中图法分类号:TP391　　**文献标识码**: A　　**文章编号**: 1006-8961(XXXX)XX-0001-20

论文引用格式:Liu Yanfeng, Wang Di, Guo Haonan, Zhang Jing, Zhang Lefei, Zhang Liangpei. Vision foundation model-guided learning for label-efficient remote sensing salient object detection[J/OL]. Journal of Image and Graphics, XXXX:1-20. DOI: 10.11834/jig.260375.(刘焰锋, 王迪, 郭昊南, 张敬, 张乐飞, 张良培. 视觉基础模型驱动的低标注遥感显著目标检测[J/OL]. 中国图象图形学报,XXXX:1-20. DOI: 10.11834/jig.260375.)[DOI:10.11834/jig.260375]

视觉基础模型驱动的低标注遥感显著目标检测

刘焰锋[1,2],王迪[1,2],郭昊南[3,2],张敬[1,2],张乐飞[1*],张良培[3,4]

1. 武汉大学计算机学院,武汉　430072; 2. 北京中关村学院,北京　100094; 3. 武汉大学测绘遥感信息工程全国重点实验室,武汉　430079; 4. 河南省科学院空天信息研究所,郑州　450046

摘　要: **目的**　显著目标检测是遥感图像理解的重要基础任务。针对像素级标注成本高、视觉基础模型的通用先验难以有效适配复杂遥感场景的问题,提出一种视觉基础模型驱动的低标注遥感显著目标检测方法,以提高无标签样本利用效率和检测精度。**方法**　首先,本文构建了ST-SAM3渐进式自训练框架,该框架利用少量标注样本对SAM3进行轻量级微调,通过迭代提示推理获得无标签样本的细化伪掩码,并构建多维度的伪标签可靠性评估策略,从而筛选高质量无标签样本并生成可靠伪标签。进一步地,本文提出DistillMatch半监督蒸馏学习框架。该框架在尺度扰动和旋转扰动下约束学生预测与教师伪标签保持几何一致;同时设计统计信息蒸馏策略,从特征分布层面增强学生模型对遥感显著目标结构和语义响应的建模能力。**结果**　本文在ORSSD、EORSSD、ORSI-4199数据集上,以5%和10%的低标注比例进行实验。结果表明,本文方法在定量指标上显著优于现有半监督方法,验证了渐进式可靠伪标签构建、几何一致性学习与统计信息蒸馏的有效性。**结论**　该研究表明,借助SAM3基础模型的通用分割先验与提示推理能力,可在无需额外人工干预下提升无标签样本的利用效率,缓解深度网络模型对大规模像素级标注的依赖,为低标注空天遥感显著目标检测提供了一种有效可行的技术路径。代码仓库及数据地址:https://github.com/lyf0801/ST-SAM3.

关键词: 遥感显著目标检测;半监督学习;视觉基础模型;渐进式自训练;几何一致性学习;统计信息蒸馏

Vision foundation model-guided learning for label-efficient remote sensing salient object detection

Liu Yanfeng[1,2], Wang Di[1,2], Guo Haonan[3,2], Zhang Jing[1,2], Zhang Lefei[1*], Zhang Liangpei[3,4]

1. School of Computer Science, Wuhan University, Wuhan 430072, China; 2. Zhongguancun Academy, Beijing 100094, China; 3. State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China; 4. Aerospace Information Research Institute, Henan Academy of Sciences, Zhengzhou 450046, China

Abstract: **Objective** Salient object detection is a fundamental task in remote sensing image understanding. However, most existing approaches are developed under a fully supervised learning paradigm, where a large number of pixel-level annotations are required for model training. This strong dependence on dense annotations limits the scalability and practical deployment of current deep models in label-scarce remote sensing scenarios. Nevertheless, existing semi-supervised

收稿日期:2026-06-27;**修回日期**:2026-10-28

*通信作者:张乐飞 zhanglefei@whu.edu.cn

基金项目:国家自然科学基金(项目编号:62431020,U25A20403);河南省科技研发联合基金(项目编号:245200810004);北京中关村学院资助(项目编号:20240308)

Supported by: National Natural Science Foundation of China (Grant No. 62431020, U25A20403); Science and Technology Research and Development Joint Fund of Henan Province (Grant No. 245200810004); the Zhongguancun Academy (Grant No. 20240308)

© 中国图象图形学报版权所有

methods still suffer from noisy pseudo-labels, error accumulation, insufficient utilization of unlabeled samples, and a considerable performance gap compared with fully supervised methods. To address these issues, this paper proposes a foundation model-guided learning method for label-efficient remote sensing salient object detection, aiming to improve the utilization of unlabeled remote sensing images and enhance detection accuracy under limited annotation conditions. **Method** The proposed framework consists of two closely related models: ST-SAM3 and DistillMatch. First, we introduce the SAM3 visual foundation model into semi-supervised remote sensing salient object detection and construct a progressive self-training framework named ST-SAM3. Instead of directly using unreliable pseudo-labels generated from a model trained with only a few labeled samples, ST-SAM3 exploits the prompt-based mask prediction capability of SAM3 to refine and evaluate pseudo-labels in an automatic manner. Specifically, a lightweight adapter is fine-tuned on a small labeled subset to adapt SAM3 to remote sensing salient object detection. The fine-tuned model then generates initial pseudo-labels for unlabeled images. Based on these pseudo-labels, foreground and background prompts are automatically extracted from connected components and background regions. These prompts are fed back into SAM3 to perform iterative prompt-based mask refinement, thereby improving the spatial completeness and boundary quality of pseudo-labels. To further assess pseudo-label reliability, we design a quality assessment strategy that jointly evaluates geometric consistency among the original pseudo-label, the prompt-refined mask, and the mask predicted from the horizontally flipped image. High-quality unlabeled samples are selected according to this reliability score, and consensus pseudo-labels are generated through voting among the initial prediction, prompt-refined prediction, and augmentation-consistent prediction. This procedure progressively expands the labeled training set without requiring additional human intervention. On the basis of the reliable pseudo-labeled samples selected by ST-SAM3, we further propose DistillMatch to exploit the remaining unlabeled data more effectively. The reliable pseudo-labeled samples are treated as an extended labeled set, while the remaining unlabeled samples are used for consistency regularization. Different from conventional consistency learning that mainly relies on photometric perturbations, DistillMatch explicitly introduces geometric perturbations, including scale transformation and rotation transformation, to enforce geometric consistency between teacher supervision and student prediction. This design helps the student model learn robust salient object representations under spatial variations that frequently appear in remote sensing images. In addition, we design a high-order statistical distillation strategy to align multi-level feature statistics between teacher and student models within multiscale salient features, enabling the student model to inherit not only pixel-level pseudo-supervision but also higher-level feature distribution knowledge from the foundation-model-based teacher. **Result** Extensive experiments are conducted on three remote sensing salient object detection datasets, namely ORSSD, EORSSD, and ORSI-4199, under 5% and 10% labeled-data settings. Experimental results show that ST-SAM3 achieves superior performance across different datasets and annotation ratios, demonstrating the effectiveness of progressive self-training and reliable pseudo-label selection. Meanwhile, DistillMatch retains competitive detection accuracy with a substantially lighter model architecture. The proposed frameworks consistently outperform existing state-of-the-art semi-supervised methods. **Conclusion** This study demonstrates that visual foundation models can provide powerful priors for label-efficient remote sensing salient object detection. By introducing SAM3 into a progressive self-training framework, pseudo-labels are automatically refined and subsequently selected according to their reliability, without requiring additional manual annotation. By further combining reliable pseudo-label supervision with geometric consistency learning and high-order statistical information distillation, the proposed DistillMatch enables a lightweight student model to effectively learn from both labeled and unlabeled remote sensing images. Overall, ST-SAM3 and DistillMatch alleviate the dependence on large-scale pixel-level annotations and provide an effective and practical solution for semi-supervised remote sensing salient object detection.

Key words: remote sensing salient object detection; semi-supervised learning; visual foundation model; progressive self-training; geometric consistency learning; statistical information distillation

论文引用格式: Liu Y F, Wang D, Guo H N, Zhang J, Zhang L F and Zhang L P. 2026. Visual foundation model-guided learning for label-efficient remote sensing salient object detection. Journal of Image and Graphics, xx(x): xxx-xxx (刘焰锋, 王迪, 郭昊南, 张敬, 张乐飞, 张良培. 2026. 视觉基础模

© 中国图象图形学报版权所有

型驱动的低标注遥感显著目标检测. 中国图象图形学报, xx(x): xxx-xxx) [DOI: 10.11834/jig.260375]

0 引 言

遥感显著目标检测旨在从光学遥感图像中准确挖掘视觉显著或任务相关的前景目标,例如飞机、储罐、车辆、船只、建筑物以及其他人造目标等(Li等,2019)。与自然场景图像不同(兰猛 等,2024),遥感图像通常具有俯视成像、目标尺度剧烈变化、背景纹理复杂、目标分布稀疏及形态结构差异显著等特点(张文凯 等,2022)。作为光学遥感图像分析中的基础任务之一,遥感显著目标检测能够为后续高层语义理解提供精确的前景先验,在空天遥感大模型面向复杂场景的高效感知与细粒度理解中具有重要意义(孙显 等,2022;Yu等,2025;Liu等,2026a)。

近年来,以视觉基础模型为代表的大模型技术正在推动计算机视觉由任务专用模型向通用视觉表征与提示驱动推理转变。对于空天遥感领域而言,基础模型所具备的大规模预训练先验、开放场景泛化能力和交互式提示能力,为低标注遥感图像理解提供了新的可能。然而,现有遥感显著目标检测方法仍主要采用全监督学习范式,即在完整训练集上利用人工标注的像素级掩码训练深度网络。该范式虽然能够提供直接、稳定的监督信号,使模型学习输入遥感图像到输出显著目标掩码之间的精确映射关系,但也带来了高昂的标注成本,限制了方法在真实遥感应用中的可扩展性和快速迁移能力。

为缓解这一问题,半监督学习提供了一种更具成本效益的替代方案:模型仅依赖少量有标注样本提供可靠监督,同时充分利用大量无标注样本中的数据分布、结构一致性和潜在语义信息,从而在显著降低标注需求的情况下逼近全监督模型的性能。这一范式与遥感应用场景具有较强契合性:一方面,空天遥感平台能够持续获取大规模影像数据,无标注样本远多于精细标注样本;另一方面,实际任务往往要求模型在有限标注条件下快速适配新的区域、目标和成像条件。与此同时,视觉基础模型依托大规模预训练形成的通用视觉先验和提示推理能力,为低标注遥感密集预测提供了新的可借鉴思路。如何将基础模型的通用先验与半监督学习机制结合起来,构建适用于复杂遥感场景的可靠监督信号,是当前低标注遥感视觉理解中值得研究的问题。

现有半监督密集预测方法主要可以分为基于伪标签自训练的方法和基于一致性正则化的方法(Jin等,2024;Fu等,2026)。伪标签自训练方法通常先利用少量标注样本训练模型,再为无标注样本生成伪标签,并将高置信的伪标签样本重新加入训练集。ST++等方法(Yang等,2022)通过重新训练和可靠样本选择提升了自训练的稳定性,使无标注数据能够转化为显式监督信号。这类方法的优势在于训练目标直接,能够让模型学习图像到掩码之间的显式映射;但其核心瓶颈在于伪标签质量极度依赖模型本身,一旦伪标签中存在目标缺失、边界偏移或背景误检,后续训练容易放大错误监督并造成噪声累积。另一类代表性方法是基于一致性正则化的半监督学习,例如FixMatch(Sohn等,2020)通过弱增强输入图像的预测输出生成高置信伪标签,并约束强增强视图下的预测与其保持一致;UniMatch(Yang等,2023)和UniMatchV2(Yang等,2025)进一步从弱到强一致性和多视图扰动等方面提升了半监督语义分割性能。这类方法能够有效利用大量无标注样本,并通过扰动一致性提升模型的鲁棒性;然而,其监督信号主要来源于不同增强视图之间的预测一致性,更多强调模型输出在扰动下的稳定性,而非直接提供高质量的像素级前景掩码监督。对于遥感显著目标检测而言,仅依靠强弱增强之间的一致性约束难以充分建模小目标、复杂边界及前景区域结构。

针对显著目标检测任务,近期研究开始探索更具任务适应性的半监督学习框架。例如,CATNet(Chen等,2025a)通过一致性引导的自适应交替训练,促进有标注监督与无标注学习之间的互补;TripleNet(Chen等,2025b)则联合互补特征与伪标签信息,以提升半监督显著目标检测的表征能力。这些工作表明,半监督显著目标检测不能仅依赖单一的伪标签或一致性约束,而应协同利用可靠监督、特征表达与无标注样本信息。近年来,以Segment Anything Model(SAM)为代表的视觉基础模型为半监督学习提供了新的技术路径。KnowSAM(Huang等,2025a)和SSMIS(Zhao等,2025)等SAM驱动的半监督医学图像分割方法表明,视觉基础模型依托大规模预训练形成的通用分割先验与提示推理能力,能够在标注稀缺场景下为无标注样本提供更具语义辨识能力的伪标签,这给半监督遥感显著目标检测

© 中国图象图形学报版权所有

挖掘无标签数据价值提供了新的思路。然而，将其直接迁移至遥感显著目标检测仍面临两方面挑战：其一，遥感场景中的显著目标常具有尺度变化大、背景纹理复杂、类别无关及前背景对比度低等特点，基础模型的通用分割结果与“突出视觉显著前景”的任务目标并不一致，易产生目标缺失、边界偏移或背景误检；若将未经甄别的预测直接用于训练，噪声将随迭代过程累积。其二，基础模型通常具有庞大的参数规模和较高计算开销，难以直接部署于资源受限的遥感应用平台。因此，如何借助基础模型的通用分割先验构建可靠伪标签，并将其有效迁移至轻量化检测网络，是半监督遥感显著目标检测亟待解决的科学问题之一。

为解决上述问题，本文提出一种面向低标注遥感显著目标检测的基础模型引导学习方法，整体由Self-Training SAM3（ST-SAM3）和DistillMatch两个阶段组成。本文的核心思想是同时吸收伪标签自训练和一致性学习的优势：一方面，通过基础模型生成并筛选高质量伪标签，使大部分无标注样本转化为可靠的显式监督信号；另一方面，将剩余无标注样本用于一致性学习，使模型能够从更大规模的数据分布中学习鲁棒表示。具体而言，本文首先引入SAM3视觉基础模型，构建ST-SAM3渐进式自训练框架。该框架利用少量标注样本对SAM3进行轻量适配器微调（Chen等，2025c），以增强其对遥感显著目标区域的响应能力。随后，基于SAM3的提示推理能力对无标注样本得到的初始伪标签进行自动细化。为了减少人工交互，本文从伪标签中自动提取前景提示和背景提示，并通过迭代式提示推理获得更精细的候选掩码。同时，结合多视图一致性响应，本文构建多维度伪标签可靠性评价策略，从无标注样本中筛选出高质量样本并生成可靠伪标签。在ST-SAM3的基础上，本文进一步提出DistillMatch框架，以更充分地利用筛选后和未筛选的无标注样本。DistillMatch将ST-SAM3筛选得到的可靠伪标签样本作为扩展标注集，与原始少量真实标注样本共同组成监督训练分支，从而为轻量级学生网络模型提供直接的图像—掩码映射监督；对于剩余无标签样本，本文采用ST-SAM3作为教师模型、DINOv3-Small作为学生模型，在尺度扰动和旋转扰动下约束学生预测与教师伪标签保持几何一致。进一步地，本文提出了一种高效的高阶统计信息蒸馏策略，对齐ST-SAM3教师模型与DINOv3-Small学生模型多层表征的一阶至四阶统计矩。这一策略使学生模型不仅关注显著目标像素级预测结果的一致性，还从特征分布层面对教师模型中的前景响应、结构模式和语义统计信息进行迁移，更充分地学习遥感显著目标的区域结构和语义

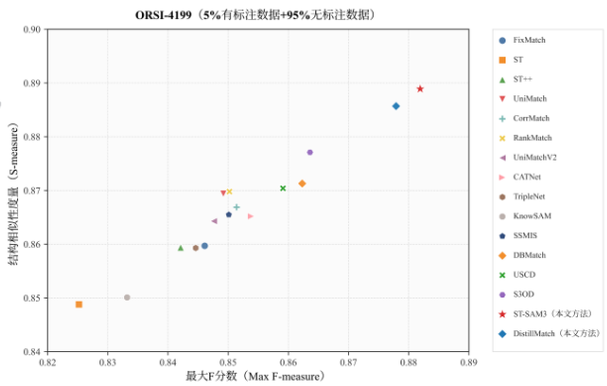


图1　不同方法在ORSI-4199数据集上的性能散点图

Fig. 1　Comparison with state-of-the-art semi-supervised methods under 5% labeled data on the ORSI-4199 dataset

表达。通过可靠伪标签监督、几何一致性学习和统计信息蒸馏的协同作用，本文所提出的ST-SAM3和DistillMatch方法旨在缓解半监督遥感显著目标检测中的伪标签噪声累积、无标注样本利用不足以及模型性能难以逼近全监督方法等问题。如图1所示，在ORSI-4199数据集5%的有标注数据设定下，本文方法相较于现有半监督方法表现出更优的性能分布，说明所提出方法能够在低标注成本条件下有效提升遥感显著目标检测性能，大大突破样本标注稀缺对模型检测能力的限制瓶颈。总而言之，本文的主要贡献可以概括如下：

1）提出了一种面向半监督遥感显著目标检测的ST-SAM3渐进式自训练方法。该方法利用少量标注样本对SAM3进行适配微调，并通过自适应提示驱动的伪标签细化和多维一致性约束的可靠性评估对无标注样本的伪标签进行细化、筛选和融合生成，从而获得更可靠的扩展伪标注集。

2）提出了一种基于几何增强和知识蒸馏的一致性学习框架DistillMatch。该框架将ST-SAM3筛选得到的可靠伪标签样本用于直接监督训练，同时将剩余无标签样本用于尺度扰动和旋转扰动下的教师—学生一致性约束，实现显式伪标签监督与无标注

© 中国图象图形学报版权所有

样本一致性学习的互补融合。

3)提出了一种简单高效的高阶统计信息蒸馏策略。考虑到教师模型的判别知识不仅体现在最终预测结果中,还蕴含于多层中间特征的统计分布内,该策略以SAM3教师模型和DINOv3-Small学生模型的多层中间表征为知识载体,通过对齐高阶统计矩,从特征分布层面增强学生模型对遥感显著目标结构、边界和语义响应的建模能力。

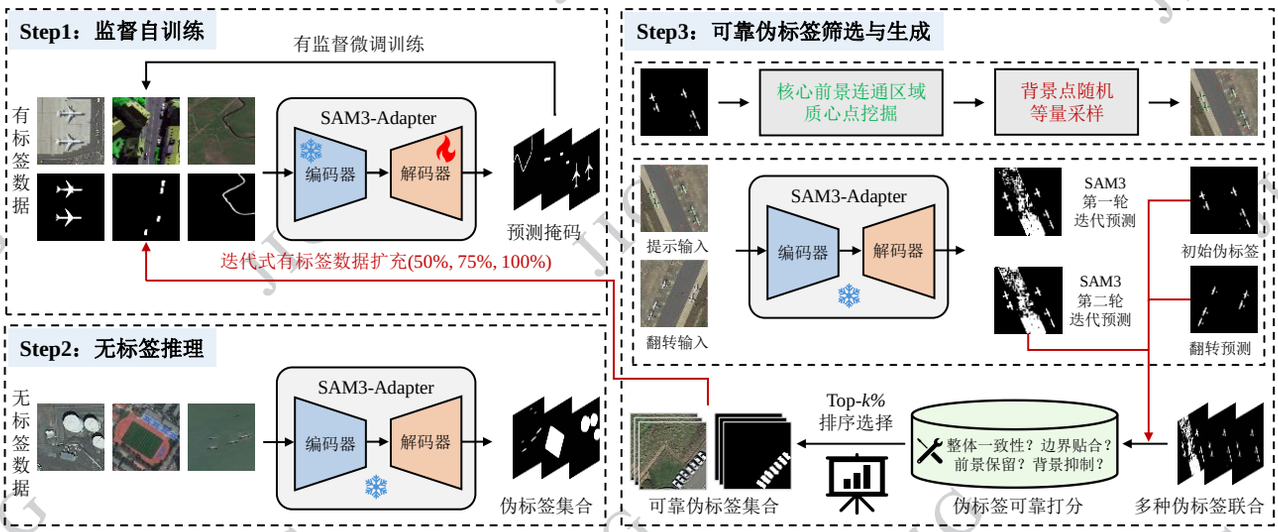


图2　本文所提出的ST-SAM3渐进式自训练方法框架图

Fig. 2　Framework overview of the proposed ST-SAM3 progressive self-training approach

1　本文方法

1.1　ST-SAM3方法介绍

针对遥感显著目标检测中像素级标注成本高、少量标注样本难以充分覆盖复杂地物外观的问题,

本文提出一种基于SAM3的半监督渐进式自训练方法ST-SAM3。该方法整体流程如图2所示,方法首先利用少量人工标注样本对视觉基础模型SAM3进行有监督微调,获得适应遥感显著目标检测任务的优化模型;随后使用该模型对无标注遥感图像生成初始伪标签;最后通过启发式可靠伪标签筛选策略,从无标注样本中选出质量较高的伪标签并加入训练集,重复执行“训练—推理—筛选—再训练”的过程,从而逐步扩充有效监督信息。

设有标注训练集为$\mathcal{D}_l=\{(\boldsymbol{I}_i,\boldsymbol{Y}_i)\}_{i=1}^{N_l}$,无标注训练集为$\mathcal{D}_u=\{\boldsymbol{I}_j\}_{j=1}^{N_u}$,其中,$\boldsymbol{I}\in\mathbf{R}^{H\times W\times 3}$表示输入遥感图像,$\boldsymbol{Y}\in\{0,1\}^{H\times W}$表示显著目标二值标注,$H$和$W$分别表示图像高度和宽度,$N_l$与$N_u$分别表示有标注样本和无标注样本的数量。ST-SAM3的目标是学习一个显著目标预测模型$f_\theta(\cdot)$,输出显著目标检测图$\boldsymbol{P}=f_\theta(\boldsymbol{I})$。本文采用简单的Adapter结构(Chen等,2025c)作为视觉基础模型的微调方法,用于对SAM3进行高效微调。具体而言,SAM3-Adapter首先对输入图像进行傅里叶变换,并通过高通滤波提取高频结构信息;随后利用轻量级MLP将频域提示特征映射到与视觉主干特征一致的维度,并注入到Transformer特征中。该过程可表示为:

$$\boldsymbol{F}'_{s,t}=\boldsymbol{F}_{s,t}+\phi_s\left(\psi_{s,t}\left(\boldsymbol{Q}_s+\eta_s(\boldsymbol{F}_{s,t})\right)\right)\tag{1}$$

式中,$\boldsymbol{F}_{s,t}$表示第s个阶段第t个Transformer块的原始特征,$\boldsymbol{F}'_{s,t}$表示注入Adapter后的增强特征,$\boldsymbol{Q}_s$表示由频域高通分支提取的提示特征,$\eta_s(\cdot)$表示特征降维映射,$\psi_{s,t}(\cdot)$表示轻量级MLP,$\phi_s(\cdot)$表示共享线性映射。由于该Adapter仅引入少量可学习参数,因此能够在冻结绝大部分SAM3图像编码器参数的情况下完成遥感域适配,降低少量样本训练时的过拟合风险。需要强调的是,SAM3-Adapter在本文中作为视觉基础模型微调使用,ST-SAM3的主要贡献在于后续的多轮自训练与可靠伪标签筛选策略。

1.1.1　多轮渐进式自训练

如图2左上部分所示,ST-SAM3首先在少量真实标注样本上进行初始有监督微调。训练过程中,模型预测概率图$\boldsymbol{P}_i=f_\theta(\boldsymbol{I}_i)$,并通过二元交叉熵损失和区域IoU损失共同优化,监督损失定义为

© 中国图象图形学报版权所有

$$\mathcal{L}_{sup} = \mathcal{L}_{bce}(\boldsymbol{P},\boldsymbol{Y}) + \mathcal{L}_{iou}(\boldsymbol{P},\boldsymbol{Y}) \tag{2}$$

式中，$\mathcal{L}_{sup}$表示有监督训练损失，$\mathcal{L}_{bce}$为二分类交叉熵损失，用于约束像素级前景/背景分类；$\mathcal{L}_{iou}$为加权交并比损失，用于增强预测区域与真实显著区域之间的整体重叠一致性。相比仅使用像素级损失，结构化区域约束更适合遥感显著目标检测中目标大小变化明显、边界形态复杂的情况（Liu等，2023a；Liu等，2023b；Wang等，2022；Liu等，2025）。

完成初始训练后，如图2左下部分所示，模型对无标注图像进行推理，生成初始伪标签。对于无标注图像$\boldsymbol{I}_j$，其第k轮伪标签定义为

$$\hat{\boldsymbol{Y}}_j^k = I\left(f_{\theta^k}(\boldsymbol{I}_j) > \tau\right) \tag{3}$$

式中，$\hat{\boldsymbol{Y}}_j^k$表示第k轮模型生成的二值伪标签，θ^k表示第k轮模型参数，τ为二值化阈值，本文取$\tau = 0.5$，$I(\cdot)$为指示函数。由于初始模型仅由少量标注样本训练得到，直接使用全部伪标签会引入大量噪声。因此，本文摒弃了将所有无标注伪标签直接加入训练集的简单思路，而采用在每轮训练后通过可靠性评估筛选高质量伪标签的策略。筛选后的可靠伪标签与原始人工标注共同构成下一轮训练集：

$$\mathcal{D}_l^{k+1} = \mathcal{D}_l \cup \{(\boldsymbol{I}_j,\hat{\boldsymbol{Y}}_j^k) \mid j \in \mathcal{R}^k\} \tag{4}$$

式中，$\mathcal{D}_l^{k+1}$表示第$k+1$轮扩展训练集，$\hat{\boldsymbol{Y}}_j^k$表示经过筛选、细化和融合后的可靠伪标签，$\mathcal{R}^k$表示第k轮被选中的可靠无标注样本集合。该过程对应图2中的红色循环路径：模型先由少量标注样本获得初始伪标签，再通过可靠样本筛选扩展监督集合，随后重新训练并继续生成更高质量的伪标签。通过这种渐进式扩展方式，ST-SAM3能够避免一次性引入低质量伪标签造成的确认偏差，同时逐轮次提高模型对遥感显著目标的适应能力。

1.1.2 自适应提示驱动的伪标签细化

初始模型在无标注图像上产生的伪标签通常包含一定的空间结构信息，但其边界完整性和目标定位精度仍受到少量标注监督的限制。若仅将初始预测直接视为伪监督信号，模型难以纠正其中的局部漏检、孔洞和背景干扰。鉴于SAM3具备利用点提示引导分割区域的能力，本文并未直接将初始伪标签用作训练标签，而是将其转化为包含目标位置与背景约束的自适应提示，进而驱动ST-SAM3对初始预测结果进行结构化修正与细化。

图2右上部分展示了自适应提示的生成过程。与为所有图像分配固定数量提示点的方式不同，本文从初始伪标签的空间结构中动态生成提示，使提示数量、位置及正负样本比例能够随图像内容变化。对于无标注图像$\boldsymbol{I}_j$的初始二值伪标签为$\hat{\boldsymbol{Y}}_j^k$，首先对其进行八邻域连通域分析，得到前景连通域集合

$$\mathcal{C}_j = \left\{ \mathcal{C}_{j,n} \mid \left|\mathcal{C}_{j,n}\right| \geqslant A_{\min} \right\}_{n=1}^{N_j} \tag{5}$$

式中，$\mathcal{C}_{j,n}$表示第n个前景连通域，$\left|\mathcal{C}_{j,n}\right|$表示该区域的像素面积，$A_{\min} = 100$表示最小有效区域面积，$N_j$表示图像$\boldsymbol{I}_j$中保留的有效前景区域数。对于每一个有效连通域，本文将其几何质心作为前景提示点，其中第n个前景点的归一化坐标可表示为

$$p_{j,n}^{+} = \left(\frac{1}{W|\mathcal{C}_{j,n}|} \sum_{(x,y)\in\mathcal{C}_{j,n}} x,\ \frac{1}{H|\mathcal{C}_{j,n}|} \sum_{(x,y)\in\mathcal{C}_{j,n}} y \right) \tag{6}$$

式中，$p_{j,n}^{+}$表示第n个前景提示点，(x,y)表示连通域中的像素坐标，H和W分别表示图像高度和宽度。质心位于区域内部且对局部边界噪声不敏感，因此能够用一个稳定点表示一个候选显著目标。由此，包含多个独立显著目标的图像将自动产生多个前景点；仅包含单个显著目标的图像则只产生一个前景点，这正是“自适应”的含义。

为约束模型不将周围复杂纹理误分为前景，本文从初始伪标签的背景区域中随机采样与前景点等数量的背景点，记为$\{p_{j,n}^{-}\}_{n=1}^{N_j}$。最终，图像$\boldsymbol{I}_j$的点提示集合可表示为

$$\mathcal{P}_j = \left\{(p_{j,n}^{+},1)\right\}_{n=1}^{N_j} \cup \left\{(p_{j,n}^{-},0)\right\}_{n=1}^{N_j} \tag{7}$$

式中，标签1和0分别表示前景点和背景点。该提示集合同时编码了“应被保留的候选目标”与“应被排除的背景区域”，使初始预测从静态掩码转化为可被SAM3先验修正的交互式条件。

在获得提示集合后，本文将其输入ST-SAM3的提示编码器和掩码解码器，得到细化掩码；随后再从细化结果中重新提取提示并进行第二次细化。该过程可概括为

$$\boldsymbol{Y}_j^{k,t} = I\left[g_{\theta^k}\left(\boldsymbol{I}_j,\mathrm{P}_j^{t-1}\right) > \tau \right],\quad t = 1,2 \tag{8}$$

式中，$g_{\theta^k}(\cdot)$表示第k轮ST-SAM3的提示式分割过程，$\mathcal{P}_j^{t-1}$表示由第$t-1$次掩码生成的提示，$\boldsymbol{Y}_j^{k,t}$表示第t次细化结果。如图2所示，本文进行两次提示迭代，在控制计算开销的同时获得更稳定的细化结果。

© 中国图象图形学报版权所有

1.1.3　多维一致性约束的可靠性评估

经过自适应提示细化后，每个无标注样本均可获得一个候选细化掩码。然而，细化结果在视觉上更平滑并不意味着其一定可靠：对于初始预测严重偏离真实目标的样本，SAM3也可能在错误提示的引导下生成结构完整但语义错误的区域。因此，本文进一步提出多维一致性约束的可靠性评估策略，其不将伪标签质量简单等同于预测置信度，而是判断初始预测、提示细化结果和测试时几何扰动预测之间是否具有稳定的结构一致性。

如图2右侧所示，本文以初始伪标签作为参照，分别考察最终SAM3细化掩码和水平翻转的测试掩码。若一个初始伪标签能够在提示驱动的细化后保持主体区域与边缘结构稳定，同时在翻转扰动下仍输出一致的显著区域，则说明该预测既符合SAM3的提示式分割先验，也具有较好的几何鲁棒性；反之，若细化结果出现明显区域漂移、边缘重构或前景大面积丢失，则该样本应被视为不可靠。基于此，本文从整体一致性、边界贴合、前景保留和背景抑制四个互补维度构建伪标签质量评估方案。

设$\boldsymbol{A}$为初始伪标签，$\boldsymbol{B}$为待评估的候选掩码，Ω为图像像素集合。本文依次取第二次提示细化后的掩码和水平翻转后恢复至原坐标系的掩码作为$\boldsymbol{B}$。首先，采用交并比衡量候选掩码与初始伪标签在目标区域上的整体一致性：

$$S_{\mathrm{iou}}(\boldsymbol{A},\boldsymbol{B})=\frac{|\boldsymbol{A}\cap\boldsymbol{B}|}{|\boldsymbol{A}\cup\boldsymbol{B}|+\varepsilon} \tag{9}$$

式中，$|\cdot|$表示掩码中前景像素的数量，ε为防止分母为零的极小常数。较高的S_{iou}表明候选掩码与初始伪标签在显著区域范围上具有较强一致性。但IoU仅反映区域重叠，难以识别区域面积相近但边界明显偏移的情况。为此，本文进一步通过Laplacian边缘响应评价两张掩码的边界贴合程度：

$$S_{\mathrm{edge}}(\boldsymbol{A},\boldsymbol{B})=1-clip\left(\frac{\left\|\Delta(\boldsymbol{A})-\Delta(\boldsymbol{B})\right\|_1}{255|\Omega|},0,1\right) \tag{10}$$

式中，$\Delta(\cdot)$表示Laplacian边缘算子，$\|\cdot\|_1$表示逐像素绝对差之和，$clip(\cdot,0,1)$表示将数值截断到[0,1]区间。由于二值掩码以0和255编码，$255|\Omega|$用于对边缘差异进行归一化。较高的S_{edge}表示候选掩码与初始预测具有更一致的目标轮廓，可抑制细化过程中出现的边界漂移。

除整体和边缘一致性外，可靠的候选掩码还应满足两个方向性的约束：一方面，不能大面积丢失初始预测中的前景区域；另一方面，也不应在初始背景中产生无依据的前景扩张。为此，本文分别定义并计算前景缺失比例和背景幻觉比例：

$$D_{\mathrm{miss}}(\boldsymbol{A},\boldsymbol{B})=\frac{\left|\left\{x\in\Omega\,\middle|\,\boldsymbol{A}(x)=1,\boldsymbol{B}(x)=0\right\}\right|}{|\boldsymbol{A}|+\varepsilon} \tag{11}$$

$$D_{\mathrm{hallu}}(\boldsymbol{A},\boldsymbol{B})=\frac{\left|\left\{x\in\Omega\,\middle|\,\boldsymbol{A}(x)=0,\boldsymbol{B}(x)=1\right\}\right|}{|\boldsymbol{A}|+\varepsilon} \tag{12}$$

式中，x表示像素位置，D_{miss}表示候选掩码在初始背景中的前景缺失比例。相应地，$1-D_{\mathrm{miss}}$表示前景保留分数，使用$1-D_{\mathrm{hallu}}$表示背景抑制分数。

对于任意候选掩码，本文将整体一致性分数S_{iou}、边缘贴合分数S_{edge}、前景保留分数$1-D_{\mathrm{miss}}$和背景抑制分数$1-D_{\mathrm{hallu}}$的算术平均值定义为该候选掩码的启发式质量分数。上述四项约束分别从区域范围、轮廓位置、前景完整性和背景纯净性刻画伪标签质量，避免仅依赖单一置信度或IoU指标所导致的误判。随后，本文分别计算初始伪标签与第二次提示细化掩码之间的质量分数，以及初始伪标签与水平翻转掩码之间的质量分数，并将两者相加作为该无标注样本的可靠性得分。前者用于验证伪标签对SAM3提示式分割先验的响应稳定性，后者用于验证其在几何变换下的预测稳定性。根据可靠性得分对全部无标注样本进行降序排序，保留Top-k%样本：具体而言，第一轮迭代训练取50%，第二轮迭代训练取75%，第三轮迭代训练取100%。

如图2右下部分所示，对于被选中的可靠样本，本文对初始预测、第二次提示细化结果和水平翻转结果实施多数投票：当至少两路预测均将某一像素判定为前景时，才将其保留为最终伪标签。最后，本文对融合后的掩码进行微小孔洞填充，并排除面积微小的孤立连通域。通过提示式细化一致性、几何扰动稳定性以及多维结构一致性约束的联合评估，ST-SAM3能够将结构更稳定、噪声更低的伪标签纳入下一轮训练集，从而降低错误伪标签监督在多轮自训练过程中的累积误导风险。

1.2　DistillMatch方法介绍

ST-SAM3的参数量和计算开销分别达到458.18 M和631.28 GMACs，限制了模型在资源受限遥感平台上的部署。为降低推理成本，本文选用仅含26.91 M参数和14.19 GMACs计算量的

© 中国图象图形学报版权所有

DINOv3-Small(Siméoni 等,2025)作为学生网络,其参数量仅为教师模型的5.87%,计算量仅为教师模型的2.25%,分别降低约94.13%和97.75%。然而,直接以 DINOv3-Small 替代 ST-SAM3 会造成严重的表征损失,尤其是在复杂背景、弱小目标和边界模糊的遥感场景中。为此,本文提出了一种蒸馏学习框架 DistillMatch,在训练阶段以冻结的 ST-SAM3 作为教师模型,通过几何一致性学习和统计信息蒸馏将其显著目标判别能力迁移至轻量级 DINOv3-Small 网络;在推理阶段仅保留学生模型,从而实现高性能与低部署成本之间的平衡。

1.2.1 几何一致性学习

为充分利用剩余无标注样本,本文构建教师—学生协同的几何一致性学习机制。如图3所示,所提出的 DistillMatch 由可靠标注监督分支、无标注几何一致性分支和统计信息蒸馏分支组成。左侧分支使用 ST-SAM3 自训练阶段获得的扩展标注集训练学生模型,其中包含少量真实标注及经可靠性筛选后的伪标签,以保证学生具备基本的显著目标定位能力。右侧分支以无标注图像为输入,融合 ST-SAM3 教师模型与指数滑动平均(Exponential Moving Average,EMA)学生模型的弱增强预测生成高置信度伪标签,并通过缩放和旋转扰动约束学生模型的几何鲁棒性。右侧分支将冻结 SAM3 教师模型的多尺度高阶统计表征迁移至学生模型,使学生不仅学习教师的一致性预测结果,还学习其对显著区域与复杂背景的特征分布级判别模式。

设冻结的 ST-SAM3 教师模型为 $g_\phi(\cdot)$,在线 DINOv3-Small 学生模型为 $f_\theta(\cdot)$,其 EMA 模型为 $f_{\bar{\theta}}(\cdot)$。其中,ϕ 为固定教师参数,θ 为待优化学生参数,$\bar{\theta}$ 为 EMA 学生参数。EMA 模型由在线学生参数持续更新,用于平滑随机优化带来的参数波动,并与教师模型共同提供更稳定的无标注伪监督。

对于任一无标注图像 $\boldsymbol{I}_u$,本文首先构造弱增强视图 $\boldsymbol{I}_u^w$,并分别输入 ST-SAM3 教师模型和 EMA 学生模型。设二者输出的显著性概率图为 $g_\phi(\boldsymbol{I}_u^w)$ 和 $f_{\bar{\theta}}(\boldsymbol{I}_u^w)$,则通过双模型融合生成伪标签:

$$\overline{\boldsymbol{P}}_u=\frac{g_\phi(\boldsymbol{I}_u^w)+f_{\bar{\theta}}(\boldsymbol{I}_u^w)}{2},\quad \hat{\boldsymbol{Y}}_u=I\left(\overline{\boldsymbol{P}}_u>0.5\right),\quad \boldsymbol{W}_u=I\left[max\left(\overline{\boldsymbol{P}}_u,1-\overline{\boldsymbol{P}}_u\right)\geqslant\delta\right] \tag{13}$$

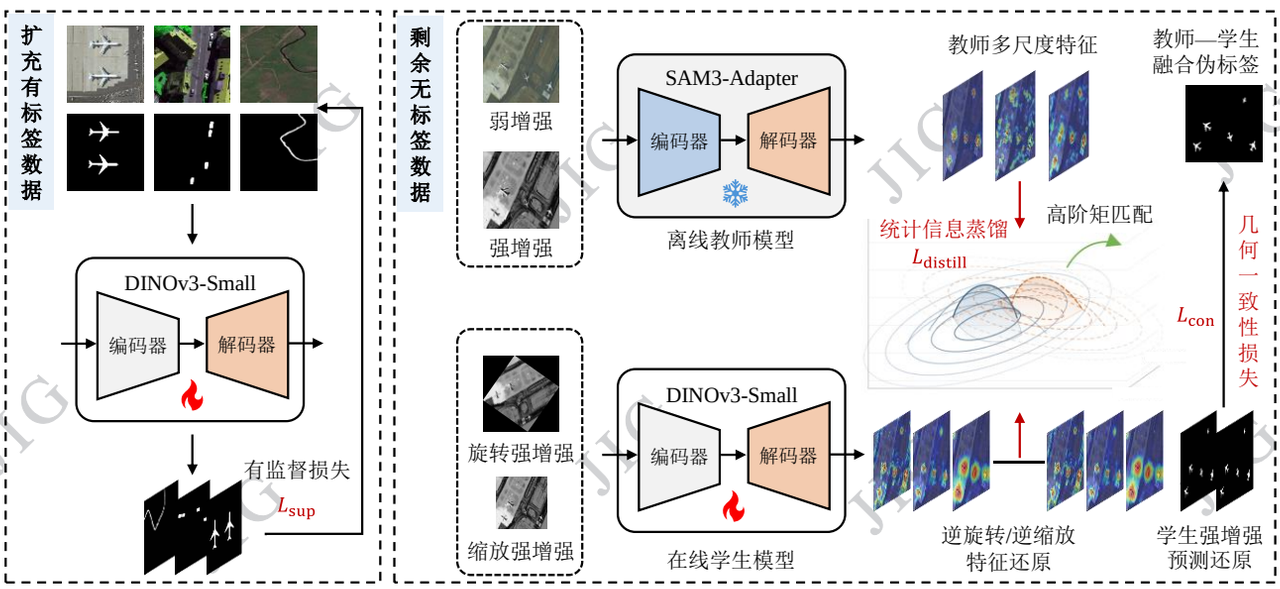


图3 本文所提出的基于几何增强的一致性学习方法 DistillMatch 框架图

Fig. 3 Framework diagram of the proposed geometry-enhanced consistency learning method DistillMatch

式中,$\overline{\boldsymbol{P}}_u$ 表示融合后的伪标签概率图,$\hat{\boldsymbol{Y}}_u$ 表示二值伪标签,$\boldsymbol{W}_u$ 表示像素级置信度掩码,δ 为置信度阈值,本文取0.95,$I(\cdot)$ 为指示函数。只有置信度满足阈值的像素参与无标注损失计算,从而减弱不可靠预测的影响(Yang 等,2023;Yang 等,2025)。

在此基础上,本文从同一无标注图像构造两路强增强视图。两路视图均依次施加颜色扰动、随机灰度化和高斯模糊;第一路进一步施加随机缩放,缩放比例从{0.25,0.5,2.0}中随机选择;第二路进一步施加随机旋转,旋转角度在[-180,180]内随机采

© 中国图象图形学报版权所有

样。教师模型输入未施加几何扰动的强增强图像，学生模型则分别接收强增强缩放视图和旋转视图。

设 $\boldsymbol{Z}_u^{\rho}$ 和 $\boldsymbol{Z}_u^{\alpha}$ 分别表示学生模型在缩放视图和旋转视图上输出的预测概率，$\mathcal{A}_{\alpha}(\cdot)$ 表示将旋转预测逆旋转并恢复至原始图像尺度的空间对齐操作。对于缩放分支，伪标签与置信度掩码采用最近邻插值同步调整至学生预测的空间尺度；为简化表示，调整后的伪标签和置信度掩码仍分别记为 $\hat{\boldsymbol{Y}}_u$ 与 $\boldsymbol{W}_u$。对于旋转分支，旋转操作保持原始图像画布大小，并以0填充超出原图范围的边缘区域。则几何一致性损失定义为

$$\mathcal{L}_{\text{con}}=\frac{1}{2}\left[\frac{1}{|\Omega_{\rho}|}\sum_{x\in\Omega_{\rho}}\boldsymbol{W}_u(x)\ell_{\text{bce}}(\boldsymbol{Z}_u^{\rho}(x),\hat{\boldsymbol{Y}}_u(x))+\frac{1}{|\Omega|}\sum_{x\in\Omega}\boldsymbol{W}_u(x)\ell_{\text{bce}}(\mathcal{A}_{\alpha}(\boldsymbol{Z}_u^{\alpha})(x),\hat{\boldsymbol{Y}}_u(x))\right] \tag{14}$$

式中，Ω_{ρ} 和 Ω 分别表示缩放视图和原始视图的像素集合，x 表示像素位置，$\ell_{\text{bce}}[\cdot,\cdot]$ 表示以预测输出和二值标签为输入的像素级二元交叉熵。第一项通过缩放后的高置信伪标签监督学生学习尺度变化下的目标结构；第二项通过将旋转预测恢复到原始坐标系，监督学生保持旋转前后的空间语义一致性。二者共同构成的 $\mathcal{L}_{\text{con}}$ 使学生模型不再依赖固定尺度或固定方向下的局部外观，而是学习对遥感目标的尺度与方向变化更稳定的显著特征响应。

1.2.2 统计信息蒸馏学习

几何一致性学习主要在输出层约束学生模型，但轻量DINOv3-Small与ST-SAM3在网络深度、特征提取机制和上下文建模能力上存在显著差异，仅依赖伪标签监督难以充分迁移教师模型的中间语义知识。尤其对于复杂背景下的弱显著目标，教师模型通常能够产生更稳定的区域响应与背景抑制特征，而学生模型的局部特征容易受到纹理干扰。为此，本文进一步从中间特征分布视角对教师和学生进行高阶语义知识蒸馏。如图3右侧所示，教师模型从经强增强、未施加几何扰动的图像中提取多尺度特征；学生模型则从对应的缩放或旋转强增强图像中提取编码器特征，并通过线性映射将其通道数对齐至ST-SAM3的通道维度，随后进行匹配。其中，对于旋转分支，学生模型的特征在匹配前需逆旋转回教师特征所在的坐标系；而对于缩放分支，则可在当前尺度的特征图上直接计算匹配损失。

设任一对齐后的特征图为 $\boldsymbol{F}\in\mathbf{R}^{C\times H\times W}$，其中 C、H 和 W 分别表示特征图通道数、高度和宽度。本文沿空间维度计算每个通道的一至四阶统计矩，以特征分布的矩统计量作为高阶语义表征知识：

$$\mathcal{M}_{r,c}(\boldsymbol{F})=\begin{cases}\dfrac{1}{HW}\displaystyle\sum_{h=1}^{H}\sum_{w=1}^{W}\boldsymbol{F}_{c,h,w}, & r=1,\\ \dfrac{1}{HW}\displaystyle\sum_{h=1}^{H}\sum_{w=1}^{W}\left[\boldsymbol{F}_{c,h,w}-\mathcal{M}_{1,c}(\boldsymbol{F})\right]^{r}, & r=2,3,4\end{cases} \tag{15}$$

式中，$\boldsymbol{F}_{c,h,w}$ 表示第 c 个通道在位置 (h,w) 的特征响应，$\mathcal{M}_{r,c}(\cdot)$ 表示第 c 个通道的第 r 阶统计矩。特别地，$\mathcal{M}_{1,c}(\cdot)$ 为第一阶矩描述通道平均激活强度。第二阶矩刻画特征响应的离散程度，第三阶矩描述响应分布的偏斜性，第四阶矩反映显著响应的尖峰与长尾特征。相较于直接进行逐像素特征匹配，高阶矩约束能够在保留通道级语义分布的同时，对异构网络间局部响应位置偏差具有更好的容忍性。

设 $\boldsymbol{F}_s^l$ 和 $\boldsymbol{F}_t^l$ 分别表示第 l 个匹配层的学生特征和教师特征，则单路视图下的统计蒸馏损失为

$$\mathcal{L}_{\text{distill}}=\sum_{r=1}^{4}\left\|\mathcal{M}_r(\boldsymbol{F}_s^l)-\mathcal{M}_r(\boldsymbol{F}_t^l)\right\|_2^2 \tag{16}$$

式中，l 表示参与蒸馏的特征层，$\|\cdot\|_2^2$ 表示均方误差。该设计使学生模型能够继承教师模型在不同尺度和方向条件下的分布级语义知识，而不要求二者具有完全一致的网络结构或局部特征响应。

1.2.3 联合优化损失函数

对于所提出的DistillMatch模型的三个训练分支承担不同且互补的作用。扩展有标注集合中的可靠标注用于建立学生模型的基本显著目标定位能力，能够为轻量学生模型提供较充分且相对稳定的显著性监督锚点；剩余无标注样本的低质量伪标签不会被直接固化为有监督目标，但其蕴含的几何一致性约束能提升学生模型对目标尺度、方向及外观扰动的鲁棒性；统计信息蒸馏则进一步将ST-SAM3教师模型的分布级语义知识迁移至轻量学生模型。若仅使用其中任一类监督，均难以同时满足准确定位、几何泛化与语义表征迁移的需求。因此，本文通过联合优化三个目标，使强监督、软监督和教师知识迁移在同一训练过程中协同发挥作用。综合上述约束，所提出DistillMatch的总训练损失定义为：

$$\mathcal{L}_{\text{total}}=\mathcal{L}_{\text{sup}}+\mathcal{L}_{\text{con}}+\lambda_{\text{stat}}\left(\mathcal{L}_{\text{distill}}^{\rho}+\mathcal{L}_{\text{distill}}^{\alpha}\right) \tag{17}$$

式中，$\mathcal{L}_{\text{distill}}^{\rho}$ 和 $\mathcal{L}_{\text{distill}}^{\alpha}$ 分别表示缩放视图与旋转视图下的统计蒸馏损失，ρ 表示缩放比例，α 表示旋转角度。

© 中国图象图形学报版权所有

$\lambda_{stat}=0.01$为缩放分支和旋转分支的统计信息蒸馏损失权重。总损失中，$\mathcal{L}_{sup}$提供可靠的任务监督，$\mathcal{L}_{con}$约束无标注样本上的预测稳定性，统计蒸馏项则补充教师模型的高阶语义知识。三者共同驱动学生模型在维持轻量化结构的同时，逐步逼近ST-SAM3教师模型的遥感显著性判别能力。

在DistillMatch训练完成后，可以不引入参数量巨大的ST-SAM3教师模型，仅仅保留参数量为26.91 M、计算量为14.19 GMACs的DINOv3-Small学生模型进行推理。由此，DistillMatch将训练阶段的强教师指导转化为轻量学生模型的推理能力，在显著降低模型存储和计算开销的同时，保持对复杂遥感显著目标的有效检测能力。

2 实验与分析

为验证本文所提出的的ST-SAM3自训练方法和DistillMatch蒸馏框架在低标注遥感显著目标检测任务中的有效性，本节将本文方法与多种半监督方法进行比较，包括FixMatch（Sohn等，2020）、ST（Yang等，2022）、ST++（Yang等，2022）、UniMatch（Yang等，2023）、CorrMatch（Sun等，2024）、RankMatch（Mai等，2024）、UniMatchV2（Yang等，2025）、CATNet（Chen等，2025a）、TripleNet（Chen等，2025b）、KnowSAM（Huang等，2025a）、SSMIS（Zhao等，2025）、DBMatch（Huang等，2025b）、USCD（Gao等，2026）和S3OD（Liu等，2026b）。上述方法涵盖了基于置信度伪标签、一致性正则化、自训练、相关性约束及基础模型适配等技术路线，从多个角度验证本文方法的性能优势。

2.1 数据集介绍

本文在ORSSD、EORSSD和ORSI-4199三个公开遥感显著目标检测数据集上开展实验。ORSSD数据集（Li等，2019）包含600幅训练图像和200幅测试图像，场景涵盖飞机、船舶、车辆、道路及建筑物等多类遥感显著目标；EORSSD数据集（Zhang等，2021）包含1400幅训练图像和600幅测试图像，具有更丰富的场景类型、目标尺度变化和背景干扰；ORSI-4199数据集（Tu等，2022）包含2000幅训练图像和2199幅测试图像，样本规模较大，且目标类别、成像条件及背景分布更为复杂。为模拟极低标注率场景，本文在每个数据集的训练集中分别按照5%和10%的标注比例构建初始有标注集合，其余训练样本作为初始无标注集合，具体划分如表1所示。

表1 三个数据集的不同样本标注划分比例的数据分布

Table 1 Data distribution of different labeled sample partition ratios across three datasets

数据集	训练集	测试集	5%样本标注	10%样本标注
ORSSD	600	200	30/570	60/540
EORSSD	1400	600	70/1330	140/1260
ORSI-4199	2000	2199	100/1900	200/1800

2.2 评价指标

为全面评价各方法的检测性能，本文采用Max F-measure、S-measure和平均绝对误差（mean absolute error，MAE）作为客观评价指标。其中，Max F-measure通过在不同阈值下计算精确率P和召回率R获得。S-measure从目标感知与区域感知两个角度评价预测掩码与真实掩码之间的结构相似性。MAE用于衡量预测显著图与真实掩码之间的平均像素误差。此外，本文绘制PR曲线和F-measure曲线，分析不同阈值下各方法在精确率与召回率之间的整体权衡关系。

2.3 实验部署细节

本文采用PyTorch2.3.1框架进行算法实现，显卡为NVIDIA A100-SXM4 GPU。所有对比方法均采用相同的训练集划分和测试集进行训练与评估；输入图像统一调整为448×448分辨率。在ST-SAM3阶段，本文利用公开的SAM3预训练权重进行初始化，并冻结图像编码器的主干参数，仅保留Adapter分支参与更新，同时训练掩码解码器，从而在保留视觉基础模型通用表征能力的基础上，适应遥感显著目标检测任务。ST-SAM3模型每个训练轮次训练30个epoch，批量大小设置为8，采用AdamW优化器，权重衰减设置为1e-4，初始学习率和最小学习率分别设置为4e-4和4e-5。全部训练阶段均采用余弦退火学习率策略，以平滑降低学习率并提升收敛稳定性。后续再训练阶段额外维护EMA模型，用于推理生成下一阶段的伪标签。

在DistillMatch阶段，加载经3轮渐进式自训练获得的ST-SAM3权重参数作为教师模型，并冻结教师模型。学生模型加载官方DINOv3-Small的权重，并进行100个epoch的训练。优化器采用AdamW，

© 中国图象图形学报版权所有

权重衰减设置为1e-4。分割头和DINOv3 backbone的初始学习率分别设置为2e-4和5e-6，并采用按迭代更新的多项式学习率衰减策略，衰减幂设置为0.9，学习率最终衰减至0。每个训练迭代包含4个扩展有标注样本和4个无标注样本，即总批量大小为8。对于无标注样本的弱增强视图，由冻结的ST-SAM3教师模型与EMA学生模型二者预测概率取平均后，以0.5为阈值进行二值化，并仅保留置信度不低于0.95的像素参与几何一致性损失的计算。统计信息蒸馏权重设为0.01。训练完成后，仅保留DINOv3-Small学生模型用于快速推理。

2.4 对比实验分析

2.4.1 定量分析

表2和表3分别给出了5%和10%有标注数据划分下，各方法在ORSSD、EORSSD及ORSI-4199数据集上的定量比较结果。总体而言，在两个标注比例、三个数据集对应的18项指标中，ST-SAM3取得了17项最优结果，仅在5%标注的EORSSD数据集上，其MAE略高于S3OD，表明ST-SAM3在不同标注比例和数据集上均具有稳定的检测优势。在5%标注设置下，与各数据集上表现最优的对比方法相比，ST-SAM3在ORSSD上的Max F-measure和S-measure分别提高2.01%和1.58%，MAE降低0.0025；在EORSSD上分别提高0.52%和0.56%；在ORSI-4199上分别提高1.83%和1.18%，MAE降低0.0048。以上表明，ST-SAM3在仅使用5%真实标注样本时，仍能保持理想的检测精度。

在10%标注设置下，ST-SAM3在三个数据集的所有指标上均取得最优结果。与已有方法中的最优结果相比，ST-SAM3在三个数据集上的Max F-measure分别提高1.23%、1.45%和1.82%；S-measure分别提高0.95%、0.97%和1.20%。该结果表明，当初始真实标签进一步增加时，可靠伪标签能够对有限的真实标注信息形成有效补充。

ST-SAM3的性能提升主要源于两个方面。其一，SAM3-Adapter继承了视觉基础模型的通用分割先验，能够在少量人工标注条件下获得较强的初始显著性预测能力；其二，本文通过多轮渐进式自训练、提示驱动细化和可靠性筛选，仅将高质量伪标签纳入扩展有标注集合，从而降低传统伪标签方法中“错误预测自我强化”的风险。相比之下，FixMatch、UniMatch、CorrMatch和RankMatch等方法主要依赖当前网络自身产生的伪标签或一致性约束，在遥感图像的复杂背景、细长目标和小尺度目标场景中更容易受到不可靠伪监督的影响。

DistillMatch虽然采用轻量DINOv3-Small作为学生网络，却在5%和10%标注设置下均稳定超过大多数对比方法，说明其能够有效继承ST-SAM3的显著目标判别知识。其优势在于：可靠扩展标注集提供强监督；将低可靠性困难样本保留为无标注数据，并通过缩放、旋转几何一致性挖掘其监督信息；高阶统计蒸馏则进一步实现师生特征分布对齐，增强学生模型的前景目标定位与背景抑制能力。

综上，5%和10%标注的实验共同验证了所提方法的有效性。具体而言，ST-SAM3借助视觉基础模型的先验知识与可靠的伪标签扩展策略，取得了优异的检测性能；DistillMatch通过对有标注和无标注样本的分层利用、几何一致性约束以及统计信息蒸馏，在采用轻量级模型的同时，实现了对ST-SAM3知识的有效迁移，并保持了理想的检测性能。

2.4.2 定性分析

图4展示了5%有标注比例下，各方法在三个数据集上的PR曲线和F-measure曲线。可以看出，ST-SAM3和DistillMatch的PR曲线在三个数据集上均更接近右上角，说明在相同召回率下能够保持更高的精确率，或在相同精确率下获得更高的召回率；同时，ST-SAM3和DistillMatch的F-measure曲线在较宽阈值范围内保持较高水平，且峰值更大，表明本文所提出的方法对阈值选择的敏感性较低，具有更稳定的显著目标预测图。尤其在EORSSD和ORSI-4199这两个更复杂的数据集上，所提方法的曲线优势更为明显，验证了可靠伪标签筛选和几何一致性学习对提升预测稳定性的作用。

图5展示了5%训练样本标注设置下在EORSSD数据集上的检测结果对比图。ST-SAM3能够更完整地覆盖真实显著目标，并保持较为清晰的边界；DistillMatch的预测结果与ST-SAM3接近，在河流主体、岛屿连通性及建筑轮廓方面均表现出更好的结构保持能力。这说明ST-SAM3的视觉基础模型先验和可靠伪标签扩展能够提升目标定位质量，而DistillMatch可通过几何一致性学习和统计知识迁移有效保留教师模型的关键显著性目标表征。

图6进一步展示了10%标注设置下ORSI-4199数据集的显著目标检测图的对比结果。图6中的典

© 中国图象图形学报版权所有

表2　不同算法在5%的有标签数据划分的三个数据集上的定量对比结果，加粗字体为每列最优值

Table 2　Quantitative comparison results of different algorithms on three datasets with 5% labeled data partitions.

/Bold values indicate the best performance in each column

方法	发表物	ORSSD数据集			EORSSD数据集			ORSI-4199数据集		
		MAE ↓	F_β ↑	S_m ↑	MAE ↓	F_β ↑	S_m ↑	MAE ↓	F_β ↑	S_m ↑
FixMatch	NeurIPS'20	0.0216	0.8603	0.8885	0.0114	0.8421	0.8938	0.0366	0.8461	0.8597
ST	CVPR'22	0.0356	0.8110	0.8560	0.0149	0.8323	0.8882	0.0426	0.8252	0.8488
ST++	CVPR'22	0.0241	0.8461	0.8879	0.0128	0.8550	0.8998	0.0375	0.8421	0.8594
UniMatch	CVPR'23	0.0214	0.8679	0.8936	0.0140	0.8465	0.8975	0.0337	0.8492	0.8694
CorrMatch	CVPR'24	0.0207	0.8620	0.8969	0.0114	0.8348	0.8914	0.0338	0.8514	0.8669
RankMatch	CVPR'24	0.0202	0.8716	0.8965	0.0109	0.8436	0.8970	0.0345	0.8502	0.8698
UniMatchV2	TPAMI'25	0.0220	0.8715	0.8936	0.0126	0.8523	0.8976	0.0353	0.8476	0.8643
CATNet	TCSVT'25	0.0204	0.8705	0.8995	0.0117	0.8549	0.9011	0.0356	0.8538	0.8652
TripleNet	TIP'25	0.0192	0.8783	0.8957	0.0118	0.8399	0.8906	0.0363	0.8446	0.8593
KnowSAM	TMI'25	0.0186	0.8794	0.8938	0.0120	0.8500	0.8847	0.0372	0.8332	0.8501
SSMIS	NeurIPS'25	0.0178	0.8969	0.9086	0.0115	0.8506	0.8972	0.0336	0.8501	0.8655
DBMatch-DINOv3	TGRS'25	0.0206	0.8732	0.8949	0.0142	0.8392	0.8894	0.0300	0.8623	0.8716
USCD-DINOv3	TGRS'26	0.0158	0.8918	0.9105	0.0107	0.8530	0.8987	0.0312	0.8591	0.8704
S3OD	ACM MM'26	0.0168	0.9064	0.9182	**0.0086**	0.8891	0.9198	0.0302	0.8636	0.8771
ST-SAM3（本文方法）	—	**0.0133**	**0.9265**	**0.9340**	0.0091	**0.8943**	**0.9254**	**0.0252**	**0.8819**	**0.8889**
DistillMatch（本文方法）	—	0.0134	0.9233	0.9283	0.0093	0.8890	0.9222	0.0267	0.8779	0.8857

注：加粗字体为每行最优值。

型样本包含飞机、蜿蜒道路和不规则拓扑的建筑，具有更强的背景干扰和目标间混淆。对于道路和小型飞机等目标，部分方法存在显著区域缺失、目标粘连或将相似背景纹理误检为前景的问题；对于不规则拓扑场景，部分方法难以同时保持显著目标的完整性。ST-SAM3在这些样本中能够更准确地定位多个独立目标，减少背景区域的错误激活，并保持道路和飞机等目标的形状完整性。

DistillMatch使用轻量学生网络，但其检测结果仍能较好地恢复目标轮廓、抑制背景噪声，并明显优于各对比方法，进一步说明几何一致性学习和高阶统计信息蒸馏有助于学生模型学习教师模型的多尺度语义判别能力。

2.4.3　模型复杂度分析

由表4可见，ST-SAM3作为基于SAM3 ViT-L的教师模型，参数量为458.18 M，计算量为631.28

© 中国图象图形学报版权所有

表3　不同算法在10%的有标签数据划分的三个数据集上的定量对比结果，加粗字体为每列最优值

Table 3　Quantitative comparison results of different algorithms on three datasets with 10% labeled data partitions.

/Bold values indicate the best performance in each column

方法	发表物	ORSSD数据集			EORSSD数据集			ORSI-4199数据集		
		MAE ↓	F_β ↑	S_m ↑	MAE ↓	F_β ↑	S_m ↑	MAE ↓	F_β ↑	S_m ↑
FixMatch	NeurIPS'20	0.0167	0.8858	0.9069	0.0120	0.8557	0.9038	0.0358	0.8505	0.8671
ST	CVPR'22	0.0240	0.8647	0.8951	0.0132	0.8457	0.8973	0.0387	0.8478	0.8607
ST++	CVPR'22	0.0178	0.8900	0.9103	0.0113	0.8669	0.9106	0.0368	0.8550	0.8639
UniMatch	CVPR'23	0.0146	0.9023	0.9182	0.0134	0.8610	0.9076	0.0350	0.8515	0.8688
CorrMatch	CVPR'24	0.0161	0.8804	0.9098	0.0103	0.8567	0.9080	0.0345	0.8570	0.8693
RankMatch	CVPR'24	0.0158	0.8964	0.9117	0.0125	0.8521	0.8960	0.0336	0.8512	0.8691
UniMatchV2	TPAMI'25	0.0157	0.9003	0.9134	0.0114	0.8630	0.9110	0.0358	0.8538	0.8703
CATNet	TCSVT'25	0.0163	0.8866	0.9116	0.0095	0.8794	0.9186	0.0353	0.8584	0.8692
TripleNet	TIP'25	0.0185	0.8889	0.9065	0.0108	0.8549	0.9039	0.0360	0.8520	0.8657
KnowSAM	TMI'25	0.0190	0.8808	0.8997	0.0104	0.8615	0.8955	0.0350	0.8480	0.8593
SSMIS	NeurIPS'25	0.0154	0.9052	0.9131	0.0095	0.8759	0.9146	0.0336	0.8568	0.8748
DBMatch-DINOv3	TGRS'25	0.0149	0.9011	0.9170	0.0103	0.8612	0.9043	0.0307	0.8617	0.8731
USCD-DINOv3	TGRS'26	0.0126	0.9136	0.9227	0.0078	0.8719	0.9118	0.0318	0.8577	0.8737
S3OD	ACM MM'26	0.0127	0.9205	0.9284	0.0086	0.8920	0.9251	0.0312	0.8649	0.8786
ST-SAM3（本文方法）	—	**0.0120**	**0.9328**	**0.9379**	**0.0073**	**0.9065**	**0.9348**	**0.0253**	**0.8831**	**0.8906**
DistillMatch（本文方法）	—	0.0129	0.9211	0.9303	0.0078	0.8964	0.9276	0.0257	0.8812	0.8883

注：加粗字体为每行最优值。

GMACs，推理速度为32 FPS，对应单张图像推理时延为31.25 ms。ST-SAM3主要用于离线任务适配、伪标签生成和可靠性筛选，并不作为最终部署模型，因此其较高的模型复杂度主要服务于教师模型的高质量表征和伪标签生成。经过DistillMatch模型蒸馏后，最终学生模型的参数量降低至26.91 M，计算量降低至14.19 GMACs，推理速度提升至83 FPS，对应单张图像推理时延约为12.05 ms。相较于ST-SAM3，DistillMatch的参数量减少约94.13%，计算量减少约97.75%，推理速度约提升至2.59倍，说明本文方法能够将高容量视觉基础模型的强大表征能力迁移至更加紧凑的轻量级学生网络中。

2.4.4　划分随机性分析

为补充单次固定划分之外的性能观察，本文在ORSSD、EORSSD和ORSI-4199三个数据集的5%与10%初始标注设置下，分别考察五种随机有标注子

© 中国图象图形学报版权所有

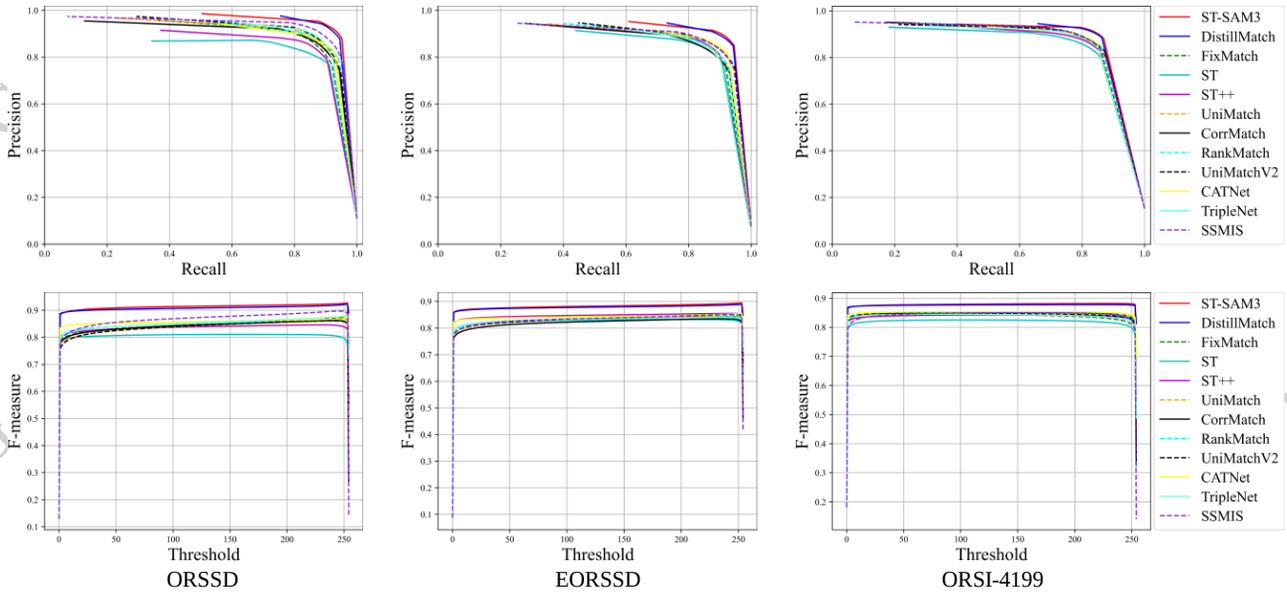


图4　不同方法在三个数据集上5%有标签训练数据下的PR曲线与F-measure曲线对比

Fig. 4　Comparison of PR curves and F-measure curves of different methods on three datasets with 5% labeled training data

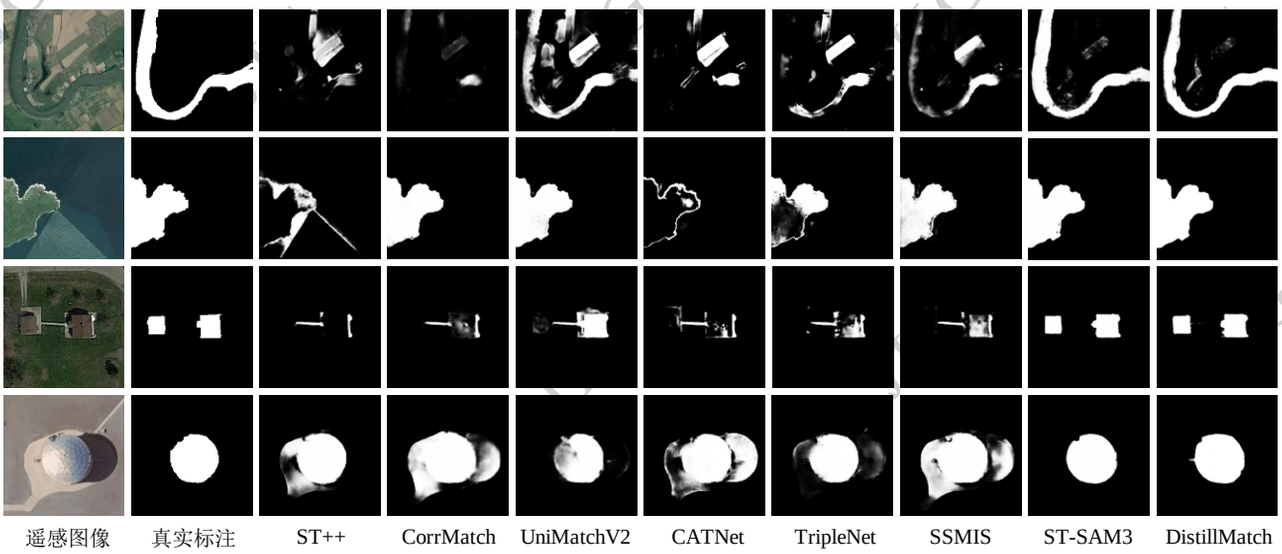


图5　不同方法在5% EORSSD数据集上的显著目标检测可视化结果对比

Fig. 5　Comparison of visual results for salient object detection of different methods on the EORSSD dataset with 5% labeled data

集划分，并报告ST-SAM3和DistillMatch的指标均值及样本标准差，结果如表5所示。

在六种数据集与标注比例组合中，ST-SAM3的Max F-measure和S-measure标准差分别为0.0010～0.0054和0.0014～0.0037，DistillMatch对应的标准差分别为0.0009～0.0043和0.0009～0.0035，表明所报告实验的检测结果总体较为集中。随着初始标注比例由5%增加至10%，ST-SAM3在ORSSD和EORSSD上的平均Max F-measure分别提高0.44和0.86个百分点，DistillMatch分别提高0.84和0.62个百分点。相比之下，ORSI-4199上的增益较小，两种标注比例的平均性能较为接近，说明增加标注样本带来的收益具有数据集依赖性，并不意味着每个划分下的各项指标均必然单调改善。上述结果量化了多划分实验的性能离散程度，为理解所提框架在不同标注子集下的实验表现提供了补充依据。

2.4.5　学生模型对比分析

为考察DistillMatch对不同学生骨干网络的适配性，表6和表7分别给出了5%和10%标注设置下的对比结果，新增了参数规模相近的SegFormer-B2

© 中国图象图形学报版权所有

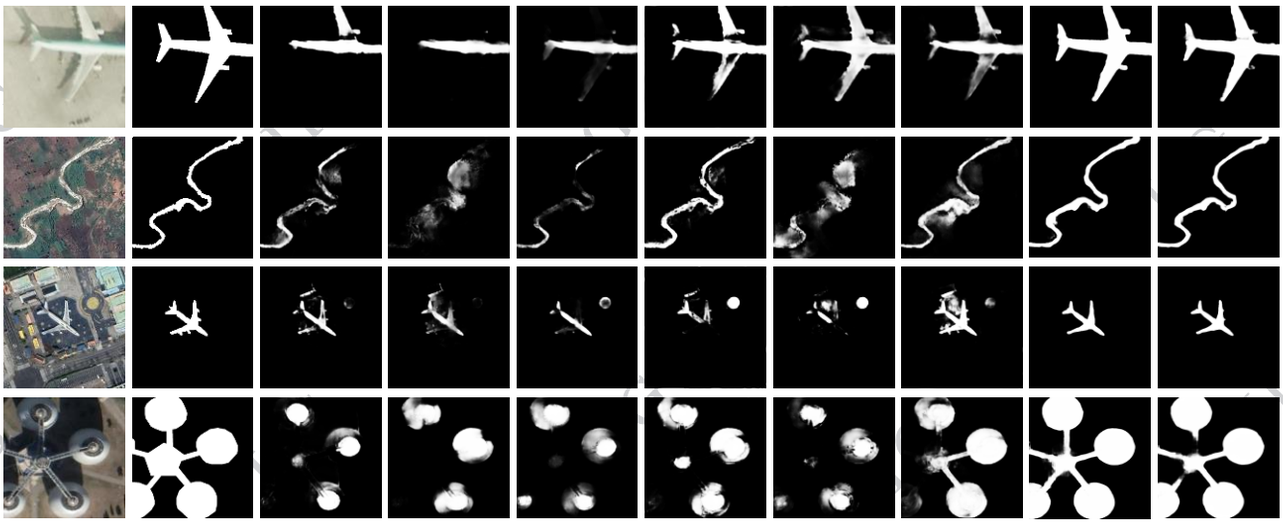


图6　不同方法在10% ORSI-4199数据集上的显著目标检测可视化结果对比

Fig. 6　Comparison of visual results for salient object detection of different methods on the ORSI-4199 dataset with 10% labeled data

表4　所提出方法与对比方法的模型复杂度对比

Table 4　Comparison of model complexity between the proposed method and comparison methods

方法	参数量M（M）	浮点次数GMACs	推理速度FPS(ms)
CATNet	27.36	20.38	117
KnowSAM	143.05	94.16	71
SSMIS	471.86	641.46	127
DBMatch-DINOv3	26.91	14.19	83
USCD-DINOv3	26.91	14.19	83
S3OD	**13.68**	**10.19**	**176**
ST-SAM3(本文方法)	458.18	631.28	32
DistillMatch(本文方法)	26.91	14.19	83

注:加粗字体为每行最优值。

和DINOv2 ViT-S/14学生模型。表中三种学生模型的部署参数量分别为24.72 M、27.36 M和26.91 M。结果表明,DistillMatch能够适配不同学生骨干,且DINOv3并非在所有指标上均占优。例如,在5%标注的ORSSD数据集上,DINOv2的S-measure达到0.9354,高于DINOv3的0.9283;在10%标注设置下,DINOv2在多个指标上也取得了更好的结果。另一方面,DINOv3学生模型的计算量为14.19 GMACs,大大低于另外两种学生模型,体现了其在检测性能与计算开销之间的权衡。本文综合考虑预训练表征能力与部署效率,选择DINOv3-Small作为最终学生骨干网络,但并不声称其是该学生框架的最优选择。与同样采用DINOv3-Small的DBMatch和USCD相比(见表2和表3),DistillMatch在三个数据集、两种标注比例下的Max F-measure和S-measure指标均更高。结合前述分析,DistillMatch的性能优势不能归因于骨干网络的预训练能力,本文提出的可靠伪标签扩充、几何一致性学习和统计信息蒸馏等训练策略发挥了更为重要的作用。

2.5　消融实验分析

表8验证了DistillMatch各模块在EORSSD数据集5%标注设置下的有效性,表明其性能增益并非

© 中国图象图形学报版权所有

表5　ST-SAM3和DistillMatch在5%和10%的三个数据集上的5种随机划分的均值与标准差

Table 5　Mean and standard deviation of ST-SAM3 and DistillMatch across five labeled-subset splits on three datasets

方法			ORSSD数据集			EORSSD数据集			ORSI-4199数据集		
			MAE ↓	F_β ↑	S_m ↑	MAE ↓	F_β ↑	S_m ↑	MAE ↓	F_β ↑	S_m ↑
5%有标签+95%无标签	ST-SAM3	均值	0.0129	0.9293	0.9352	0.0084	0.8990	0.9289	0.0254	0.8814	0.8900
		±标准差	0.0012	0.0043	0.0037	0.0006	0.0054	0.0036	0.0006	0.0010	0.0014
	DistillMatch	均值	0.0142	0.9168	0.9274	0.0088	0.8891	0.9227	0.0263	0.8790	0.8871
		±标准差	0.0013	0.0043	0.0035	0.0005	0.0040	0.0023	0.0003	0.0010	0.0009
10%有标签+90%无标签	ST-SAM3	均值	0.0116	0.9337	0.9386	0.0070	0.9076	0.9344	0.0255	0.8830	0.8908
		±标准差	0.0012	0.0051	0.0035	0.0010	0.0024	0.0027	0.0010	0.0023	0.0019
	DistillMatch	均值	0.0116	0.9252	0.9324	0.0074	0.8953	0.9271	0.0263	0.8800	0.8873
		±标准差	0.0012	0.0038	0.0027	0.0009	0.0030	0.0022	0.0006	0.0009	0.0009

表6　不同学生骨干网络下DistillMatch在三个数据集5%标注设置下的模型复杂度与检测性能对比

Table 6　Model complexity and detection performance of DistillMatch with different student backbones with 5% labeled data

方法	参数量（M）	计算量（GMACs）	ORSSD数据集			EORSSD数据集			ORSI-4199数据集		
			MAE ↓	F_β ↑	S_m ↑	MAE ↓	F_β ↑	S_m ↑	MAE ↓	F_β ↑	S_m ↑
ST-SAM3	458.18	631.28	**0.0133**	**0.9265**	0.9340	0.0091	**0.8943**	**0.9254**	**0.0252**	**0.8819**	**0.8889**
SegFormer-DistillMatch	**24.72**	16.36	0.0143	0.9163	0.9285	0.0091	0.8887	0.9207	0.0302	0.8712	0.8784
DINOv2-DistillMatch	27.36	23.82	0.0134	0.9231	**0.9354**	**0.0087**	0.8889	0.9207	0.0261	0.8793	0.8871
DINOv3-DistillMatch	26.91	**14.19**	0.0134	0.9233	0.9283	0.0093	0.8890	0.9222	0.0267	0.8779	0.8857

注：加粗字体为每行最优值。

表7　不同学生骨干网络下DistillMatch在三个数据集10%标注设置下的模型复杂度与检测性能对比

Table 7　Model complexity and detection performance of DistillMatch with different student backbones with 10% labeled data

方法	参数量（M）	计算量（GMACs）	ORSSD数据集			EORSSD数据集			ORSI-4199数据集		
			MAE ↓	F_β ↑	S_m ↑	MAE ↓	F_β ↑	S_m ↑	MAE ↓	F_β ↑	S_m ↑
ST-SAM3	458.18	631.28	**0.0120**	**0.9328**	**0.9379**	0.0073	**0.9065**	**0.9348**	**0.0253**	**0.8831**	**0.8906**
SegFormer-DistillMatch	**24.72**	16.36	0.0125	0.9263	0.9353	0.0072	0.8950	0.9267	0.0283	0.8767	0.8839
DINOv2-DistillMatch	27.36	23.82	0.0124	0.9276	0.9338	**0.0071**	0.9013	0.9301	**0.0253**	0.8830	0.8895
DINOv3-DistillMatch	26.91	**14.19**	0.0129	0.9211	0.9303	0.0078	0.8964	0.9276	0.0257	0.8812	0.8883

来自单一模块，而是由“可靠强监督—几何鲁棒性—语义迁移”逐步累积形成。首先，仅引入标签扩充便使MAE从0.0133降至0.0098，降幅约为26.3%，同时Max F-measure提升4.56个百分点，说明在5%初始标注条件下，增加可靠监督能够明显改善学生模型性能；ST-SAM3提供的可靠伪标签有效扩大了可用于直接监督的样本覆盖范围。进一步引入几何一致性后，Max F-measure在标签扩充基础上继续增加

© 中国图象图形学报版权所有

0.59个百分点，且S-measure增加0.54个百分点。该结果表明，缩放和旋转等几何扰动下的一致性学习能够进一步利用无标注样本，并增强模型对目标空间结构变化的适应能力。最后，统计信息蒸馏使Max F-measure提升0.45个百分点，并将S-measure进一步提升0.30个百分点，说明统计信息蒸馏能够在输出一致性约束之外提供额外的特征层监督，有助于改善显著区域定位和复杂背景抑制能力。相较于基线模型，最终DistillMatch的MAE总降幅达到30.1%，上述结果进一步支持了三个模块在监督覆盖、空间几何结构和特征语义层面的互补作用。

表9比较了不同SAM版本的视觉基础模型在5%标注比例下的检测性能。可以看出，ST-SAM3在EORSSD数据集上的三项指标均取得最优结果，其MAE、Max F-measure和S-measure分别达到0.0091、0.8943和0.9254。与ST-SAM1相比，ST-SAM3的MAE降低了0.0021，Max F-measure提高了3.15%，S-measure提高了1.90%；与ST-SAM2相比，MAE进一步降低了0.0022，Max F-measure和S-measure分别提高了2.40%和1.34%。上述结果表明，在极低标注场景下，SAM3能够提供更加有效的视觉表征和下游迁移能力。与SAM1和SAM2相比，SAM3在预训练数据规模、数据多样性和视觉表征能力方面进行了进一步增强，并具有更强的提示驱动掩膜预测能力，这是其在遥感显著目标尺度变化、复杂背景和边界细节处理方面表现更优的重要原因。尤其是在仅使用5%真实标注数据的条件下，ST-SAM3仍在三项指标上保持一致的优势。

表10表明，扩展有标注样本与剩余无标注样本之间并非“越多伪标签越好”的简单关系，而需要在强监督质量和软监督覆盖之间取得平衡。与仅使用5%原始标注、95%无标注样本进行一致性学习的设置相比，75%/25%划分使Max F-measure和S-measure分别提高0.63%和0.60%，说明将高可靠伪标签转化为强监督比单纯扩大无标注一致性学习规模更有效。然而，当全部训练样本均作为扩展有标注数据时，Max F-measure和S-measure分别较75%/25%划分下降0.28%和0.38%，说明低可信度困难样本若被直接赋予硬标签，会重新引入伪标签噪声。值得注意的是，25%/75%扩展比例下取得最低MAE（0.0087），而75%/25%划分在Max F-measure和S-measure上取得最优结果，反映出不同指标之间存在轻微权衡。综合考虑三项指标，本文最终选择75%/25%划分方案将可靠伪标签用于强监督训练，将相对困难的样本保留在无标注集合中，并通过一致性约束进一步利用其潜在监督信息。上述结果支持了“可靠样本用于强监督、困难样本用于一致性学习”的分层训练策略。

表8　DistillMatch在5%的EORSSD上的消融实验对比

Table 8　Ablation study comparison of DistillMatch on the EORSSD dataset with 5% labeled data

标签扩充	几何一致	统计蒸馏	MAE ↓	F_β ↑	S_m ↑
—	—	—	0.0133	0.8330	0.8878
√	—	—	0.0098	0.8786	0.9138
√	√	—	0.0096	0.8845	0.9192
√	√	√	**0.0093**	**0.8890**	**0.9222**

注：加粗字体为每行最优值。√表示模块已加入，—表示模块未加入。

表9　基于不同SAM版本模型在5%的EORSSD数据集的对比 Table 9 Comparison of different SAM-based foundation models on the EORSSD dataset with 5% labeled data

方法	骨干网络	MAE ↓	F_β ↑	S_m ↑
ST-SAM1	ViT-H	0.0112	0.8628	0.9064
ST-SAM2	ViT-Hiera-L	0.0113	0.8703	0.9120
ST-SAM3（本文算法）	ViT-L+	**0.0091**	**0.8943**	**0.9254**

注：加粗字体为每行最优值。

表10　DistillMatch在5% EORSSD上不同伪标签划分的对比

Table 10　Comparison of DistillMatch with different pseudo-label partitions on EORSSD with 5% labeled data

扩充后有标签占比	剩余无标签占比	MAE ↓	F_β ↑	S_m ↑
5%（不扩充）	95%（不扩充）	0.0098	0.8827	0.9162
25%	75%	**0.0087**	0.8851	0.9199
50%	50%	0.0096	0.8824	0.9188
60%	40%	0.0093	0.8856	0.9212
75%	25%	0.0093	**0.8890**	**0.9222**
100%	0%	0.0095	0.8862	0.9184

注：加粗字体为每行最优值。

© 中国图象图形学报版权所有

2.6 可视化分析

本小节从ST-SAM3的伪标签质量、ST-SAM3的样本评估机制和DistillMatch的失败预测表现这几个方面开展可视化分析。

图7展示了在5%样本标注设置下不同方法在ORSI-4199数据集上的伪标签对比结果。其中，白色区域表示正确前景，黑色区域表示正确背景，绿色区域表示漏检前景，红色区域表示错检背景。可见，部分对比方法未能完整覆盖细长目标，产生明显漏检；部分对比方法则将相似纹理、阴影或邻近地物误识别为前景，导致背景错检。相比之下，ST-SAM3的绿色漏检和红色错检区域显著减少，尤其能够较完整地恢复多个相邻目标、细长目标及边界不规则目标。这说明自适应提示细化和多维一致性约束能够筛选出更可靠的伪标签，从源头上降低错误伪标签在多轮监督自训练过程中的累积。

图8揭示了多维一致性约束如何识别潜在噪声伪标签。图中每一行依次给出原始图像及自适应正负点提示、初始伪标签、两次SAM提示细化结果、水平翻转预测和真实掩码标注，每一行下方标注了综合一致性得分、SAM细化一致性得分、水平翻转扰动一致性得分及其全体排序。可见，这些样本在不同预测分支间均存在明显的不稳定性：对于密集储罐目标，SAM3的第二轮迭代细化结果使得预测不完整储罐目标的完整性变好，但是水平翻转预测仍然存在储罐目标的前景缺失；对于图9中的飞机与桥梁等细长目标，提示细化造成了前景的局部缺失或虚假扩张；对于小尺度目标和复杂纹理背景，初始预测与后续分支之间还会出现额外虚假前景区域或显著区域消失。这些不一致现象会同时降低SAM细化一致性和扰动一致性得分，使样本获得较低的综合分数和靠后的排序，最终不被纳入扩展有标注集合。该结果说明，本文提出的可靠性评估机制能够主动排除结构响应不稳定的困难样本，避免错误伪标签在后续自训练中被当作硬监督使用。

图9展示了DistillMatch在5%标注EORSSD数据集上的典型失败检测案例。方法在外观干扰强、目标尺度极小或拓扑结构高度不规则的样本中仍存在漏检或错检。例如，部分船舶周围的水花、尾迹，模型可能将其错误扩张为前景。对于极小目标，其有效像素数量有限，经过特征下采样后目标响应容易被复杂背景淹没，从而产生漏检。与此同时，针对弯曲道路、河流或不规则拓扑结构的遥感目标，模型可能出现结构断裂、缺失检测等现象。上述失败案例表明，DistillMatch对低对比度目标、微小目标及复杂拓扑结构的建模能力仍有待提升。

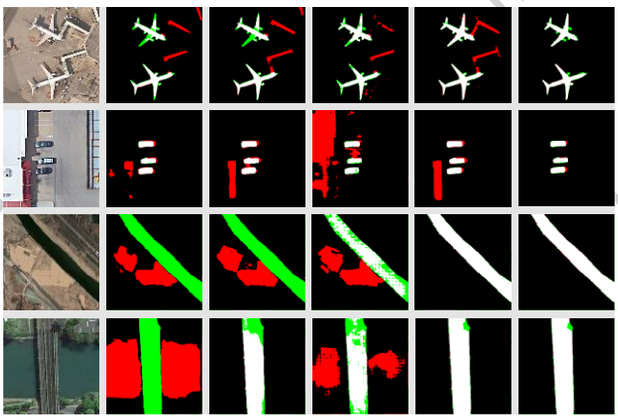


图7 不同方法在5% ORSI-4199数据集上的伪标签可视化

Fig. 7 Visualization of pseudo-labels generated by different methods on the ORSI-4199 dataset with 5% labeled data

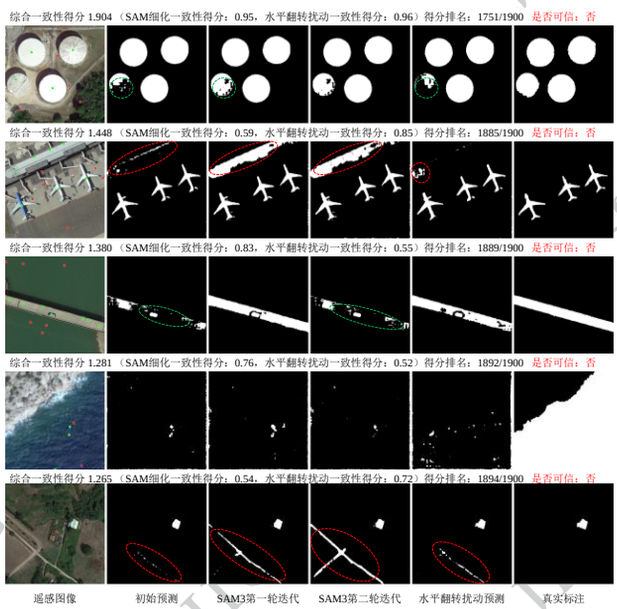


图8 ST-SAM3的多维一致性约束可靠性评估可视化

Fig. 8 Reliability assessment visualization of ST-SAM3 via multi-dimensional consistency constraints

3 结 论

本文针对半监督遥感显著目标检测中标注稀缺、伪标签噪声累积及视觉基础模型部署成本高的问题，提出由ST-SAM3和DistillMatch组成的两阶段

© 中国图象图形学报版权所有

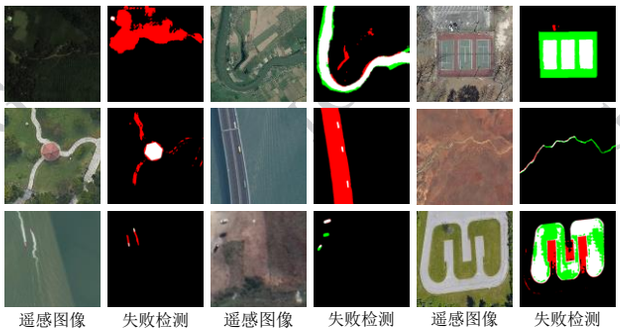


图9　DistillMatch在5% EORSSD数据集上的失败检测案例

Fig. 9　Failure cases of DistillMatch on 5% labeled EORSSD

学习框架。ST-SAM3结合渐进式自训练、自适应提示细化和多维一致性评估，构建可靠伪标签；DistillMatch通过几何一致性学习与高阶统计信息蒸馏，将教师模型的显著目标检测能力迁移至轻量级学生网络。在ORSSD、EORSSD和ORSI-4199数据集上的实验表明，所提方法在低标注条件下取得了显著优于现有半监督对比方法的检测性能。消融实验与可视化分析验证了可靠伪标签扩充、几何一致性学习和统计信息蒸馏的互补作用，说明视觉基础模型先验能够有效缓解监督不足，并通过蒸馏兼顾检测性能与部署成本。本文所提方法在极小目标、低对比度、阴影及复杂结构场景中仍可能出现漏检、错检和结构断裂。后续将探索更高效的伪标签更新策略，并结合不确定性建模、边界约束与拓扑先验，提升复杂遥感场景下的低标注视觉感知鲁棒性。

参考文献(References)

Chen L Y, Liu W, Wang H, Jeon S W, Jiang Y L and Zheng Z L. 2025a. Consistency-guided adaptive alternating training for semi-supervised salient object detection. IEEE Transactions on Circuits and Systems for Video Technology, 35(7): 7033-7046 [DOI: 10.1109/TCSVT.2025.3539471]

Chen L Y, Yang M H, Pu J and Zheng Z L. 2025b. TripleNet: Exploiting complementary features and pseudo-labels for semi-supervised salient object detection. IEEE Transactions on Image Processing, 34: 5628-5641 [DOI: 10.1109/TIP.2025.3601334]

Chen T R, Cao R L, Yu X D, Zhu L Y, Ding C T, Ji D, et al. 2025c. SAM3-Adapter: Efficient adaptation of Segment Anything 3 for camouflage object segmentation, shadow detection, and medical image segmentation [EB/OL]. [2025-11-24]. https://arxiv.org/pdf/2511.19425.pdf

Fu C B, Zhuang Q Y, Lei S, Tian A H and Shi Z W. 2026. CPDM: Class prototypes with dynamic multidomain mixed perturbation consistency for semi-supervised image segmentation. IEEE Transactions on Geoscience and Remote Sensing, 64: 1-16 [DOI: 10.1109/TGRS.2025.3645047]

Gao Y P, Li T Y, Yin P X, Hu Y W, Chen S and Zhang W. 2026. Uncertainty-guided semisupervised change detection for remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 64: 1-17 [DOI: 10.1109/TGRS.2026.3710667]

Huang K W, Zhou T, Fu H Z, Zhang Y Z, Zhou Y, Gong C, et al. 2025a. Learnable prompting SAM-induced knowledge distillation for semi-supervised medical image segmentation. IEEE Transactions on Medical Imaging, 44(5): 2295-2306 [DOI: 10.1109/TMI.2025.3530097]

Huang W, Gu Z Q, Shi Y L, Xiong Z T and Zhu X X. 2025b. Semi-supervised building footprint extraction using debiased pseudo-labels. IEEE Transactions on Geoscience and Remote Sensing, 63: 1-14 [DOI: 10.1109/TGRS.2024.3524300]

Jin J D, Lu W X, Yu H F, Rong X E, Sun X and Wu Y R. 2024. Dynamic and adaptive self-training for semi-supervised remote sensing image semantic segmentation. IEEE Transactions on Geoscience and Remote Sensing, 62: 1-14 [DOI: 10.1109/TGRS.2024.3407142]

Lan M, Zhang L F, Du B, Zhang L P. 2024. Spatio-temporal hierarchical query for referring video object segmentation. Scientia Sinica Informationis, 54(3): 674-691 (兰猛, 张乐飞, 杜博, 张良培. 2024. 基于时空层级查询的指代视频目标分割. 中国科学: 信息科学), 54(3): 674-691 [DOI: 10.1360/SSI-2023-0030]

Li C Y, Cong R M, Hou J H, Zhang S Y, Qian Y and Kwong S. 2019. Nested network with two-stream pyramid for salient object detection in optical remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 57(11): 9156-9166 [DOI: 10.1109/TGRS.2019.2925070]

Liu C Y, Zhang J F, Chen K Y, Wang M, Zou Z X and Shi Z W. 2026a. Remote sensing spatiotemporal vision-language models: A comprehensive survey. IEEE Geoscience and Remote Sensing Magazine, 14(1): 383-423 [DOI: 10.1109/MGRS.2025.3598283]

Liu Y F, Xiong Z T, Yuan Y and Wang Q. 2023a. Distilling knowledge from super-resolution for efficient remote sensing salient object detection. IEEE Transactions on Geoscience and Remote Sensing, 61: 1-16 [DOI: 10.1109/TGRS.2023.3267271]

Liu Y F, Xiong Z T, Yuan Y and Wang Q. 2023b. Transcending pixels: Boosting saliency detection via scene understanding from aerial imagery. IEEE Transactions on Geoscience and Remote Sensing, 61: 1-16 [DOI: 10.1109/TGRS.2023.3298661]

Liu Y F, Zhang X, Guo W and Zhang L F. 2025. Structured cross-resolution distillation for remote sensing salient object detection. IEEE Transactions on Geoscience and Remote Sensing, 63: 1-14 [DOI: 10.1109/TGRS.2025.3598956]

Liu Y F, Chen X Y and Zhang L F. 2026b. S3OD: Bridging Segment Anything Model and semi-supervised learning for salient object

© 中国图象图形学报版权所有

detection in aerial imagery//Proceedings of the 34th ACM International Conference on Multimedia. Rio de Janeiro, Brazil: ACM: [DOI: 10.1145/3767308.3836228]

Mai H Y, Sun R, Zhang T Z and Wu F. 2024. RankMatch: Exploring the better consistency regularization for semi-supervised semantic segmentation//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE: 3391-3401 [DOI: 10.1109/CVPR52733.2024.00326]

Siméoni O, Vo H V, Seitzer M, Baldassarre F, Oquab M, Jose C, et al. 2025. DINOv3 [EB/OL]. [2026-08-28]. https://arxiv.org/pdf/2508.10104.pdf

Sohn K, Berthelot D, Carlini N, Zhang Z, Zhang H, Raffel C A, et al. 2020. FixMatch: Simplifying semi-supervised learning with consistency and confidence//Advances in Neural Information Processing Systems. 33: 596-608

Sun B Y, Yang Y Q, Zhang L, Cheng M M and Hou Q B. 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: IEEE: 3097-3107 [DOI: 10.1109/CVPR52733.2024.00299]

Sun X, Meng Y, Diao W H, Huang L J, Zhang X, Luo J C, Gao L R, Wang P J, Yan Z Y, Gao L J, Dong W, Feng Y C, Li J H and Fu K. 2022. The review of AI-based intelligent remote sensing capabilities. Journal of Image and Graphics, 27(6): 1799-1822 (孙显, 孟瑜, 刁文辉, 黄丽佳, 张新, 骆剑承, 高连如, 王佩瑾, 闫志远, 郜丽静, 董文, 冯瑛超, 李霁豪, 付琨. 2022. 智能遥感:AI赋能遥感技术. 中国图象图形学报, 27(6): 1799-1822) [DOI: 10.11834/jig.220161]

Tu Z Z, Wang C, Li C L, Fan M H, Zhao H F and Luo B. 2022. ORSI salient object detection via multiscale joint region and boundary model. IEEE Transactions on Geoscience and Remote Sensing, 60: 1-13 [DOI: 10.1109/TGRS.2021.3101359]

Wang Q, Liu Y F, Xiong Z T and Yuan Y. 2022. Hybrid feature aligned network for salient object detection in optical remote sensing imagery. IEEE Transactions on Geoscience and Remote Sensing, 60: 1-15 [DOI: 10.1109/TGRS.2022.3181062]

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//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, LA, USA: IEEE: 4268-4277 [DOI: 10.1109/CVPR52688.2022.00423]

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//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, BC, Canada: IEEE: 7236-7246 [DOI: 10.1109/CVPR52729.2023.00699]

Yang L H, Zhao Z and Zhao H S. 2025. UniMatch V2: Pushing the limit of semi-supervised semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 47(4): 3031-3048 [DOI: 10.1109/TPAMI.2025.3528453]

Yu Z P, Liu C Y, Liu L Q, Shi Z W and Zou Z X. 2025. MetaEarth: A generative foundation model for global-scale remote sensing image generation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 47(3): 1764-1781 [DOI: 10.1109/TPAMI. 2024.3507010]

Zhang Q J, Cong R M, Li C Y, Cheng M M, Fang Y M, Cao X C, et al. 2021. Dense attention fluid network for salient object detection in optical remote sensing images. IEEE Transactions on Image Processing, 30: 1305-1317 [DOI: 10.1109/TIP.2020.3042084]

Zhang W K, Liu W J, Sun X, Xu G L and Fu K. 2022. Multi-source features adaptation fusion network for semantic segmentation in high-resolution remote sensing images. Journal of Image and Graphics, 27(8): 2516-2526 (张文凯, 刘文杰, 孙显, 许光銮, 付琨. 2022. 多源特征自适应融合网络的高分遥感影像语义分割. 中国图象图形学报, 27(8): 2516-2526) [DOI: 10.11834/jig.210054]

Zhao H F, Li H Y, Ma L L and Sun D D. 2025. Segment Anything Model meets semi-supervised medical image segmentation: A novel perspective//Advances in Neural Information Processing Systems. 38: 170128-170148

作者简介

刘焰锋,男,博士研究生,主要研究方向为多模态遥感信息处理、视觉基础模型、多模态大模型。E-mail: liuyanfeng99@whu.edu.cn

张乐飞,通信作者,男,教授,博士生导师,主要研究方向为人工智能、多模态遥感大模型。E-mail:zhanglefei@whu.edu.cn

王迪,男,博士后,主要研究方向为空天信息大模型。E-mail: d_wang@whu.edu.cn

郭昊南,男,博士后,主要研究方向为时空智能大模型。E-mail:haonan.guo@whu.edu.cn

张敬,男,教授,博士生导师,主要研究方向为视觉大模型、多模态大模型、医疗具身智能、遥感智能解译。E-mail:jingzhang.cv@whu.edu.cn

张良培,男,教授,博士生导师,主要研究方向为高光谱遥感、计算智能及其在遥感影像处理中的应用、高分辨率影像处理。E-mail:zlp62@whu.edu.cn

© 中国图象图形学报版权所有