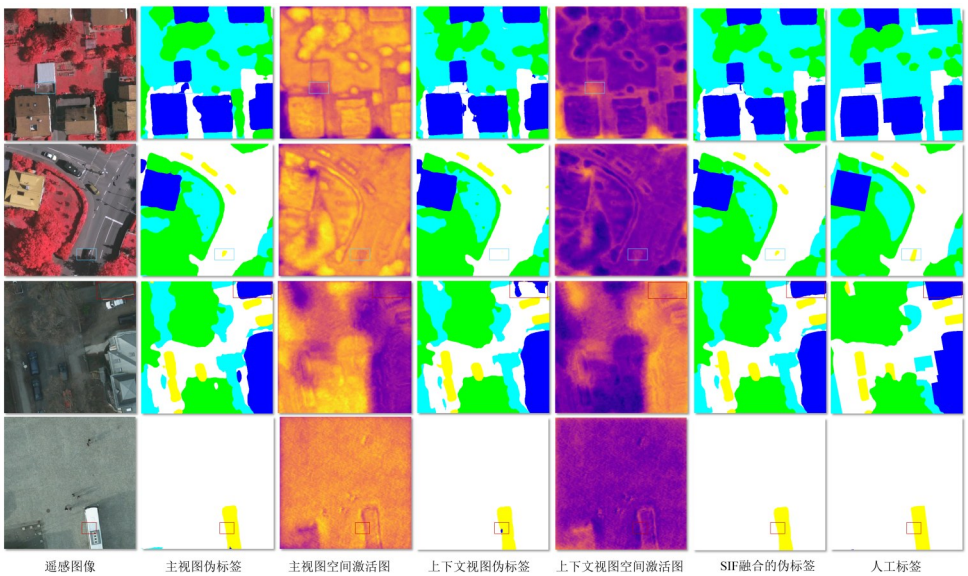
(5)输入分辨率消融实验. 本文探索了输入分辨率大小对所提出方法性能的影响,实验结果如表8所示. 可以看出,随着输入分辨率的提升,模型性能呈现逐步提升的趋势. 然而,即便输入分辨率为256 × 256,本文提出的方法依然取得了较为优异的性能;当提升至768 × 768时性能进一步提高,但训练时间也明显增加. 综合考虑性能与训练时间的平衡,我们最终选取512 × 512作为主视图及上下文视图的窗口大小.

(6)主视图采样数量消融实验. 与传统将整幅高分辨率遥感图像裁切为固定图块的训练策略不同,本文在每次迭代中需同时采样一张主视图与若干上下文视图,因此我们需要在所有无监督高分辨率图像中先随机选取一张,再从中采样出用于训练的主视图和上下



（a）每轮次的伪标签平均精度对比曲线 （b）每次迭代的伪标签精度对比曲线 （c）SIF融合伪标签的累计精度增长曲线

图4 SIF模块融合伪标签与初始伪标签的精度对比曲线



注：从上到下每两行分别对应Vaihingen和Potsdam数据集.

图5 空间交互感知模块(SIF)的分割结果与空间激活图可视化

文视图. 为评估每轮次主视图采样数量对性能与计算开销的影响，我们在Potsdam数据集上进行了消融实验. 实验结果如表9所示，采样数量对模型表现与训练时间均有显著影响：当采样数量极少（如100）时，样本量不足导致模型训练不充分，性能较低；将采样数量提升到500可带来稳步提升，但训练时间同步增长；增至1 000时，性能出现明显跃升，表明该样本量能显著改善模型性能；而当进一步提高到1 500或2 000时，性能增益趋于平缓，但每轮训练时间显著增加. 综合考虑精度与训练效率，我们将每轮主视图采样数量设置为1 000.

表8 在Potsdam数据集输入图像分辨率大小的消融实验

输入分辨率	每轮训练时间/min	mIoU/%	mF1/%	OA/%	Kappa/%
256×256	**0.85**	78.45	87.35	86.93	83.35
512×512	1.50	79.92	88.49	88.07	84.42
768×768	2.56	**80.13**	**88.61**	**88.34**	**84.72**

注：加粗数据表示最优结果.

表9 在Potsdam数据集主视图采样数量的消融实验

采样数量/轮	每轮训练时间/min	mIoU/%	mF1/%	OA/%	Kappa/%
100	**0.42**	78.35	87.21	86.75	83.21
500	0.72	78.97	87.52	87.12	83.63
1 000	1.50	79.92	88.49	88.07	84.42
1 500	2.20	79.64	88.10	87.70	84.05
2 000	2.38	**80.11**	**88.60**	**88.25**	**84.70**

注：加粗数据表示最优结果.

5 结论

本文发现在标注数据有限的场景下,当前高分辨率遥感图像半监督语义分割方法面临的主要挑战之一是难以有效建模上下文信息. 为此,提出了一种基于跨视图上下文感知的高分辨率遥感图像半监督语义分割方法,以提升模型在有限标注条件下对上下文信息的建模能力. 该方法引入空间感知交互融合模块,通过生成空间注意力图,自适应融合不同视图的预测结果,有效提升伪标签的质量,从而为训练过程提供更可靠的监督信号. 同时,设计了多重跨视图上下文一致性约束,利用多个重叠区域的一致性约束增强不同视图间的语义一致性,有助于模型在多视角下学习更广泛的上下文信息. 在两个遥感图像分割基准数据集Vaihingen和Potsdam上的大量实验证明,该方法在不同划分协议下均显著优于现有主流方法,在标注数据有限的情况下,能有效提升模型的分割精度和泛化能力,展现出明显的优势.

参考文献

[1] ZHANG L F, ZHANG L P. Artificial Intelligence for Remote Sensing Data Analysis: A review of challenges and opportunities[J]. IEEE Geoscience and Remote Sensing Magazine, 2022, 10(2): 270-294.

[2] 梁燕, 易春霞, 王光宇, 等. 基于多尺度语义编解码网络的遥感图像语义分割[J]. 电子学报, 2023, 51(11): 3199-3214. LIANG Y, YI C X, WANG G Y, et al. Semantic segmentation of remote sensing image based on multi-scale semantic encoder-decoder network[J]. Acta Electronica Sinica, 2023, 51(11): 3199-3214. (in Chinese)

[3] VAN ENGELEN J E, HOOS H H. A survey on semi-supervised learning[J]. Machine Learning, 2020, 109(2): 373-440.

[4] QIAO S Y, SHEN W, ZHANG Z S, et al. Deep Co-training for semi-supervised image recognition[M]//Computer Vision - ECCV 2018. Cham: Springer International Publishing, 2018: 142-159.

[5] WANG D, ZHANG X Q, FAN M Y, et al. Semi-supervised dictionary learning via structural sparse preserving[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2016, 30(1): 2137-2144.

[6] YANG L H, ZHUO W, QI L, et al. ST++: Make self-training-Work better for semi-supervised semantic segmentation[C]// 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 4258-4267.

[7] TARVAINEN A, VALPOLA H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results[C]//Advances in Neural Information Processing Systems. San Diego: NeurIPS, 2017: 4268-4277.

[8] ZHANG Y X, LI W, SUN W D, et al. Single-source domain expansion network for cross-scene hyperspectral image classification[J]. IEEE Transactions on Image Processing, 2023, 32: 1498-1512.

[9] ZHANG Y X, LI W, JIA W, et al. Cross-domain hyperspectral image classification based on bi-directional domain adaptation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2025, 35(12): 12038-12051.

[10] WANG J X, CHEN S B, DING C H Q, et al. Semi-supervised semantic segmentation of remote sensing images with iterative contrastive network[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 2504005.

[11] XU Y Z, YAN L L, JIANG J. EI-HCR: An efficient end-to-end hybrid consistency regularization algorithm for semisupervised remote sensing image segmentation[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 4405015.

[12] HUANG W, SHI Y L, XIONG Z T, et al. Decouple and weight semi-supervised semantic segmentation of remote sensing images[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2024, 212: 13-26.

[13] SOHN K, BERTHELOT D, CARLINI N, et al. FixMatch: Simplifying semi-supervised learning with consistency and confidence[C]//Advances in Neural Information Processing Systems 33. San Diego: NeurIPS, 2020: 596-608.

[14] YANG L H, QI L, FENG L T, et al. Revisiting weak-to-strong consistency in semi-supervised semantic segmentation[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 7236-7246.

[15] SUN B Y, YANG Y Q, ZHANG L, et al. CorrMatch: Label propagation via correlation matching for semi-supervised semantic segmentation[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 3097-3107.

[16] WANG H N, ZHANG Q X, LI Y, et al. AllSpark: Reborn labeled features from unlabeled in transformer for semi-supervised semantic segmentation[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 3627-3636.

[17] LV L, ZHANG L F. ScaleMatch: Multi-scale consistency enhancement for semi-supervised semantic segmentation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2025, 39(6): 5910-5918.

[18] WANG J X, CHEN S B, DING C H Q, et al. RanPaste: Paste consistency and pseudo label for semisupervised remote sensing image semantic segmentation[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 2002916.

[19] JIN J D, LU W X, YU H F, et al. Dynamic and adaptive self-training for semi-supervised remote sensing image semantic segmentation[J]. IEEE Transactions on Geosci-

ence and Remote Sensing, 2024, 62: 5639814.
[20] CAI M X, CHEN H, ZHANG T, et al. Consistency regularization based on masked image modeling for semisupervised remote sensing semantic segmentation[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2024, 17: 17442-17460.
[21] XIN Y, FAN Z D, QI X Y, et al. Confidence-weighted dual-teacher networks with biased contrastive learning for semi-supervised semantic segmentation in remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 5614416.
[22] LI Z H, CHEN H, WU J J, et al. SegMind: Semisupervised remote sensing image semantic segmentation with masked image modeling and contrastive learning method[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 4408917.
[23] WANG Z C, ZHAO Z, XING X X, et al. Conflict-based cross-view consistency for semi-supervised semantic segmentation[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 19585-19595.
[24] FU Y J, WANG M Y, VIVONE G, et al. An alternating guidance with cross-view teacher-student framework for remote sensing semi-supervised semantic segmentation[J]. IEEE Transactions on Geoscience and Remote Sensing, 2025, 63: 4402012.
[25] LIU R Z, LUO T Z, HUANG S G, et al. CrossMatch: Cross-view matching for semi-supervised remote sensing image segmentation[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 5650515.
[26] LAI X, TIAN Z T, JIANG L, et al. Semi-supervised semantic segmentation with directional context-aware consistency[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 1205-1214.
[27] DANG B, LI Y S, ZHANG Y J, et al. Progressive learning with cross-window consistency for semi-supervised semantic segmentation[J]. IEEE Transactions on Image Processing, 2024, 33: 5219-5231.
[28] 兰猛, 张乐飞, 杜博, 等. 基于时空层级查询的指代视频目标分割[J]. 中国科学: 信息科学, 2024, 54(3): 674-691. LAN M, ZHANG Z, DU B, et al. Spatio-temporal hierarchical query for referring video object segmentation[J]. Scientia Sinica (Informationis), 2024, 54(3): 674-691. (in Chinese)
[29] ROTTENSTEINER F, SOHN G, JUNG J, et al. The isprs benchmark on urban object classification and 3d building reconstruction[J]. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2012, I-3: 293-298.

作者简介

吕 亮 男,1994年7月出生于甘肃省庆阳市. 现为武汉大学计算机学院博士研究生. 主要研究方向为遥感图像语义分割、半监督学习.
E-mail: lianglyu@whu.edu.cn

兰 杰 男,2000年12月出生于湖北省武汉市. 现为武汉大学计算机硕士研究生. 主要研究方向为半监督学习.
E-mail: jie_lan@whu.edu.cn

兰 猛 男,1996年2月出生于湖北省黄梅县. 现为香港科技大学电子与计算机工程系博士后研究员. 主要研究方向为计算机视觉.
E-mail: eemenglan@ust.hk

卢宪凯 男,1990年4月出生于山东省济南市. 现为山东大学软件学院研究员、博士生导师. 主要研究方向为视频目标分割、遥感图像分析等.
E-mail: carrierlxk@gmail.com

张乐飞 男,1986年6月出生于湖北省武汉市. 现为武汉大学计算机学院教授、博士生导师. 主要研究方向为计算机视觉、遥感信息处理、多模态遥感基础模型等. 中国电子学会会员编号:E190182786M.
E-mail: zhanglefei@whu.edu.cn