

基于弱监督实例分割的遥感图像低成本细粒度解译方法

陈满, 黄勇杰, 徐磊, 潘志松

陆军工程大学 指挥控制工程学院, 南京 210007

摘要: 遥感图像的实例分割能够同时实现感兴趣物体的目标级定位与像素级分类, 是一项重要且极具挑战性的任务。当前大多遥感图像实例分割方法依赖精细的像素级注释, 其制作成本高昂。此外, 遥感图像混杂的前背景和复杂的目标轮廓同样增加分割的难度。为了迎接这些挑战, 本文构建出一套适用于遥感图像弱监督实例分割任务的先验信息驱动体系, 并提出一种基于多重先验驱动的遥感图像弱监督实例分割网络。具体地, 将遥感图像弱监督实例分割任务中的先验信息按照其来源分为任务先验和图像先验, 其中任务先验源自与实例分割紧密相关的边界框检测任务, 图像先验来源于对图像本身信息的归纳与挖掘。进一步地, 设计框—掩码投影一致性约束、像素区分难度表征函数和中心位置先验约束3个具体组件实例化任务先验信息, 驱动网络确定掩码的尺寸并使其充分关注图像中的重点像素和区域; 设计邻域视觉一致性约束和梯度一致性约束两个组件构建图像先验信息, 使网络高效地进行前背景区分且适应遥感图像中复杂的目标轮廓。在光学和SAR遥感图像数据集上的实验结果表明, 所提方法无需任何像素级注释即可实现52.5和54.1的AP值, 优于现有的弱监督分割方法, 且达到全监督Mask R-CNN的89.3%和84.3%。该方法能够为遥感图像的细粒度解译提供一种高性能且低成本的解决方案。

关键词: 遥感图像, 实例分割, 细粒度解译, 弱监督学习, 先验信息, 驱动体系, 目标轮廓, 注释成本

中图分类号: TP751/P2

引用格式: 陈满, 黄勇杰, 徐磊, 潘志松. 2025. 基于弱监督实例分割的遥感图像低成本细粒度解译方法. 遥感学报, 29(4): 1002-1016

Chen M, Huang Y J, Xu L and Pan Z S. 2025. Low-Cost Fine-Grained interpretation method of remote sensing images based on weakly supervised instance segmentation. National Remote Sensing Bulletin, 29(4): 1002-1016 [DOI:10.11834/jrs.20243407]

1 引言

地球观测技术的进步使得遥感图像的获取方式更加多元化, 且图像质量也得到了显著的提高。对遥感图像进行解译, 有助于相关部门高效地获取大量有价值的信息, 在污染治理 (Huang 等, 2022)、海洋监测 (吕军 等, 2022) 和灾害分析 (李梦倩 等, 2022) 等多项工作中具有重要的应用价值。根据遥感图像解译方法的细化程度, 可以将其分为图像级 (Miao 等, 2022)、目标级 (张涛 等, 2022; 汪西莉 等, 2023) 和像素级 (吴强强 等, 2022) 3种类型。实例分割作为一种细粒度的像素级解译方法, 能够同时定位遥感图像中感兴趣

目标的位置并进行像素级分类, 是一项重要且极具挑战性的任务 (Chen 等, 2022)。

近年来, 遥感图像的实例分割方法得到了一定的研究, 经历了基于自然光实例分割方法的简单推广, 结合遥感图像特点改进, 以及标签受限情况下的探索3个阶段。早期的遥感图像实例分割方法大多是对自然光实例分割方法的简单推广。其中, Waqas Zamir 等 (2019) 通过在Mask R-CNN (He 等, 2020) 上进行微调初步实现遥感图像的实例分割。HQ-ISNet (Su 等, 2020) 将分割性能更强的Cascade Mask R-CNN (Cai 和 Vasconcelos, 2018) 引入遥感图像分割领域, 且在原始结构的基础上引入包含多级特征图的高分辨率FPN (Feature

收稿日期: 2023-09-20; 预印本: 2024-01-22

基金项目: 国家自然科学基金(编号: 62076251)

第一作者简介: 陈满, 研究方向为遥感图像处理。E-mail: chenmanhjd@163.com

通信作者简介: 潘志松, 研究方向为模式识别与机器学习。E-mail: panzhisong@aeu.edu.cn

Pyramid Network) (Lin 等, 2017), 以提升其对光学和 SAR (Synthetic Aperture Radar) 遥感图像的适应能力。随着研究的不断深入, 越来越多方法开始关注遥感图像自身的特点, 并根据具体的特点改进网络结构, 以实现更好的实例分割性能。Zhao 等 (2021) 关注到遥感图像背景复杂的特点, 设计一种基于协同注意力的 SAR 遥感图像实例分割方法, 该方法吸收了全局注意力和语义注意力等多种注意力的优势, 能够准确地从 SAR 遥感图像中提取出舰船的实例掩码。SS-Net (Zhang 等, 2022) 聚焦于遥感图像自身的特性, 通过引入语义注意力机制增强网络对感兴趣目标的激活, 还设计一种尺度互补的掩码预测分支, 对遥感图像中的目标掩码能够实现精确预测。近期出现的遥感图像分割方法中, LFG-Net (Wei 等, 2022) 通过在特征提取和融合阶段充分关注分辨率较高的低层特征, 并将超分辨率去噪技术补充到网络结构中, 在多个 SAR 遥感图像实例分割数据集上取得了令人满意的分割性能。近期, 部分学者对标签受限情况下的分割方法进行了探索, 旨在降低遥感图像实例分割任务中高昂的注释成本。具体地, MGWI-Net (Chen 等, 2023) 结合人类观察目标时的认知习惯将遥感图像实例分割过程拆解为“由粗到细”的感知步骤, 并设计了相应的实例化组件对感兴趣目标的掩码进行感知, 实现了对弱监督分割的初步探索。本研究聚焦于弱监督范式的实例分割任务, 以降低模型对昂贵的像素级标签的依赖, 实现视觉处理效果和注释成本之间的平衡。

现有的弱监督实例分割方法可按照其标签注释的类型细分为图像级监督和框级监督两种。图像级监督方法旨在通过简单的图像级标签训练网络实现目标的实例分割 (Ahn 等, 2019, 2020)。尽管这种方法能够显著降低标签制作成本, 但图像级标签能够提供的监督信息过于薄弱, 难以适应复杂的遥感图像。相比之下, 框级监督方法通过监督信息相对丰富的边界框标签训练网络, 更容易实现标记成本与监督能力的平衡, 因此本研究重点关注此类方法。目前, 大部分框级监督方法如 BBAM (Lee 等, 2021) 和 DiscoBox (Lan 等, 2021) 通过辅助任务或无监督方法生成伪标签, 然后使用伪标签指导网络进行训练。这些方法的训练步骤较为繁琐, 且最终的分割效果与全监督

方法相比存在一定差距。最近提出的一些方法 BoxInst (Tian 等, 2021)、DBIN (Li 等, 2022) 和 MGWI-Net (Chen 等, 2023) 等通过挖掘实例分割任务中的先验信息对目标掩码进行提取, 但是这些方法往往未能对先验信息进行充分地利用, 因此仍然具有一定的改进空间。此外, 相比于精细的像素级标签, 框级标签在网络的训练过程中提供的监督信息是粗糙的, 会导致网络对目标的感知不够细致, 这无疑会给实例分割任务带来挑战 (Hsu 等, 2019)。因此, 如何在网络学习的过通过粗糙的边界框注释获取高质量的目标轮廓, 也是弱监督实例分割方法需要关注的重点问题 (Hao 等, 2021)。

遥感图像分割任务的关键就是区分出图像中的前背景, 其中前景指的是待分割的物体, 背景则是指待分割物体的周围环境等非感兴趣区域。不同于自然图像, 光学遥感图像的背景中包含了建筑、植被和水体等多种要素, 这些要素与感兴趣目标 (即前景) 混杂在一起, 给前背景的区分带来了难度 (吕少云 等, 2023)。SAR 遥感图像存在斑点噪声高、信噪比低的挑战, 且图像中的目标也容易受到海岸和噪声等不利因素的影响, 因此前背景区分的难度也明显高于自然图像 (李宗凌 等, 2021)。另一方面, 光学和 SAR 遥感图像中待分割物体的轮廓往往较为复杂, 这也会给掩码的提取带来困难 (Yang 等, 2021)。因此, 在挖掘和利用先验信息对网络进行驱动时, 考虑这些限制因素非常有必要。

为了解决遥感图像实例分割任务中面临的标记成本高、框级标签监督信息粗糙、前背景混杂和目标轮廓复杂等问题, 本文建立一套适用于弱监督实例分割任务的先验信息驱动体系, 并结合该体系提出一种基于多重先验驱动的弱监督实例分割网络 MPD-WSIS-Net (Multi-Prior Driven Weakly Supervised Instance Segmentation Network)。首先, 对遥感图像弱监督实例分割任务中的先验信息进行归纳, 并将其按照来源划分为任务先验和图像先验两部分。然后, 设计的框—掩码投影一致性约束、像素区分难度表征函数和中心位置先验约束 3 个具体组件实例化任务先验信息, 驱动网络初步确定掩码的尺寸并使其充分关注图像中的重点像素和区域。接着, 设计邻域视觉一致性约束和

梯度一致性约束两个组件构建图像先验信息，驱动网络高效地区分前背景并适应遥感图像中复杂的目标边界。最后，通过整合上述5个组件建立完整的先验信息驱动体系，并将该体系融入到具体的网络结构中，得到MPD-WSIS-Net，实现遥感图像的弱监督实例分割任务。

2 研究方法

本节首先对遥感图像弱监督实例分割任务中先验信息的来源和总体架构进行详细描述，然后分别介绍任务先验和图像先验的具体实现方式，最后对先验信息驱动体系和MPD-WSIS-Net的网络结构进行说明。

2.1 先验信息的来源和总体架构

现有的大部分框级监督方法侧重于通过辅助任务或无监督方法如GrabCut (Rother等, 2004) 和MCG (Pont-Tuset等, 2017) 等生成伪标签，然后使用伪标签指导网络进行训练。这种方式涉及复杂的训练步骤，难以进行端到端训练，且最终的分割效果与全监督方法相比仍有一定的差距。近期的一些方法通过挖掘先验信息指导网络实现弱监督实例分割，但是对先验信息的利用程度不足，难以应用于遥感图像中。因此，本研究建立了如图1所示的先验信息驱动体系，对弱监督实例分割任务中的先验信息进行更深层次的挖掘与利用，在有限的监督条件下实现更好的分割效果。

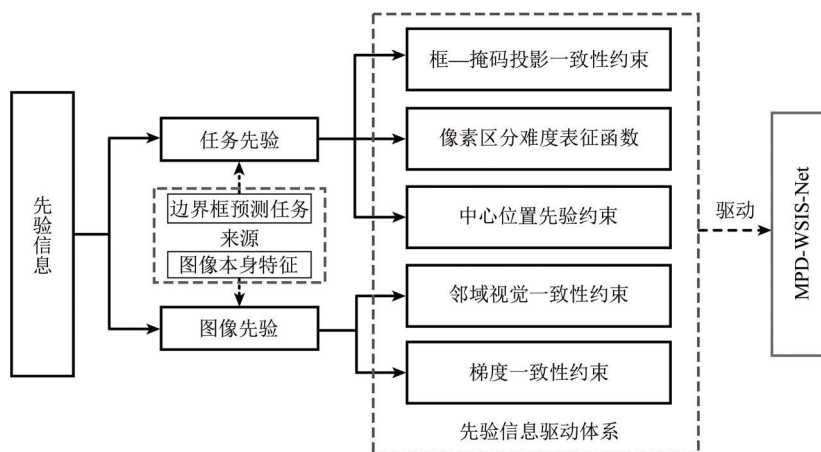


图1 先验信息的来源和总体架构

Fig. 1 Sources and overall architecture of the priori information

本文通过对弱监督实例分割任务中的先验信息进行细致地观察，发现其先验知识主要源于两个角度：任务角度和图像角度。任务角度的先验信息源自与实例分割密切相关的边界框检测任务。在实例分割中，目标的掩码预测任务和边界框预测任务是紧密关联的，因此可从边界框预测任务中获取一定的先验信息，以促进无像素级注释条件下的掩码预测。具体地，论文通过精心设计的框-掩码投影一致性约束、像素区分难度表征函数和中心位置先验约束3个组件对边界框预测中的先验信息进行实例化。其中，框-掩码投影一致性约束能够使预测掩码的最小边界矩形与相应的边界框保持一致，进而初步确定掩码的尺寸信息；像素区分难度表征函数充分考虑了遥感图像中前背景区分难度大的问题，能够使网络在训练过程中对不确定程度较高的像素点进行更多的关注；

中心位置先验约束能够促进模型对高置信度前景区域进行合理的预测，进一步改善分割性能。

图像角度的先验信息源于对图像本身信息的归纳或挖掘，论文重点关注图像中邻域像素点之间的关系和目标的梯度信息。邻域像素点关系方面，本研究构建了邻域视觉一致性约束，该约束倾向于将图像中邻近且具有较高视觉相似度的像素点分配到同一类别中，并能够与像素区分难度表征函数进行结合，共同实现高效的前背景的区分。梯度信息方面，构建了梯度一致性约束，这项约束能够使预测掩码的边界与图像本身的梯度对齐，进而适应框级标签监督信息粗糙和遥感图像目标边界复杂的问题。如图1所示，先验信息驱动体系整合了上述约束和像素区分难度表征函数，能够为MPD-WSIS-Net提供完善的高质量监督，实现遥感图像实例分割任务。接下来，将分别介

绍任务先验和图像先验的具体实现方式。

2.2 任务先验

2.2.1 框—掩码投影一致性约束

在实例分割任务中，目标的掩码预测任务和其边界框预测任务是紧密关联的。不难发现，图像中目标的掩码在某一方向（水平或垂直方向）上的投影应该与其边界框在相同方向上的投影相同（Tian 等，2021）。据此，设计框—掩码投影一致性约束，使预测掩码的最小边界矩形与对应的边界框相匹配，进而确定掩码尺寸。

如图2所示，对于待分割图像 I ，将其注释边界框和对应的实例分割结果分别表示为 B_{GT} 和 M 。然后，分别计算注释框和实例分割结果在水平和垂直方向上的投影，并将注释框在水平和垂直方向上的投影分别表示为 B_h 和 B_v ，将预测掩码 M 在水平和垂直方向上的投影分别表示为 M_h 和 M_v 。接着，通过最小化注释框水平投影 B_h 和预测掩码水平投影 M_h 的差别来构建水平方向上的约束，同理，使用注释框垂直投影 B_v 和预测掩码垂直投影 M_v 建立垂直方向上的约束，并将两个方向上的投影结合，共同构成框—掩码投影一致性约束。在MPD-WSIS-Net中，通过损失函数 L_p 具体化该约束，具体计算方式如下：

$$L_p = L_{dic}(B_h, M_h) + L_{dic}(B_v, M_v) \quad (1)$$

式中， L_{dic} 为Dice损失函数（Milletari 等，2016），用于适应遥感图像中存在的样本不平衡的影响。

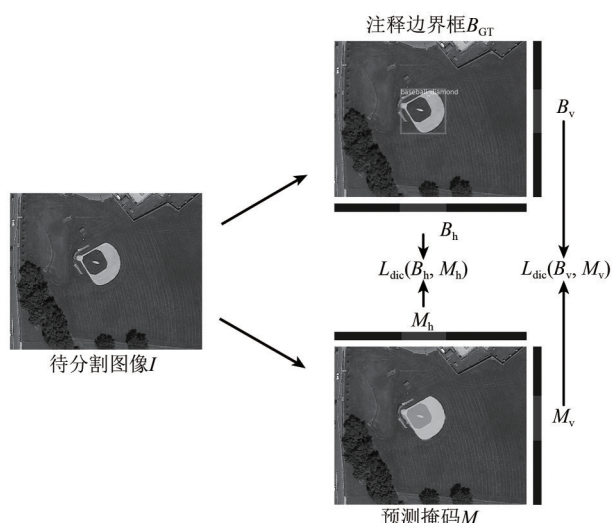


图2 框—掩码投影一致性约束示意图

Fig. 2 Schematic representation of the box-mask projection consistency constraint

2.2.2 像素区分难度表征函数

相比于自然图像，遥感图像中前景和背景的混杂程度更高，会给它们的区分带来难度。进一步观察可知，图像中难以区分的像素点大多集中在目标边界附近和多个目标的堆叠区域，因此有必要在掩码预测过程中对这些不确定程度较高的像素点给予较多的关注（Ma 等，2022）。由于目标的掩码预测任务与其边界框预测任务具有较大的关联性，因此可以考虑使用边界框近似代替目标，即直接关注边界框附近和多个边界框堆叠区域的像素点。

论文通过边界框注释构建像素区分难度表征函数，用于反映目标像素点的不确定程度。以像素点 i 为例，其像素区分难度表征函数 $F(i)$ 主要包含两个部分，即像素点与边界框注释的距离因素 $D(i)$ 和重叠在像素点上的边界框注释的数量因素 $N(i)$ 。其中，距离因素 $D(i)$ 通过高斯核函数计算，其计算方式如式（2）所示：

$$D(i) = 1 + C(i, r) * G(r, \sigma^2) \quad (2)$$

式中， $C(i, r)$ 表示以像素点 i 为中心， r 为半径的计数区域； $*$ 表示卷积操作； $G(r, \sigma^2)$ 表示半径为 r ，方差为 σ 的高斯核。数量因素 $N(i)$ 则简单地表示为

$$N(i) = 1 + N_{\text{box}}(i) \quad (3)$$

式中， $N_{\text{box}}(i)$ 表示包含像素点 i 的边界框注释的数量。结合 $D(i)$ 和 $N(i)$ 可得到最终的像素区分难度表征函数 $F(i)$ ，如下所示

$$F(i) = 1 + \mu \left(f \left(\frac{D(i) \cdot N(i)}{\max_{p \in P_{\text{ins}}} (D(i) \cdot N(i))} \right) - 0.5 \right) \quad (4)$$

式中， f 为Sigmoid函数， P_{ins} 表示边界框注释内部所有的像素点的集合， μ 为缩放比例。在网络训练过程中，该函数能够与邻域视觉一致性约束进行组合，促进网络对边界框附近和多个边界框堆叠区域的像素点给予较多的关注。在2.3.1节中，将对邻域视觉一致性约束的实现方式进行详细介绍。

2.2.3 中心位置先验约束

通过进一步观察可以发现，遥感图像中目标的中心位置有极大概率属于前景，因此可考虑在分割过程中利用这一先验条件构建中心位置先验约束，促进目标和背景分离。

考虑到弱监督实例分割任务缺乏精确的目标

掩码,因此在构建中心位置先验约束的过程中仍然使用边界框近似反映目标。如图3所示,首先在注释边界框 B_{GT} 的中心位置构建一个直径为 d_{cen} 的圆形区域,记为 S_{cen} ,并将该区域视为高置信度的前景区域。相应地, B_{GT} 中的剩余部分即为前背景共存的区域。接着,通过损失函数 L_c 具体化中心位置先验约束,使高置信度的前景区域尽量分配为目标,表示为

$$L_c = \frac{2 \sum_{i \in P_{cen}} M(i) \cdot B(i)}{\sum_{i \in P_{cen}} (M^2(i) + B^2(i))} \quad (5)$$

式中, P_{cen} 为 S_{cen} 中所有像素点的集合。该损失函数能够驱动MPD-WSIS-Net对高置信度的区域进行合理的掩码预测,是任务先验组件的重要组成部分。

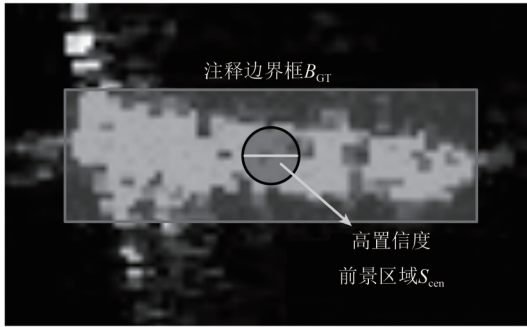


图3 高置信度前景区域示意图

Fig. 3 Schematic representation of the high-confidence foreground region

2.3 图像先验

2.3.1 邻域视觉一致性约束

除了注释先验之外,图像本身的信息也能为弱监督实例分割任务提供丰富的先验信息。直观地,图像中目标和背景通常存在一定的视觉差异,但是目标(或背景)本身的视觉差异性往往较小。因此,对于位置邻近的两个像素位置,若它们的视觉相似性较高,则极有可能同属于目标(或背景)(Chen等,2023)。基于上述的图像本身的视觉关系可构建邻域视觉一致性约束,该约束倾向于将图像中邻近且具有较高视觉相似度的像素点分配到同一类别中。此外,本研究还将任务先验中的像素区分难度表征函数融入到邻域视觉一致性约束中,对图像中的感兴趣目标和背景进行高

效地区分。

对于待分割图像 I 中的每个像素点 i ,均存在由8个相邻像素点组成的邻域。若将该邻域定义推广到图像 I 中的全体像素点,可得到8张邻域图像 $\tilde{I}_n$,其中 $n = 1, 2, \dots, 8$ 。然后,设置视觉相似度阈值 ω ,作为判断两个像素位置是否是属于同一类别的依据。若图像 I 和其邻域图像 $\tilde{I}_n$ 在像素点 i 处的视觉相似度大于 ω ,则认为它们在像素点 i 处属于相同类别,即图像 I 和其邻域图像 $\tilde{I}_n$ 在像素位置 i 处具有相同类别的概率 $P'_n(i)$ 为1。同理,对于预测掩码 M ,同样构建由8个不同的像素位置组成的邻域,并同样推广全体像素点得到8张邻域图像 $\tilde{M}_n$ 。接着,计算预测掩码 M 和其邻域图像 $\tilde{M}_n$ 在像素点 i 处具有相同类别的概率 $P_n^M(i)$,计算方式如下:
 $P_n^M(i) = M(i) \cdot \tilde{M}_n(i) + (1 - M(i)) \cdot (1 - \tilde{M}_n(i))$ (6)

进一步地,在 $P'_n(i)$ 和 $P_n^M(i)$ 的基础上,构建邻域视觉一致性约束,驱动MPD-WSIS-Net在无像素级注释的条件下实现感兴趣目标和背景的区别。值得注意的是,本研究在构建邻域视觉一致性约束时巧妙地结合了注释先验中的像素区分难度表征函数 $F(i)$,用于对不确定程度较高的像素点进行更多的关注,以适应遥感图像中前背景区分难度大的问题。最终的邻域视觉一致性约束通过损失函数 L_n 进行实例化,表示为

$$L_n = -\frac{1}{N_{all}} \sum_n \sum_i F(i) P'_n(i) \log(P_n^M(i)) \quad (7)$$

式中, N_{all} 为注释框中被判定为属于同一类别的像素对的总数目, P_{ins} 为边界框注释内部所有的像素点的集合, $F(i)$ 为像素点 i 处的像素区分难度表征函数。

2.3.2 梯度一致性约束

遥感图像中的船舶等物体通常存在形状不规则和轮廓复杂的特点,而弱监督模式下边界框标签在网络的训练过程中能提供的监督信息是粗糙的,会导致网络对边缘的敏感性不足,从而对实例分割的准确性造成影响。考虑到目标的梯度信息能在一定程度上反映其轮廓(Hao等,2021),因此本研究对图像中目标的梯度信息进行挖掘,建立如图4所示的梯度一致性约束。该约束能够使预测掩码的梯度与图像的梯度对齐,驱动网络预测出合适的目标轮廓,进而更好地适应对边缘敏感程度不足的边界框注释。

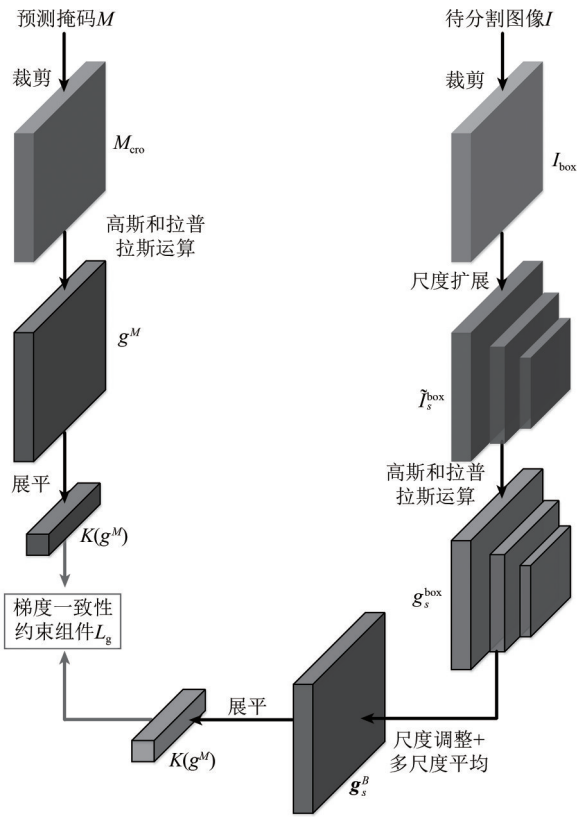


图4 梯度一致性约束构建流程示意图

Fig. 4 Schematic representation of the gradient consistency constraint construction process

如图4所示,对于预测掩码 M ,首先以其对应的预测边界框 B_{pre} 为基础进行裁剪,得到裁剪后的掩码 M_{cro} 。然后,采用高斯和拉普拉斯运算提取掩码的梯度信息,提取过程可表示为式(8):

$$\mathbf{g}^M = |L(G(M_{cro}))| \quad (8)$$

式中, $\mathbf{g}^M$ 为掩码 M_{cro} 对应的梯度信息, G 和 L 分别表示高斯和拉普拉斯运算。

类似地,对于待分割图像 I ,首先以边界框注释为基础对其进行裁剪,得到边界框尺寸的目标图像 I_{box} 。然后,将目标图像 I_{box} 进行尺度扩展,得到 S 个不同尺度的目标图像 $\tilde{I}_s^{box}$,其中 $s = 1, 2, \dots, S$ 。相比于原始单一尺度的目标图像 I_{box} ,尺度扩展后的目标图像 $\tilde{I}_s^{box}$ 可以提供更加丰富的表征。接着,采用高斯和拉普拉斯运算提取目标图像 $\tilde{I}_s^{box}$ 的梯度信息,即:

$$\mathbf{g}_s^{box} = R\left(\left|L\left(G\left(\tilde{I}_s^{box}\right)\right)\right|\right) \quad (9)$$

式中, $\tilde{I}_s^{box}$ 表示第 s 尺度的目标图像, $\mathbf{g}_s^{box}$ 为第 s 尺度的目标图像 $\tilde{I}_s^{box}$ 对应的尺度信息, R 为基于双线性插值的尺度调整操作,用于将变换后的图像尺

度调整为与掩码 M_{cro} 的尺度相同。进一步地,采用简单的多尺度取平均的方式获取最终的梯度信息 $\mathbf{g}_s^B$,表示为

$$\mathbf{g}_s^B = \frac{1}{S} \sum_{s=1}^S \mathbf{g}_s^{box} \quad (10)$$

式中, S 是用于梯度提取的目标图像的尺度数量。

最后,结合预测掩码和待分割图像的梯度信息,构建梯度一致性约束。具体地,使用 $\mathbf{g}^M$ 和 $\mathbf{g}_s^B$ 的负余弦相似度建立如式(11)所示的损失函数 L_g 来具体化梯度一致性约束:

$$L_g = -\frac{K(\mathbf{g}^M)^T K(\mathbf{g}_s^B)}{\|\mathbf{g}^M\| \|\mathbf{g}_s^B\|} \quad (11)$$

式中, K 为展平操作,负号用于确保训练过程中预测掩码和待分割图像两方面的梯度信息的点积最大化。在训练过程中,通过最小化上述梯度一致性约束,能够使预测掩码和待分割图像的梯度保持对齐,驱动网络预测出合适的目标轮廓,进而适应边界框注释对边缘敏感程度不足的限制。

2.4 先验信息驱动体系

通过整合框—掩码投影一致性约束、像素区分难度表征函数、中心位置先验约束、邻域视觉一致性约束和梯度一致性约束,可将MPD-WSIS-Net中的先验信息驱动体系 L_{pri} 表示为

$$L_{pri} = L_p + L_c + L_n + L_g \quad (12)$$

因此,MPD-WSIS-Net中的先验信息驱动体系由一系列约束和像素区分难度表征函数进行实例化,全面地包含注释先验和任务先验中的多种先验信息。

在网络训练的过程中,MPD-WSIS-Net首先将待分割图像 I 进行前向传递,得到对应的实例分割结果 M 。然后,利用实例分割结果 M 和对应的注释边界框 B_{GT} ,计算出式(12)中的 L_{pri} 。接着,将 L_{pri} 视为掩码损失函数,并据此计算梯度信息,将误差从网络的输出层向输入层传播。在反向传播过程中,根据随机梯度下降法SGD (Stochastic Gradient Descent) 更新网络的参数,实现对网络的监督。重复上述前向传递、 L_{pri} 计算和反向传播的步骤,直到达到指定的训练迭代次数时停止。上述训练过程通过实例分割结果 M 和对应的注释边界框 B_{GT} 计算 L_{pri} ,并结合反向传播优化MPD-WSIS-Net,能够在无像素级注释的条件下为网络提供掩码方面的监督信息,驱动网络实现遥感图

像的分割。

2.5 网络结构

网络结构方面, MPD-WSIS-Net 是以经典的 FCOS 目标检测器 (Tian 等, 2019) 为基础实现的, 主要由特征提取网络和任务网络两部分组成。其中, 特征提取网络用于对输入图像进行多尺度特征提取, 获取多尺度特征图; 任务网络则用于对多尺度特征图进一步处理。在网络的训练过程中, 先验信息驱动体系能够为网络提供任务先验和图像先验两方面的监督信息, 促进网络对遥感图像中感兴趣目标的掩码进行预测。

如图 5 所示, 特征提取网络由常规的 ResNet (He 等, 2016) 和 FPN (Lin 等, 2017) 组成; 任务网络由分类分支、中心点分支、边界框分支和掩码分支组成。其中, 类别分支、中心点分支和边界框分支用于实现目标检测部分的功能, 其结

构与 FCOS 目标检测器相同。掩码分支的结构主要继承自 CondInst (Tian 等, 2020), 主要由控制器、底层特征转换器和紧凑 FCN (Fully Convolutional Network) (Long 等, 2015) 3 部分组成。其中, 控制器重复用于 FPN 输出的不同尺度的特征图, 能够预测出紧凑 FCN 的参数; 特征转换器与 FPN 的最底层特征图 P3 相连接, 由 4 个 3×3 卷积和一个 1×1 卷积组成, 能够将底层特征图 P3 转变为紧凑 FCN 的输入; 紧凑 FCN 包含参数由控制器预测的 3 个 1×1 卷积层, 能够将特征转换器转换后的特征图进行进一步处理, 得到相应的分割结果。不同于 Mask R-CNN (He 等, 2020) 的分割分支直接通过 FCN 进行掩码提取, 本研究采取的掩码分支能够根据不同的实例动态地调整紧凑 FCN 的权重, 具有良好的灵活性。图 5 还展示了 MPD-WSIS-Net 的先验信息驱动过程, 该过程除了通过前向传递预测目标的掩码外, 还包含 L_{pri} 计算和反向传播步骤。

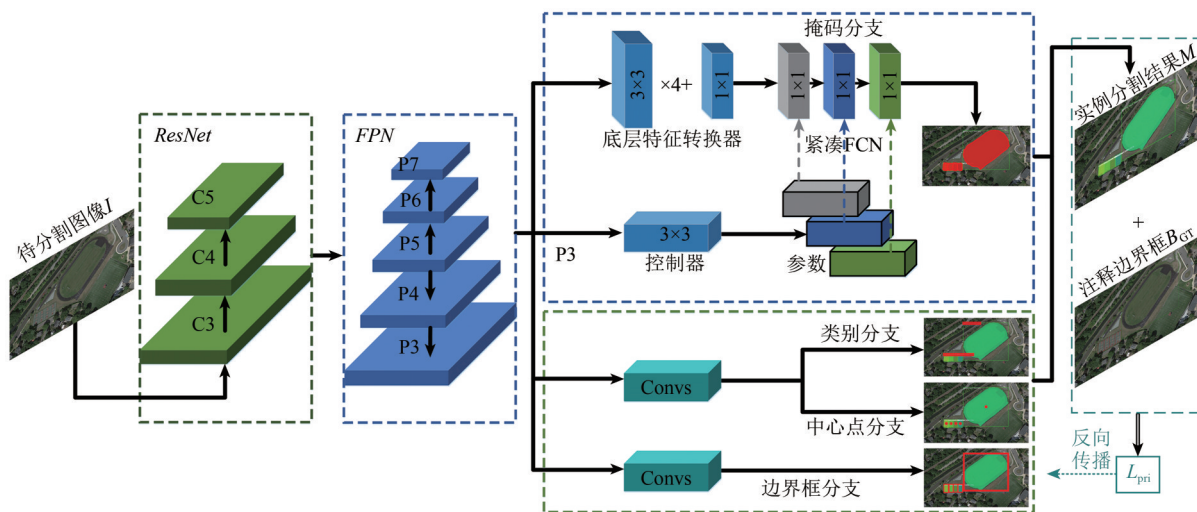


图 5 MPD-WSIS-Net 的网络结构和先验信息驱动过程示意图

Fig. 5 Schematic diagram of the network structure of MPD-WSIS-Net and the driving process of the prior information

3 实验结果与分析

3.1 实验数据集

采用 NWPU VHR-10 数据集 (Cheng 等, 2016) 和 SSDD 数据集 (Su 等, 2020) 验证所提的弱监督实例分割方法。其中, NWPU VHR-10 数据集共包含 800 张光学遥感图像, 包括 10 个类别, 即飞机、棒球场、地面田径场、车辆、船只、网球场、海港、储油罐、篮球场和桥梁。SSDD 数据集包括 1160 张 SAR 遥感图像, 涵盖多种极化方式。类似

于 MGWI-Net (Chen 等, 2023), 两个数据集在实验过程中均按照 7 : 3 的比例随机分配训练集和测试集。此外, 像素级标签仅在 MPD-WSIS-Net 的测试过程中使用, 训练过程不使用像素级标签。

3.2 实验设置

实验硬件是处理器为 Intel Xeon Gold 6226R CPU@2.90 GHz 的服务器, GPU 为 NVIDIA Tesla V100, 骨干网络中的 ResNet (He 等, 2016) 为在 ImageNet 上预训练的 ResNet-101。本研究的对比实验遵循文献 (Chen 等, 2023)。网络的动量为

0.9，权重衰减为 $1\text{E-}4$ 。共进行12轮的训练，批处理大小设为4，基础学习率设置为 $2.5\text{E-}3$ ，在第8轮和第11轮时学习率分别减少10倍。考虑到遥感图像中目标的尺寸变化较大，在训练阶段将输入图像的短边随机调整为600、700、800、900和1000中的值，长边统一设置为1333。邻域视觉一致性约束中视觉相似度的计算方式与MGWI-Net (Chen 等, 2023) 相同，阈值 ω 取值为0.3；像素区分难度表征函数中的缩放比例 μ 设置为5；中心位置先验约束中的高置信度前景区域 S_{cen} 的直径 d_{cen} 设置为其对应注释边界框的短边的十分之一。

本研究在可见光和SAR遥感图像数据集上进行了广泛的实验，以充分验证所提出的MPD-WSIS-Net。具体而言，实验分为对比实验和消融实验两部分。在对比实验方面，首先将MPD-WSIS-Net与其他先进的弱监督实例分割方法进行比较，主要关注分割效果、处理速度和参数量等方面的差异，以验证该方法的总体性能。接着，本研究还将MPD-WSIS-Net与标注成本更高的混合监督 and 全监督的实例分割方法进行比较，以进一步说明该方法在实例分割方面的能力。在消融实验方面，本研究对MPD-WSIS-Net中任务和图像两部分的先验信息组件进行了详细的消融研究，以考察这些组件在遥感图像实例分割任务中的贡

献。此外，本研究还对MPD-WSIS-Net中关键的参数进行了消融研究，以讨论这些参数对实例分割精度的影响。

3.3 评价指标

本研究采用常见的MS COCO (Lin 等, 2014) 评价指标对光学和SAR遥感图像的实例分割方法进行评估，主要包括AP、 AP_{50} 、 AP_{75} 、 AP_s 、 AP_m 和 AP_L ，其中 AP_s 、 AP_m 和 AP_L 分别为针对小目标、中等目标和大目标分别设计的评价指标。此外，本研究还报告了模型的性能。和参数量 N_{par} ，以全方面地反映模型的性能。

3.4 与各类型方法的对比实验

3.4.1 与弱监督方法的对比实验

为了验证MPD-WSIS-Net在遥感图像实例分割任务上的效果，将其与多种弱监督实例分割方法进行对比，定量和定性实验结果分别如表1和图6所示。本研究选取的用于对比的弱监督实例分割方法包括BoxInst (Tian 等, 2021)、DiscoBox (Lan 等, 2021)、DBIN (Li 等, 2022) 和MGWI-Net (Chen 等, 2023)。值得注意的是，考虑到本文未涉及域自适应方面的研究，因此对DBIN中域自适应部分的组件做出了删减。

表1 与弱监督方法对比的定量实验结果

Table 1 Quantitative experimental results comparing with weakly supervised methods

数据集	方法	AP/%	$\text{AP}_{50}/\%$	$\text{AP}_{75}/\%$	$\text{AP}_s/\%$	$\text{AP}_m/\%$	$\text{AP}_L/\%$	FPS/ N_{par}
NWPU VHR-10	BoxInst	47.6	78.9	49.0	33.8	43.9	55.5	10.6
	DiscoBox	46.2	79.7	47.4	29.4	42.9	57.1	11.0
	DBIN	48.3	80.2	50.5	34.5	46.1	57.0	10.1
	MGWI-Net	51.6	81.3	53.3	37.6	48.2	59.1	10.4
	MPD-WSIS-Net	52.5	81.9	54.6	38.0	48.9	60.3	10.7
SSDD	BoxInst	49.9	90.1	52.7	50.6	52.3	—	53.5
	DiscoBox	48.4	90.2	50.4	47.2	50.6	—	65.0
	DBIN	50.6	91.7	52.8	51.3	52.0	—	55.6
	MGWI-Net	53.0	92.4	57.1	53.7	54.9	—	53.7
	MPD-WSIS-Net	54.1	92.7	58.9	55.2	56.5	—	53.6

注：由于SSDD数据集中大尺寸目标的数量少，导致 AP_L 指标有较大的随机性，故不报告该指标，相应位置用“—”表示。

如表1所示，在NWPU VHR-10数据集上，MPD-WSIS-Net相比于其他弱监督实例分割方法能够取得一定的分割优势。在用于对比的分割方法中，BoxInst、DBIN和MGWI-Net只能通过不全面

的先验信息驱动网络，因此对框监督实例分割任务中的先验信息无法进行充分的挖掘和利用，分割效果相比于MPD-WSIS-Net较差；DiscoBox通过联合任务辅助的方式进行弱监督实例分割，该方

法在光学遥感图像上的分割效果相比于BoxInst等利用先验信息的方法存在一定的差距。另一方面,得益于其从任务和图像两个角度对先验信息的全面利用,MPD-WSIS-Net在SSDD数据集上的表现相比于其他专门的弱监督实例分割方法也更好,AP值比BoxInst高4.2。此外,表1中最右列的上半部分记录了各方法的处理速度,下半部分报告了

方法的参数量。可以看出,MPD-WSIS-Net在处理速度和参数量方面也保持在合理的水平。结合本研究中设计的先验信息驱动体系可知,MPD-WSIS-Net能够从任务和图像两个方面对先验信息进行充分且合理的挖掘和利用,因此能够对光学和SAR遥感图像实现良好的实例分割效果。

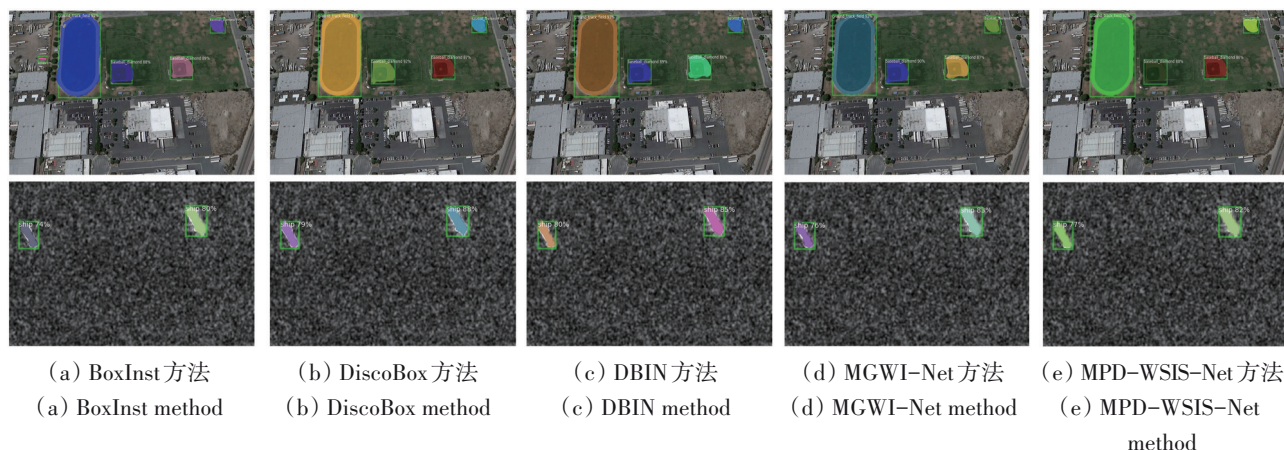


图6 定性实验结果(第一行和第二行分别来自NWPU VHR-10和SSDD数据集)

Fig. 6 Qualitative experimental results (The first and second rows are from the NWPU VHR-10 and SSDD datasets, respectively)

图6中展示了NWPU VHR-10和SSDD数据集上各方法的定性实验结果。图6(a)–(d)分别为用于对比的弱监督方法中BoxInst、DiscoBox、DBIN和MGWI-Net的分割效果,图6(e)为MPD-WSIS-Net的分割效果。得益于全面的先验信息驱动体系,MPD-WSIS-Net相较于其他专门的弱监督实例分割方法能够实现更加优越的细节处理效果,其预测掩码能够精确地覆盖目标,未出现明显的欠分割和过分割现象。总体来说,在相同的标签制作成本下,MPD-WSIS-Net能够具有更好的分割性能。

3.4.2 与混合监督 and 全监督方法的对比实验

除了弱监督实例分割方法之外,还设计了混合监督 and 全监督的实例分割方法,用于与本研究提出的MPD-WSIS-Net进行全面地对比,定量和定性实验结果分别见表2和图7、图8。其中,混合监督实例分割方法是弱监督方法和全监督方法之间的过渡方法,其注释成本在二者之间。全监督实例分割方法在训练期间的像素级注释占比 R_{pix} 为100%,注释成本最高,是目前遥感图像实例分割领域应用最多的方法。在与混合监督 and 全

监督方法的对比实验中,本研究选取YOLACT (Bolya等, 2019)、Mask R-CNN (He等, 2020)和CondInst (Tian等, 2020)作为基准。其中,YOLACT为具有代表性的快速分割方法;Mask R-CNN为经典的实例分割方法;CondInst以FCOS目标检测器为基础构建的典型分割方法,这3种方法均具有良好的代表性。

表2上半部分展示了混合监督 and 全监督实例分割方法在NWPU VHR-10数据集上的分割表现,其中 R_{pix} 为训练期间像素级注释占比。相比于像素级注释占比为50%和75%时的混合监督实例分割方法,所提的MPD-WSIS-Net能够以更少的注释成本获得更高的AP值,这进一步体现了MPD-WSIS-Net优秀的弱监督分割性能。此外,相比于全监督的YOLACT,MPD-WSIS-Net也展现出了一定的分割优势,其AP值能够达到全监督Mask R-CNN的89.3%,表明这种由先验信息驱动的弱监督分割方法在分割表现上接近全监督方法。表2下半部分报告了SSDD数据集中混合监督 and 全监督方法的分割效果。可以看出,MPD-WSIS-Net的分割表现显著优于像素级标签占比为50%时的混合监督实例分割方法,且与75%像素级注释条件下的Mask

R-CNN 和 CondInst 接近。另一方面，相比于全监督的 YOLACT，MPD-WSIS-Net 拥有显著的优势，并且能够达到全监督 Mask R-CNN 的 84.3%，这说明了所提的 MPD-WSIS-Net 在不依赖昂贵的像素级标签的条件下即可对 SAR 遥感图像数据集实现良好的实例分割效果。综上所述，相比于注释成本更加昂贵的混合监督和全监督实例分割方法，MPD-WSIS-Net 也具有强大的竞争力。

表 2 与混合监督和全监督方法对比的定量实验结果

Table 2 Quantitative experimental results of comparison with hybrid supervised and fully supervised methods

数据集	方法	R_{pix}	AP	AP ₅₀	AP ₇₅	AP _s	AP _m	AP _L
NWPU VHR-10	YOLACT	50	22.5	49.7	17.0	9.6	19.9	31.5
	Mask R-CNN	50	35.5	70.8	31.3	24.6	34.2	39.9
	CondInst	50	34.5	73.4	27.6	23.7	34.1	35.9
	YOLACT	75	27.5	54.4	27.4	12.1	25.9	34.2
	Mask R-CNN	75	49.3	82.6	51.7	36.9	47.0	53.9
	CondInst	75	49.5	85.1	50.3	35.9	48.6	53.7
	YOLACT	100	35.6	68.4	36.4	14.8	33.3	56.0
	Mask R-CNN	100	58.8	86.6	65.2	47.1	57.5	62.4
	CondInst	100	58.5	90.1	62.9	29.4	56.8	71.3
	MPD-WSIS-Net	0	52.5	81.9	54.6	38.0	48.9	60.3
SSDD	YOLACT	50	28.6	76.7	9.0	32.1	34.2	—
	Mask R-CNN	50	39.3	86.2	28.0	42.7	44.4	—
	CondInst	50	38.4	87.4	28.6	41.3	43.8	—
	YOLACT	75	39.0	79.9	32.5	40.3	45.5	—
	Mask R-CNN	75	54.6	90.2	63.0	56.6	57.1	—
	CondInst	75	54.1	93.0	59.6	54.6	56.8	—
	YOLACT	100	44.6	86.6	41.0	45.3	48.5	—
	Mask R-CNN	100	64.2	94.9	80.1	62.0	64.7	—
	CondInst	100	63.0	95.9	78.4	63.7	63.6	—
	MPD-WSIS-Net	0	54.1	92.7	58.9	55.2	56.5	—

注：由于 SSDD 数据集中大尺寸目标的数量少，导致 AP_L 指标有较大的随机性，故不报告该指标，相应位置用“—”表示。

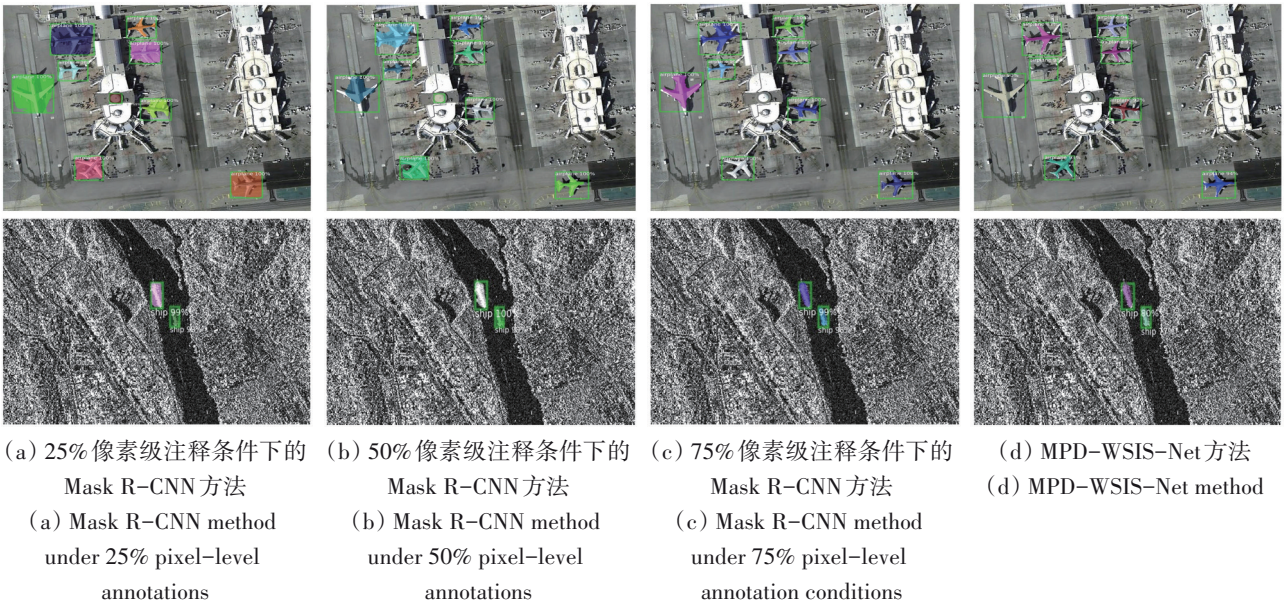


图 7 与混合监督方法对比的定性实验结果(第一行和第二行分别来自 NWPU VHR-10 和 SSDD 数据集)

Fig. 7 Qualitative experimental results of comparison with mixed supervised methods (The first and second rows are from the NWPU VHR-10 and SSDD datasets , respectively)

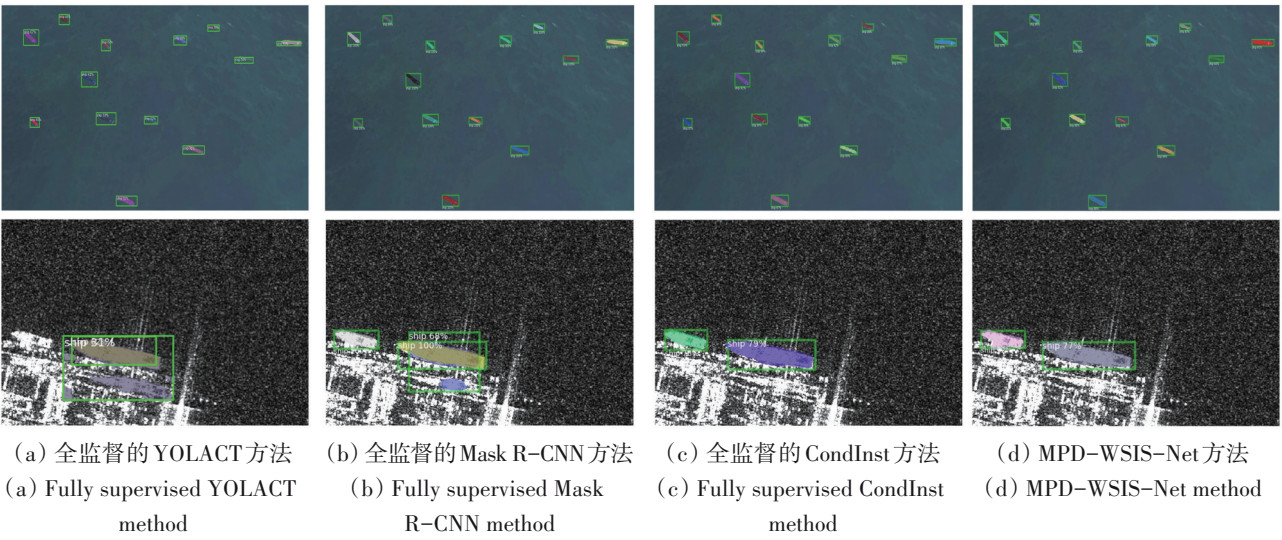


图8 与全监督方法对比的定性实验结果(第一行和第二行分别来自NWPU VHR-10和SSDD数据集)
Fig. 8 Qualitative experimental results of comparison with fully supervised methods (The first and second rows are from the NWPU VHR-10 and SSDD datasets, respectively)

图7直观地展示了MPD-WSIS-Net与混合监督范式方法在光学和SAR遥感数据集中的分割效果对比,图7(a)、(b)和(c)分别为像素级标签占比为25%、50%和75%时的分割效果。可以看出,在像素级注释占比为25%时,混合监督范式的Mask R-CNN出现了明显的过分割现象,其掩码质量较差。随着像素级标签的增加,混合监督方法的实例掩码质量逐渐改善,但是跟图7(d)所示的MPD-WSIS-Net相比仍有一定的差距。图8反映了MPD-WSIS-Net和全监督范式方法在光学和SAR遥感图像数据集中的定性实验结果,图8(a)–(c)分别为全监督的YOLACT、Mask R-CNN和CondInst的分割效果。由于全监督范式的像素级标签丰富,因此该范式下的实例分割效果良好。相比于这些方法,MPD-WSIS-Net在无像素级注释的

情况下提取出的掩码也非常准确,无明显的过分割和欠分割现象。总体来说,MPD-WSIS-Net能够使光学和SAR遥感图像的实例分割不再依赖于复杂的像素级标签,实现了标签制作成本和分割效果之间的平衡,是一种低成本的实例分割方案。

3.5 消融实验

本研究也在光学和SAR遥感图像数据集上进行了详细地消融实验,以考察MPD-WSIS-Net中的各种先验信息对应组件在遥感图像实例分割任务方面的贡献,实验结果见表3和表4。考虑到MPD-WSIS-Net中的像素区分难度表征函数与邻域视觉一致性约束组合在一起,因此设计退化版的邻域视觉一致性约束组件 $L_{n-basic}$,该组件在原始邻域视觉一致性约束 L_n 的基础上移除了像素区分难度表征函数。

表3 任务先验信息组件对MPD-WSIS-Net分割效果的影响

Table 3 Impact of task prior components on MPD-WSIS-Net segmentation effectiveness									
数据集	模型	$F(i)$	L_c	AP/%	AP ₅₀ /%	AP ₇₅ /%	AP _s /%	AP _M /%	AP _L /%
NWPU VHR-10	基线	—	—	49.8	80.7	51.0	35.2	46.1	57.4
	组件	√	—	50.4	81.1	51.9	35.6	46.7	58.2
		—	√	50.7	80.9	51.4	36.2	46.5	57.8
		√	√	51.3	81.2	52.5	36.9	47.4	58.9
SSDD	基线	—	—	51.8	91.9	54.0	52.7	53.2	—
	组件	√	—	52.3	92.0	55.8	53.3	54.1	—
		—	√	52.2	92.3	55.5	53.6	54.8	—
		√	√	53.2	92.5	56.7	54.1	55.3	—

注: 由于SSDD数据集中大尺寸目标的数量少,导致AP_L指标有较大的随机性,故不报告该指标,相应位置用“—”表示。

表4 图像先验信息组件对MPD-WSIS-Net分割效果的影响
Table 4 Impact of image prior components on MPD-WSIS-Net segmentation effectiveness

数据集	模型	AP	AP ₅₀ /%	AP ₇₅ /%	AP _s /%	AP _M /%	AP _L /%
NWPU VHR-10	基线	29.8	62.9	25.8	14.7	27.4	39.0
	+ $L_{n-basic}$	49.8	80.7	51.0	35.2	46.1	57.4
	+ $L_{n-basic}$ + L_g	52.0	81.4	53.9	37.3	48.4	59.7
SSDD	基线	38.1	88.5	21.8	39.8	42.7	—
	+ $L_{n-basic}$	51.8	91.9	54.0	52.7	53.2	—
	+ $L_{n-basic}$ + L_g	53.5	92.4	57.8	54.3	55.9	—

注：由于SSDD数据集中大尺寸目标的数量少，导致AP_L指标有较大的随机性，故不报告该指标，相应位置用“—”表示。

任务先验信息各组件在NWPU VHR-10数据集上的消融实验结果如表3上半部分所示，其基线为框—掩码投影一致性约束和退化版邻域视觉一致性约束共同驱动的MPD-WSIS-Net。由实验结果可知，随着各种先验信息对应组件的引入，MPD-WSIS-Net在NWPU VHR-10数据集上的表现得到逐步提升。当MPD-WSIS-Net只包含框—掩码投影一致性约束组件 L_p 和退化版的邻域视觉一致性约束组件 $L_{n-basic}$ 时，网络对先验信息的挖掘不够充分，因此仅能取得49.8的AP值。引入像素区分难度表征函数后，MPD-WSIS-Net能够通过该组件对不确定程度较高的像素点给与更多的关注，因此在光学遥感图像数据集上的分割表现得到一定的提升。引入中心位置先验约束后，MPD-WSIS-Net的分割性能也能得到改进。表3下半部分报告的在SSDD数据集中消融实验结果也能得到类似的结论，说明任务先验信息组件对遥感图像弱监督实例分割任务的积极作用。

图像先验信息的消融实验结果如表4所示，其基线为框—掩码投影一致性约束单独驱动的MPD-WSIS-Net。由实验结果可知，将退化版的邻域视觉一致性约束组件 $L_{n-basic}$ 引入到MPD-WSIS-Net后，各项实例分割指标得到大幅提升，表明图像先验中的领域视觉一致性约束对目标的分割起到重要作用。继续引入梯度一致性约束组件 L_g 后，MPD-WSIS-Net在光学和SAR遥感图像数据集中的AP值分别提升了2.2和1.7，表明这种使预测掩码与图像的梯度对齐的方式，能够使网络得到更好的分割效果。综上所述，MPD-WSIS-Net中的各先验信息组件对光学遥感图像的实例分割性能均有着积极的影响，由框—掩码投影一致性约束、像素区分难度表征函数、中心位置先验约束、邻

域视觉一致性约束和梯度一致性约束共同构成的先验信息驱动体系能够驱动MPD-WSIS-Net实现光学和SAR遥感图像的实例分割任务。

进一步地，本研究还对MPD-WSIS-Net中关键的参数进行消融研究，以讨论这些参数对实例分割精度的影响。具体地，考虑到邻域视觉一致性约束对目标的分割起到重要作用，本文对该约束中的关键参数视觉相似度阈值 ω 进行消融研究，结果见表5。此外，本文还对梯度一致性约束中用于梯度提取的目标图像尺度数量 S 进行消融研究，以探究尺度扩展对分割精度的影响，实验结果如表6所示。

表5 邻域视觉一致性约束中视觉相似度阈值的消融实验
Table 5 Ablation study of visual similarity thresholds in the neighborhood visual consistency constraint

数据集	$\omega=0.1$	$\omega=0.2$	$\omega=0.3$	$\omega=0.4$	$\omega=0.5$
NWPU VHR-10	51.3	52.6	52.5	51.7	51.1
SSDD	53.2	53.9	54.1	53.1	52.9

表6 梯度一致性约束中目标图像的尺度数量的消融实验
Table 6 Ablation study of the number of scales of the target image in the gradient consistency constraint

数据集	$S=1$	$S=2$	$S=3$	$S=4$	$S=5$
NWPU VHR-10	51.0	51.8	52.8	52.7	52.5
SSDD	52.8	53.5	54.1	54.0	54.3

由表5可知，视觉相似度阈值 ω 过大或者过小时，MPD-WSIS-Net的分割效果都会变差。通过对邻域视觉一致性约束进行分析可知：当 ω 较小时，像素对的区分能力会变弱，这可能导致两个属于不同类别的像素点被误认为属于同一类别；当 ω 较大时，则可能会使属于同一类别的像素对判定失误。因此，视觉相似度阈值 ω 应该取适中的值。综合可见光和SAR遥感图像数据集的实验结果， ω

的取值为0.2或0.3时较为适合,本研究选择0.3作为MPD-WSIS-Net的视觉相似度阈值。

如表6可知,随着梯度一致性约束中目标图像的尺度数量 S 的增加,MPD-WSIS-Net的AP值总体上呈上升趋势,这说明尺度扩展对分割效果的提升具有一定的积极作用。结合梯度一致性约束的设计思路可知,多尺度的目标图像能够产生得更加丰富的表征,使提取出的梯度信息更接近目标真实的梯度信息,从而提升MPD-WSIS-Net的分割性能。进一步观察可知,当 S 取值大于3时,分割效果对该参数的增长不明显。因此,尺寸数量 S 可从3—5中取值。在可见光数据集训练时的 S 的取值为5,SAR遥感图像数据集中选取3作为梯度一致性约束中目标图像的尺寸数量。

4 结 论

本研究针对遥感图像实例分割任务的挑战,构建了一套适用于遥感图像弱监督实例分割任务的先验信息驱动体系,并且提出了一种基于多重先验信息的弱监督实例分割网络。驱动体系将先验信息按照其来源分为任务先验和图像先验两部分,实例分割网络则通过这两部分先验实例化构建出特定的约束和函数,使网络在无像素级注释的条件下完成实例分割任务。在光学和SAR遥感图像数据集上的实验结果表明,该方法无需任何像素级注释即可实现优秀的分割效果,优于现有的弱监督方法,并且达到了与全监督Mask R-CNN接近的水平。在未来的研究中,将尝试将提出的先验信息驱动体系与更先进的实例分割网络结构相结合,并通过进一步挖掘丰富的先验信息来完善该体系,以缩小弱监督与全监督方法的差距,实现更好的分割效果。

参考文献(References)

- Ahn J, Cho S and Kwak S. 2019. Weakly supervised learning of instance segmentation with inter-pixel relations//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE: 2204-2213 [DOI: 10.1109/CVPR.2019.00231]
- Arun A, Jawahar C V and Kumar M P. 2020. Weakly supervised instance segmentation by learning annotation consistent instances//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 254-270 [DOI: 10.1007/978-3-030-58604-1_16]
- Bolya D, Zhou C, Xiao F Y and Lee Y J. 2019. YOLACT: real-time instance segmentation//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 9156-9165 [DOI: 10.1109/ICCV.2019.00925]
- Cai Z W and Vasconcelos N. 2018. Cascade R-CNN: delving into high quality object detection//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE: 6154-6162 [DOI: 10.1109/CVPR.2018.00644]
- Chen M, Zhang Y, Chen E P, Hu Y H, Xie Y F and Pan Z S. 2023. Meta-knowledge guided weakly supervised instance segmentation for optical and SAR image interpretation. Remote Sensing, 15(9): 2357 [DOI: 10.3390/rs15092357]
- Chen Z Q, Shang Y H, Python A, Cai Y X and Yin J W. 2022. DB-BlendMask: decomposed attention and balanced BlendMask for instance segmentation of high-resolution remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 60: 5615915 [DOI: 10.1109/TGRS.2021.3138913]
- Cheng G, Zhou P C and Han J W. 2016. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 54(12): 7405-7415 [DOI: 10.1109/TGRS.2016.2601622]
- Hao S Y, Wang G A and Gu R S. 2021. Weakly supervised instance segmentation using multi-prior fusion. Computer Vision and Image Understanding, 211: 103261 [DOI: 10.1016/j.cviu.2021.103261]
- He K M, Gkioxari G, Dollár P and Girshick R. 2020. Mask R-CNN. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(2): 386-397 [DOI: 10.1109/TPAMI.2018.2844175]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- Hsu C C, Hsu K J, Tsai C C, Lin Y Y and Chuang Y Y. 2019. Weakly supervised instance segmentation using the bounding box tightness prior//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver, BC, Canada: Curran Associates Inc.: 591 [DOI: 10.5555/3454287.3454878]
- Huang L, Han X Y, Wang X L, Zhang Y D, Yang J F, Fen A P, Li J G and Zhu N H N W. 2022. Coupling with high-resolution remote sensing data to evaluate urban non-point source pollution in Tongzhou, China. Science of the Total Environment, 831: 154632 [DOI: 10.1016/j.scitotenv.2022.154632]
- Lan S Y, Yu Z D, Choy C, Radhakrishnan S, Liu G L, Zhu Y K, Davis L S and Anandkumar A. 2021. DiscoBox: weakly supervised instance segmentation and semantic correspondence from box supervision//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, QC, Canada: IEEE: 3386-3396 [DOI: 10.1109/ICCV48922.2021.00339]
- Lee J, Yi J H, Shin C and Yoon S. 2021. BBAM: bounding box attribution map for weakly supervised semantic and instance segmentation//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, TN, USA: IEEE: 2643-2651 [DOI: 10.1109/CVPR46437.2021.00267]

- Li M Q, Fang S B, Zhu Y C, Wu Y J, Cao Y, Zhuo W and E Y H. 2022. Spatial and temporal distributions of waterlogging disasters in the summer of 2021 in Mainland China and their possible impacts. *National Remote Sensing Bulletin*, 26(9): 1886-1894 (李梦倩, 房世波, 朱永超, 武英洁, 曹云, 卓文, 魏有浩. 2022. 2021年夏季中国大陆涝渍灾害时空分布分析. *遥感学报*, 26(9): 1886-1894 [DOI: 10.11834/jrs.20221782])
- Li Y H, Xue Y, Li L F, Zhang X J and Qian X M. 2022. Domain adaptive box-supervised instance segmentation network for mitosis detection. *IEEE Transactions on Medical Imaging*, 41(9): 2469-2485 [DOI: 10.1109/TMI.2022.3165518]
- Li Z L, Wang L Y, Jiang S, Wu Y H and Zhang Q J. 2021. On orbit extraction method of ship target in SAR images based on ultra-light-weight network. *National Remote Sensing Bulletin*, 25(3): 765-775 (李宗凌, 汪路元, 蒋帅, 吴雨航, 张庆君. 2021. 超轻量网络的SAR图像舰船目标在轨提取. *遥感学报*, 25(3): 765-775) [DOI: 10.11834/jrs.20210160]
- Lin T Y, Dollár P, Girshick R, He K M, Hariharan B and Belongie S. 2017. Feature pyramid networks for object detection//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, USA: IEEE: 936-944 [DOI: 10.1109/CVPR.2017.106]
- Lin T Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P and Zitnick C L. 2014. Microsoft COCO: common objects in context//*Proceedings of the 13th European Conference on Computer Vision*. Zurich, Switzerland: Springer: 740-755 [DOI: 10.1007/978-3-319-10602-1_48]
- 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, MA, USA: IEEE: 3431-3440 [DOI: 10.1109/CVPR.2015.7298965]
- Lyu J, Wang S Q, Sun D Y, Nie J W, Jiao H B, Zhang H L and Liang H W. 2022. Remote sensing of spatial and temporal variations of euphotic zone depth in the Bohai Sea and Yellow Sea during recent 20 years (2002—2020). *National Remote Sensing Bulletin*, 26(12): 2507-2517 (吕军, 王胜强, 孙德勇, 聂君伟, 焦红波, 张海龙, 梁涵玮. 2022. 近20年(2002—2020年)渤海黄真光层深度的时空变化遥感分析. *遥感学报*, 26(12): 2507-2517) [DOI: 10.11834/jrs.20210414]
- Lyu S Y, Li J T, A X H, Yang C, Yang R C and Shang X M. 2023. Res_ASPP_UNet++: building an extraction network from remote sensing imagery combining depthwise separable convolution with atrous spatial pyramid pooling. *National Remote Sensing Bulletin*, 27(2): 502-519 (吕少云, 李佳田, 阿晓葵, 杨超, 杨汝春, 尚晓梅. 2023. Res_ASPP_UNet++: 结合分离卷积与空洞金字塔的遥感影像建筑物提取网络. *遥感学报*, 27(2): 502-519) [DOI: 10.11834/jrs.20210477]
- Ma T Q, Wang Q L, Zhang H Z and Zuo W M. 2022. Delving deeper into pixel prior for box-supervised semantic segmentation. *IEEE Transactions on Image Processing*, 31: 1406-1417 [DOI: 10.1109/TIP.2022.3141878]
- Miao W, Geng J and Jiang W. 2022. Semi-supervised remote-sensing image scene classification using representation consistency Siamese network. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 5616614 [DOI: 10.1109/TGRS.2022.3140485]
- Milletari F, Navab N and Ahmadi S A. 2016. V-Net: fully convolutional neural networks for volumetric medical image segmentation//*Proceedings of the 2016 Fourth International Conference on 3D Vision*. Stanford, CA, USA: IEEE: 565-571 [DOI: 10.1109/3DV.2016.79]
- Pont-Tuset J, Arbeláez P, Barron J T, Marques F and Malik J. 2017. Multiscale combinatorial grouping for image segmentation and object proposal generation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(1): 128-140 [DOI: 10.1109/TPAMI.2016.2537320]
- Rother C, Kolmogorov V and Blake A. 2004. "GrabCut": interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics (TOG)*, 23(3): 309-314 [DOI: 10.1145/1015706.1015720]
- Su H, Wei S J, Liu S, Liang J D, Wang C, Shi J and Zhang X L. 2020. HQ-ISNet: high-quality instance segmentation for remote sensing imagery. *Remote Sensing*, 12(6): 989 [DOI: 10.3390/rs12060989]
- Tian Z, Shen C H and Chen H. 2020. Conditional convolutions for instance segmentation//*Proceedings of the 16th European Conference on Computer Vision*. Glasgow, UK: Springer: 282-298 [DOI: 10.1007/978-3-030-58452-8_17]
- Tian Z, Shen C H, Chen H and He T. 2019. FCOS: fully convolutional one-stage object detection//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*. Seoul, Korea (South): IEEE: 9626-9635 [DOI: 10.1109/ICCV.2019.00972]
- Tian Z, Shen C H, Wang X L and Chen H. 2021. BoxInst: high-performance instance segmentation with box annotations//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville, TN, USA: IEEE: 5439-5448 [DOI: 10.1109/CVPR46437.2021.00540]
- Wang X L, Liang Z Y and Liu T. 2023. Feature attention pyramid-based remote sensing image object detection method. *National Remote Sensing Bulletin*, 27(2): 492-501 (汪西莉, 梁正印, 刘涛. 2023. 基于特征注意力金字塔的遥感图像目标检测方法. *遥感学报*, 27(2): 492-501) [DOI: 10.11834/jrs.20235011]
- Waqas Zamir S, Arora A, Gupta A, Khan S, Sun G L, Shahbaz Khan F, Zhu F, Shao F, Xia G S and Bai X. 2019. iSAID: a large-scale dataset for instance segmentation in aerial images. *arXiv*: 1905.12886 [DOI: 10.48550/arXiv.1905.12886]
- Wei S J, Zeng X F, Zhang H, Zhou Z C, Shi J and Zhang X L. 2022. LFG-Net: low-level feature guided network for precise ship instance segmentation in SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 5231017 [DOI: 10.1109/TGRS.2022.3188677]
- Wu Q Q, Wang S, Wang B and Wu Y L. 2022. Road extraction method of high-resolution remote sensing image on the basis of the spatial information perception semantic segmentation model. *National Remote Sensing Bulletin*, 26(9): 1872-1885 (吴强强, 王帅, 王彪, 吴艳兰. 2022. 空间信息感知语义分割模型的高分辨率遥感影

- 像道路提取. 遥感学报, 26(9): 1872-1885 [DOI: 10.11834/jrs.20210021]
- Yang F, Yuan X Y, Ran J, Shu W Q, Zhao Y, Qin A Y and Gao C Q. 2021. Accurate instance segmentation for remote sensing images via adaptive and dynamic feature learning. *Remote Sensing*, 13(23): 4774 [DOI: 10.3390/rs13234774]
- Zhang T, Yang X G, Lu X Q, Lu R T and Zhang S X. 2022. Ship detection in remote sensing image based on dense RFB and LSTM. *National Remote Sensing Bulletin*, 26(9): 1859-1871 (张涛, 杨小冈, 卢孝强, 卢瑞涛, 张胜修. 2022. Dense RFB 和 LSTM 遥感图像舰船目标检测. 遥感学报, 26(9): 1859-1871) [DOI: 10.11834/jrs.20211042]
- Zhang T Y, Zhang X R, Zhu P, Tang X, Li C, Jiao L C and Zhou H Y. 2022. Semantic attention and scale complementary network for instance segmentation in remote sensing images. *IEEE Transactions on Cybernetics*, 52(10): 10999-11013 [DOI: 10.1109/TCYB.2021.3096185]
- Zhao D P, Zhu C B, Qi J, Qi X H, Su Z H and Shi Z W. 2021. Synergistic attention for ship instance segmentation in SAR images. *Remote Sensing*, 13(21): 4384 [DOI: 10.3390/rs13214384]

Low-Cost Fine-Grained interpretation method of remote sensing images based on weakly supervised instance segmentation

CHEN Man, HUANG Yongjie, XU Lei, PAN Zhisong

College of Command and Control Engineering, Army Engineering University of PLA, Nanjing 210007, China

Abstract: Remote sensing image interpretation has essential application values in various fields, such as urban management, maritime monitoring, and resource planning. As an important and challenging task in remote sensing image interpretation, instance segmentation of remote sensing images can achieve target-level localization and pixel-level classification of objects of interest with fine granularity, making it a current research hotspot. However, most existing remote sensing image instance segmentation methods adopt the fully supervised paradigm and require expensive pixel-level labels. Moreover, remote sensing images often have issues such as mixed foreground and background and complex target contours, making segmentation challenging.

To overcome these challenges, we propose a prior information-driven system suitable for weakly supervised instance segmentation tasks in remote sensing images and a multiprior driven weakly supervised instance segmentation network (MPD-WSIS-Net) to address these challenges. The prior information of weakly supervised instance segmentation can be divided into task prior and image prior. The task prior mainly comes from the bounding box detection task, which is closely related to instance segmentation. Specifically, the study obtains prior information from three components: box - mask projection consistency constraint, pixel discrimination difficulty representation function, and center position prior constraint. The image prior comes from summarizing or excavating information about the image itself. In this study, we focus on the relationships between neighboring pixels in the image and the gradient information of targets. By integrating these constraints and the pixel discrimination difficulty representation function, we establish a complete prior information driving system to enable MPD-WSIS-Net effectively to perform instance segmentation tasks under weakly supervised conditions in remote sensing images.

MPD-WSIS-Net was compared with weakly supervised methods, hybrid supervised methods, and fully supervised methods on optical and SAR remote sensing image datasets. Compared with weakly supervised methods, MPD-WSIS-Net achieved improved segmentation results. Compared with hybrid supervised methods, the segmentation performance of MPD-WSIS-Net on optical and SAR remote sensing image datasets considerably surpasses that of Mask R-CNN and CondInst of 50% pixel-level annotations. It is also competitive compared with Mask R-CNN and CondInst under 75% pixel-level annotation conditions. Compared with fully supervised methods trained with pixel-level labels, MPD-WSIS-Net can achieve 89.3% of fully supervised Mask R-CNN's AP value on optical and 84.3% on SAR remote sensing image datasets. Furthermore, we have demonstrated the positive impact of each prior information component in MPD-WSIS-Net on instance segmentation performance in optical remote sensing images through ablation experiments.

This study constructs a prior information-driven system consisting of task and image priors through a detailed analysis of the prior information in weakly supervised instance segmentation tasks. The specification of prior information is achieved through the box-mask projection consistency constraint, pixel discrimination difficulty representation function, center position prior constraint, neighborhood visual consistency constraint, and gradient consistency constraint. This research can enable MPD-WSIS-Net to complete instance segmentation tasks without pixel-level annotations and provide high-performance and low-cost prescription for fine-grained interpretation of optical and SAR remote sensing images.

Key words: remote sensing image, instance segmentation, fine-grained interpretation, weakly supervised learning, priori information, driven system, target contour, annotation cost

Supported by National Natural Science Foundation of China (No. 62076251)