E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035
中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS

中图法分类号:TP751; TP391　　**文献标识码**: A　　**文章编号**: 1006-8961(2026)01-0212-31

论文引用格式:Chen S Q, Yang X, Zhu R Q Liao N, and Zhao W W. 2026. Parameter-efficient fine-tuning for remote sensing image interpretation: a survey. Journal of Image and Graphics, 31(1):0212-0242(陈诗琪, 杨学, 朱荣强, 廖宁, 赵卫伟. 2026. 面向遥感图像解译的参数高效微调研究综述. 中国图象图形学报, 31(1):0212-0242)[DOI:10.11834/jig.250105]

面向遥感图像解译的参数高效微调研究综述

陈诗琪[1],杨学[2],朱荣强[1*],廖宁[3],赵卫伟[1]

1. 信息支援部队工程大学,武汉　430030; 2. 上海交通大学自动化与感知学院,上海　200240;
3. 上海交通大学计算机学院,上海　200240

摘　要: 海量遥感数据的获取和AI大模型的发展极大程度地推动了智能化遥感图像解译的下游应用落地。“预训练+微调”是视觉语言基础大模型适配下游领域的经典范式,能有效将基础模型的知识迁移至新任务中。尽管遥感大模型发展如火如荼且在下游任务中表现突出,扩展的模型规模和高昂的训练成本使其难以适用于资源受限、标签不足、需求动态的实际应用场景。为使模型快速适应特定下游任务且有效避免额外训练资源消耗,参数高效微调方法得以广泛研究,并逐渐应用于遥感图像解译当中,成为当下的研究热点。本文面向不同类型的参数高效微调方法和解译任务,对提示词微调、适配器微调和低秩自适应微调三大类方法展开调研并梳理了现有研究工作。此外,本文收集归纳并总结了多个代表性数据集上30余种用于遥感图像解译任务的参数高效微调方法的性能,并从模型精度、训练参数量和推理耗时角度综合评估了方法性能,有助于启发研究者提出新方法并进行公平比较。最后,本文结合当前现状从多模态生成式任务、模型可解释性、边缘端部署应用的角度,展望并讨论了该交叉领域的未来研究方向,旨在为打造“AI+遥感”的下游应用生态提供理论参考与研究思路。

关键词: 视觉语言大模型;参数高效微调(PEFT);遥感图像解译;提示词;适配器;低秩自适应

Parameter-efficient fine-tuning for remote sensing image interpretation: a survey

Chen Shiqi[1], Yang Xue[2], Zhu Rongqiang[1*], Liao Ning[3], Zhao Weiwei[1]

1. Information Support Force Engineering University, Wuhan　430030, China;
2. School of Automation and Intelligent Sensing, Shanghai Jiao Tong University, Shanghai　200240, China;
3. School of Computer Science, Shanghai Jiao Tong University, Shanghai　200240, China

Abstract: Vision language model (VLM) currently plays a fundamental role in the research field located at the fusion of computer vision and natural language processing, demonstrating its application potential in various downstream vision tasks. With the continuous development of remote sensing data acquisition and processing technologies, the substantial amount of accumulated remote sensing images offers a solid data support for downstream remote sensing task. Remote sensing information extraction, fusion, and intelligent interpretation are crucial in various fields, such as high-resolution earth observation, land resources and energy exploration, environmental investigation, and military reconnaissance. Imperatively, using the generalization, versatility, and high-precision advantages of AI models can elevate the intelligence and

收稿日期:2025-04-23;**修回日期**:2025-05-25;**预印本日期**:2025-06-01

*通信作者:朱荣强　zhurq91@163.com

基金项目:国家自然科学基金项目(62401581,62401574)

Supported by:National Natural Science Foundation of China(62401581,62401574)

automation level of remote sensing image interpretation. However, the performance of the state-of-the-art VLMs is highly related to its expanded billions or even trillions training parameters, the conventional pretraining and full fine-tuning from scratch paradigm become untenable due to its excessive computational costs and high storage requirements. As a cost-effective tuning method, the recent emergence of new paradigms based on parameter-efficient fine-tuning (PEFT) offers a practical solution through the adaptation of large VLM models to specialized domain or task without introducing additional parameters. A meaningful research direction focuses on reducing the cost of fine-tuning the foundational models for downstream remote sensing tasks while still maintaining excellent performance. The constructed interpretation models should possess efficient training and downstream task adaptation capabilities under low-resource conditions to address the increasingly refined application demands and reduce the computational burden of full-parameter fine-tuning. Existing reviews mainly concentrate on the basic large vision-language models in remote sensing and provide limited attention on the parameter-efficient fine-tuning of remote sensing pre-trained models, thereby neglecting the systematic summary of the corresponding mainstream methods. With the rapid development of intelligent interpretation models for remote sensing images, systematically elaborating and analyzing parameter-efficient fine-tuning methods for remote sensing image interpretation is of considerable importance to fully leverage the dominant advantages of remote sensing large models, lower the barrier for applying remote sensing large models, and promote low-cost training, low-latency inference, and lightweight deployment applications of large models. Aiming to mitigate this gap, this survey systematically analyzes and reveals the potential of PEFT in remote sensing image interpretation to surpass the performance of full fine-tuning through minimal parameter modifications. First, the formal definition of PEFT is introduced, and its mainstream algorithms are explored. Prompt tuning, adapter tuning, and low-rank adaption tuning are the most prevalently used PEFT methods in the field of visual domain. Therefore, this survey divides existing methods into three categories and meticulously analyzes the advancements in their applications. Furthermore, this survey discusses some basic theoretical knowledge on influential large pre-trained vision and vision-language models. The survey also analyzes and summarizes the challenges of remote sensing downstream applications in terms of parameter-efficient fine-tuning methods from the perspective of dense prediction tasks, feature utilization, and training efficiency. A comprehensive investigation of relevant PEFT methods for diverse interpretation tasks, such as scene recognition, object classification, semantic segmentation, change detection, and object detection, is then provided. The contribution of each method, as well as the connections, similarities and differences among various methods on each downstream task, are comprehensively summarized and analyzed. Furthermore, existing remote sensing datasets suitable for each downstream task are collected and organized. All the datasets are provided with detailed information such as category number, data quantity, image size, and corresponding reference. Particularly, the prevailing evaluation criterion for each task is also provided for fair comparison. The collection, organization, and analysis of experimental results are performed considering datasets such as UC-Merced, AID, NWPU-RESISC45, UCAS-AOD, FAIR1M, LoveDA, and LEVIR-CD. Experimental results on several typical remote sensing datasets validate the superior performance of PEFT methods compared to full fine-tuning methods in terms of training parameters and memory cost. Building upon an extensive survey of the current state-of-the-art PEFT methods in remote sensing field, the related challenges are also discussed, and development trends are suggested for further research. Future research should include the following: 1) PEFT methods for generative tasks and multimodal data: PEFT techniques are conducive to empowering downstream applications of remote sensing generative foundation models that integrate multimodal content such as text, image, audio, and video. Exploring the application of PEFT methods in generative tasks within the multimodal remote sensing domain holds notable potential. 2) Research on the interpretability of PEFT methods: aiming to enhance the decision-making and cognitive capabilities of remote sensing large models, integrating prior information or multisource knowledge is necessary to guide the models. The robustness and trustworthiness of models can be largely improved by revealing the internal mechanisms of how models process multimodal information. 3) Research on the deployment and application of PEFT methods: designing large-scale foundational model architectures for efficient computation, combined with breakthrough in technologies such as model pruning, quantization, and inference acceleration, can help ultimately achieve low-cost training, low memory consumption, efficient real-time inference, and lightweight deployment applications for remote sensing large models. The authors hope that this survey can help establish a methodological framework for benchmarking PEFT algorithms in remote

sensing image interpretation, providing researchers in this field with detailed and valuable perceptions for further investigation.

Key words: vision-language large model; parameter-efficient fine-tuning (PEFT); remote sensing image interpretation; prompt; adapter; low rank adaptation

0 引 言

遥感技术能够远距离且高精度地探测地表和大气层中的信息,拍摄得到的遥感影像一直是自然资源调查、监测和管理的重要数据源。遥感图像解译指通过对遥感图像中目标特征进行建模、分析和推理,进而获取关于图像中地物类型、空间分布以及特征属性等信息的过程,是遥感技术应用的关键技术环节,能为城市规划、环境监测、灾害预警与评估、战场态势感知以及军事情报侦察等领域提供决策支持(郑向涛 等,2024)。传统的遥感图像解译以目视解译方法为主,需要由专业解译者根据其认知程度对目标进行探测、识别和鉴定,该解译方法具有直观、灵活且适应性强等优点,但主观性较强且解译效率低下。

随着对地观测技术的不断发展,遥感影像数据呈现爆炸式增长,可构建的海量优质样本数据集不断涌现。人工智能(artificial intelligence, AI)技术的兴起为遥感影像的高效处理与特征提取提供了有效契机。进入AI时代以来,随着算力的不断突破和算法的日益成熟,对海量遥感影像数据的深度分析和学习能提高数据处理的效率和解译的准确性。为此,亟须充分结合现有数据资源和深度学习先进技术打造“AI + 遥感”应用范式,提升遥感图像解译的智能化和自动化水平,推动人工智能与遥感解译的应用广度与深度。

随着自然语言大模型、视觉基础大模型和多模态视觉语言大模型的涌现和发展,人工智能预训练大模型(简称AI大模型)正逐渐落地应用。由于其具有较好的通用性、泛化性和实用性,训练后模型精度高,可以适配各类下游任务。面向场景识别、地物分类、变化检测和目标检测等典型遥感图像解译任务,国内外商业遥感公司、高校和科研院所已陆续推出如SkySense、空天—灵眸、秦岭—西电遥感脑等遥感大模型平台。基于海量数据的“遥感基础模型 + 下游任务应用”模式在遥感领域备受关注,已有综述文献系统阐述并总结了遥感基础大模型的结构设计或预训练方法。

然而,预训练大模型前所未有的规模也伴随着高昂的计算成本,对包含上亿参数的模型进行训练需要耗费大量计算资源。由于下游任务之间存在差异,针对不同任务或场景需要构建任务或场景特定的解译模型,在任务特定数据集上需要进行有监督微调以适应特定任务的需求,该方式面临解译效率低下和多任务泛化应用困难的弊病。其次,在资源受限的硬件平台上定制并部署下游模型时也面临模型规模庞大和算力紧缺的挑战,对于不同的数据集和任务还需要维护独立的模型权重,随着实际应用中任务数量增加,存储和计算资源难以覆盖现实的迫切需求,以上问题极大限制了遥感图像智能解译模型的大规模多任务应用。受限于大模型的规模,为实现有限计算资源下预训练大模型到特定任务的匹配,轻量化的大模型迁移学习技术,即参数高效微调(parameter-efficient fine-tuning, PEFT)(Houlsby 等,2019;He 等,2022a)应运而生。目前,对大模型进行微调时普遍采用“调整有限数量参数,同时固定其余参数不变”的策略,该技术已广泛应用于自然语言处理、计算机视觉以及视觉语言跨模态领域的微调,对于推动大模型在垂直领域的应用具有重要意义。

鉴于遥感领域大模型的构建亟须专业知识,许多学者借鉴视觉、语言以及视觉语言联合大模型的思路发展遥感大模型,通过微调现有开源视觉大模型、构建预训练大模型或构建多模态遥感大模型形式,研究高质量多模态数据构建、自监督式预训练以及标注数据微调等技术,并提出许多关于遥感领域基础大模型的综述(张永军 等,2024;张帅豪和潘志刚,2025),以适应不同模态、不同时相、不同场景和不同要素的智能解译需求。为适应日益细化的业务需求并降低全参数微调的算力负担,构建的解译模型应在低资源条件下具备高效训练和下游任务适配能力。然而,现有综述主要集中在遥感基础大模型上,对遥感预训练模型的参数高效微调关注有限,未能成体系地总结相应方法。在遥感图像智能解译模型迅猛发展的黄金时期,系统阐述并分析面向遥感图像解译的参数高效微调方法,对于充分发挥遥感大

模型显著优势，降低遥感大模型使用门槛并推动大模型的低成本训练、低延迟推理和轻量化部署应用具有指示性意义。

鉴于PEFT和“AI + 遥感”大模型的广阔发展前景，本文深入调研了参数高效微调方法在遥感视觉领域的应用，从参数高效微调方法的分类、主要方法详解以及参数高效微调技术在遥感图像解译的下游应用等层面梳理和分析现有工作，同时展望了基于参数高效微调的遥感图像智能解译应用未来的发展方向。

1 微调方法概述

1.1 参数高效微调主要思想

随着深度学习技术的发展和计算机性能的提升，预训练模型参数量呈现快速增长趋势，虽然模型参数量的增长会带来一定程度的性能增益但也存在诸多隐患。一是模型参数量的激增会导致模型微调成本的激增，传统预训练模型中使用的两阶段学习法难以适用，使用海量数据进行微调也无法充分利用微调样本信息；二是传统微调方法需要针对不同下游任务展开，需要存储关于任务的模型权重，极大消耗存储资源，且进一步训练优化模型整体参数对算力也提出极高要求。相较于训练所有网络参数的全参数微调方法，参数高效微调方法能通过在相应数据集上对小规模参数进行对应下游任务的训练优化，达到减少资源消耗并精细化适配下游子任务的目的。PEFT方法最早源于自然语言处理中的大语言模型，随后逐渐运用于计算机视觉领域，包括视觉Transformer结构、扩散模型等，该方法能有效降低模型微调的计算成本和大模型的使用门槛，拓宽了大模型在各类垂直领域的应用范围。本节将依次综述每类高效参数微调方法的现状，并重点针对常用PEFT方法总结分析优缺点。

1.2 主要方法及分类

当前主流的分类方法将参数高效微调分为附加式、指定式和重参数化式三大类方法。图1展示了不同PEFT方法的示意图。附加式微调将小规模的网络模块或可学习参数插入整体模型，并通过对少部分参数进行微调实现下游任务的高效适配；指定式微调既不引入新参数也不改变网络原始结构，直接对与任务相关的指定参数进行优化；重参数化微调方法首先增加参数进行微调，然后将新增参数融合至现有模型。

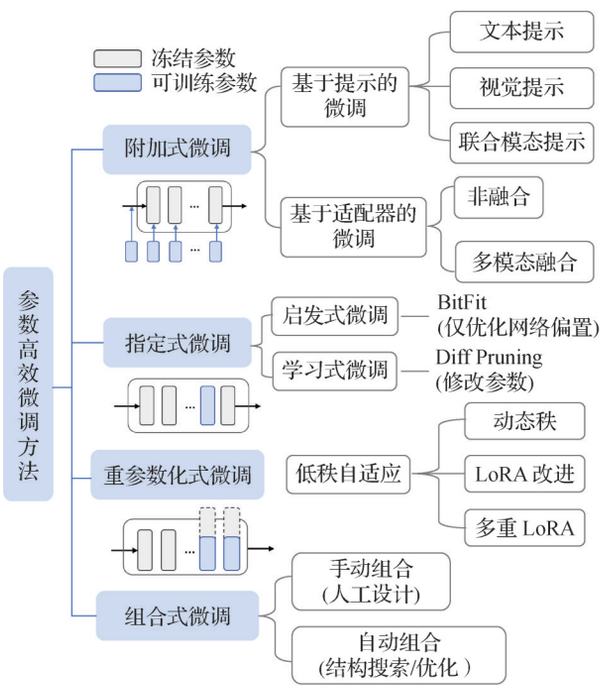


图1 参数高效微调方法结构与分类示意图

Fig. 1 Schematic diagram of the structure and classification of parameter-efficient fine-tuning methods

为选取应用于遥感领域下游任务时最具代表性的几类方法，本文主要关注提示微调、适配器微调和低秩自适应微调这三小类方法。下面针对目前常用的3类PEFT方法进行详细阐述。

1.2.1 提示微调

基于提示的微调最初用于使大模型适配特定的下游自然语言处理任务。在计算机视觉领域中，提示微调方法通过模板将下游任务重构为与模型预训练相近的形式，对原始输入进行包装以指导模型生成符合特定任务需求的输出，降低了微调的存储和运算资源消耗。提示主要分为硬提示和软提示，前者也称为离散提示，其强烈依赖于人工针对具体任务手动设计。而软提示将提示特征向量的生成作为任务进行参数化学习，具有显式语义信息且能在连续空间内参与训练优化(Xing等，2024)。上述向输入序列添加可调整向量的过程可表示为

$$\boldsymbol{X}^{(l)} = [s_1^{(l)}, \cdots, s_{N_s}^{(l)}, x_1^{(l)}, \cdots, x_{N_x}^{(l)}] \tag{1}$$

式中，$\boldsymbol{X}^{(l)}$表示第l层经过提示调整后的输出序列，包括软提示$\text{tokens}_i^{(l)}$和原始输入$\text{token}x_i^{(l)}$。

Lester等人(2021)提出提示词微调方法(prompt

tuning)，仅在预训练模型输入的嵌入向量前拼接连续可学习的张量$\boldsymbol{P} \in \mathbf{R}^{l \times h}$以适配下游任务。随着模型容量增加，该方法的参数节省比例增高(Su等，2022)。Li和Liang(2021)提出基于附加式的微调方法，即前缀调优prefix-tuning，在预训练模型的每一个多头注意力层(multi-head attention，MHA)中构造连续且与任务相关的虚拟词元前缀，但前缀微调法不仅将前缀序列拼接到模型输入层向量，还拼接到模型中所有隐层的前端。对于提示微调方法而言，训练过程中保持预训练模型参数不变，仅优化特定任务的prefix。具体将构建两组随机初始化的前缀矩阵$\boldsymbol{P}_k \in \mathbf{R}^{l \times d}$和$\boldsymbol{P}_v \in \mathbf{R}^{l \times d}$置于原始输入序列进行拼接，再通过3个矩阵变换得到查询Query，键Key和值Value的取值。

由于直接优化$\boldsymbol{P}$容易导致模型训练不稳定，可考虑先对$\boldsymbol{P}$进行变换然后拼接至嵌入向量或隐层前端。

图2展示了提示微调法(prompt tuning，Pro-T)和前缀微调法(prefix-tuning，Pre-T)两者的示意图，前缀微调属于提示微调方法的一种，两种形式均采用与任务相关的词元对提示初始化，训练时在下游任务上优化可学习张量，由于仅在输入端嵌入提示，Pro-T相比Pre-T添加的参数量更少。

视觉提示微调方法(visual prompt tuning，VPT)通过在Transformer的输入层或者每层前引入有限数量的可训练参数即提示，引导预训练视觉模型执行各项下游任务(Jia等，2022)。考虑到VPT运用于视觉任务时缺乏上下文关联性，改进视觉提示微调方法(Yoo等，2023)在不同层的提示标记之间添加连接和交互关系，还有部分工作侧重于通过设计子网络产生视觉提示。Nie等人(2024)设计了轻量级提示模块以完成预训练提示调整方法。Wang等人(2024a)向预训练模型的首尾分别添加两个隐层以生成视觉提示。Zhu等人(2023)提出的视觉提示多模态跟踪方法同时将RGB和辅助模态图像作为网络输入，然后用图像块嵌入模块处理以生成相应的提示标记。视觉查询微调(visual query-tuning，VQT)(Tu等，2023)仅将提示模块与模型的中间查询特征进行交互，通过添加可学习的查询标记来聚合Transformer模型中各层的Query表征。有效且高效的视觉提示微调方法(Han等，2023)则在视觉提示上附加键Key和值Value的约束，在Transformer中自注意力层和输入层的提示中引入额外的可学习键和值提示。由于单模态提示方法存在局限性，Zang等人(2022)提出统一提示微调方法，拟结合不同模态优势实现性能的平衡。

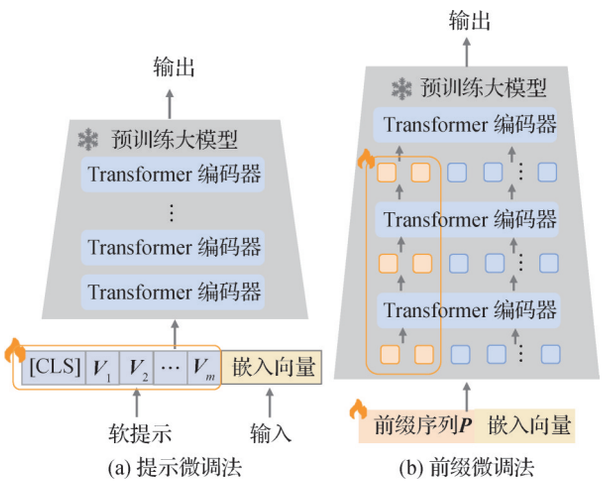


图2 提示微调法与前缀微调法示意图(Xing等，2024)

Fig. 2 Schematic diagram of prompt tuning method and prefix-tuning method (Xing et al., 2024) ((a) prompt tuning; (b) prefix-tuning)

1.2.2 适配器微调

适配器(adapter)作为PEFT的一种典型技术，其通过向基于Transformer的预训练模型中添加可学习的网络模块，即适配器以适应下游任务(Houlsby等，2019)。适配器的主要思想是在已冻结的预训练模型中引入额外可训练参数，在执行下游任务时仅微调适配器的参数而无需对预训练参数进行全模型调整，由此完成对特定任务知识的学习并且避免灾难性遗忘。

图3展示了适配器与预训练模型融合的方式以及适配器内部的主要结构。通常Adapter模块的插入位置为前馈层后，每个Adapter模块包括两个前馈网络(向下和向上投影)、一个非线性层和一个残差连接。传统Adapter采用简单的下采样—上采样瓶颈结构，还有其余适配器采用下采样—二维卷积—上采样结构。由于适配器模块中包含参数较少，且仅更新Adapter模块权重，极大降低了算力要求。

图4从模块嵌入位置的角度展示了适配器与预训练模型共同作用的方式。如图4(b)所示，顺序式适配器(Houlsby等，2019)通过在Transformer模块中插入两个轻量级适配器层来引入额外的可训练参数，插入位置在自注意力层和前馈神经网络层之后。

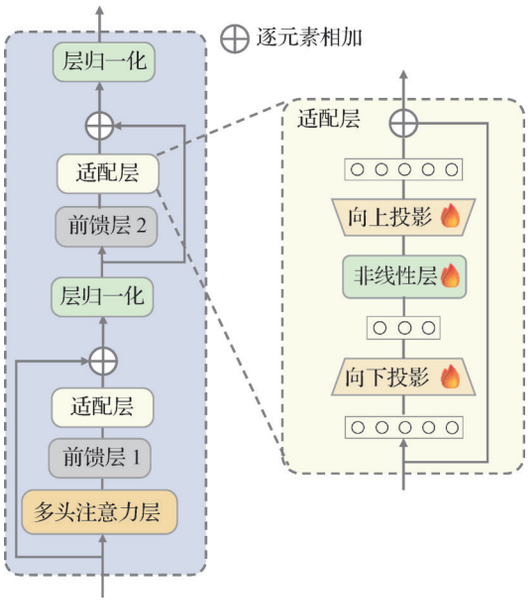


图 3　适配器法结构示意图（Houlsby 等，2019）

Fig. 3　Schematic diagram of the adapter structure (Houlsby et al. , 2019)

残差适配器（Lin 等，2020）在前馈和标准化操作之后插入适配器进一步提高效率。卷积适配器（Chen 等，2024c）是专门为卷积网络设计的适配器模块，仅针对骨干网络的中间特征进行适配。

不同于顺序式适配器，如图 4（c）所示的并行式适配器（He 等，2022a）提出适配器并行侧网络结构，激活层在模块层与适配子层之间并行传递。这类适配器方法与条件独立注意力转移（Zhu 等，2021）和 KronA 适配器（Edalati 等，2022）原理类似，通过给原始模型架构添加额外的层进行高效微调，部分结构还能进行多任务学习（Wang 等，2022a；Chronopoulou 等，2023）。Chen 等人（2022a）提出的 ViT-Adapter 将视觉变换器（vision Transformer，ViT）作为整体框架的骨干网络，通过并行插入适配器模块引入与输入图像相关的归纳偏置，提升模型在下游任务上的性能。AdaptFormer（Chen 等，2022b）用 AdaptMLP 代替 Transformer 结构中的多层感知器模块，包括两个并行的冻结分支和可训练分支，在视觉任务上使用并行适配器结构学习速度更快且效果更好。

混合式适配器（adapter fusion）（Pfeiffer 等，2021）通过在“相加与归一化”层后集成多个任务的适配器层，实现跨任务的知识传输并提高计算效率。该方法能根据给定输入为不同任务对应的适配器分配不同权重，有效缓解了灾难性遗忘、不同任务间干扰和训练不稳定的问题，但应用中由于需要在组合层中添加额外参数，计算成本较高。

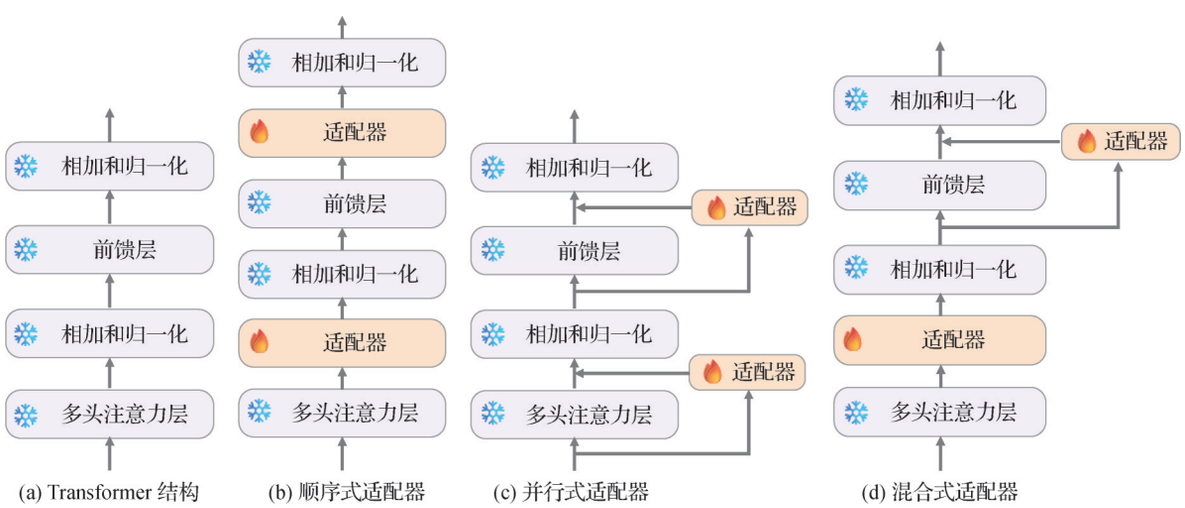


图 4　适配器的几种变体（Feng 等，2024）

Fig. 4　Several variants of the adapter structure (Feng et al. , 2024)

((a) original transformer block; (b) sequential adapter; (c) parallel adapter; (d) mixed adapter)

1.2.3　低秩自适应微调

低秩自适应（low-rank adaptation，LoRA）（Hu 等，2021）是参数高效微调中最常见的重参数化技术，通过矩阵乘法更新现有参数，由于不存在额外参数，不会增加推理延迟，模型的灵活性也能得到有效保证。研究人员认为预训练模型权重中的参数矩阵通常满秩（Li 等，2018a），可通过限制矩阵维度以减少参数冗余，降低微调过程中需要更新的参数量。

LoRA 提出针对自注意力模块中原始权重矩阵的低秩分解，通过搜索低秩矩阵使其既能保留原始

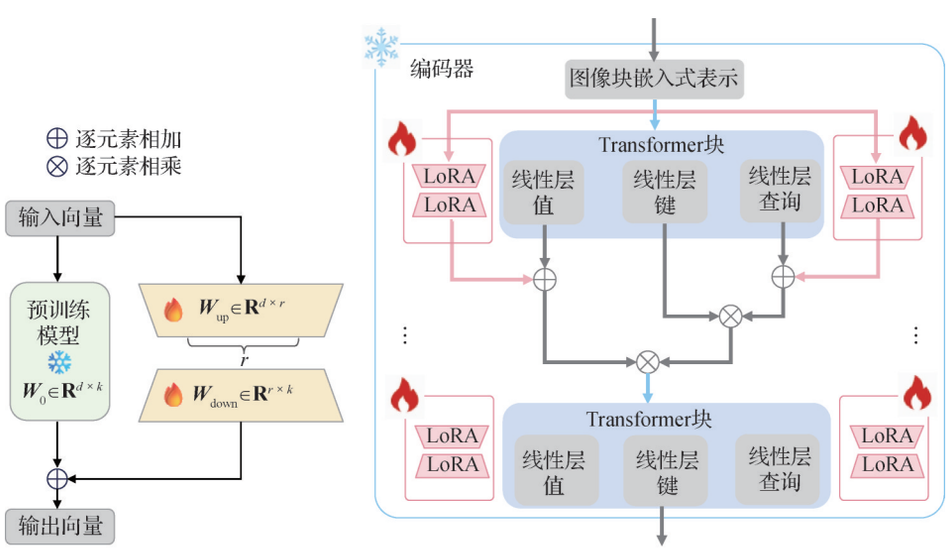
信息，同时通过更新的增量权重降低计算复杂性和内存使用。对于预训练模型的权重矩阵 $\boldsymbol{W}_0 \in \mathbf{R}^{d\times k}$，LoRA使用低秩分解 $\Delta\boldsymbol{W} = \boldsymbol{W}_{\text{down}}\boldsymbol{W}_{\text{up}}$，其中，$\boldsymbol{W}_{\text{down}} \in \mathbf{R}^{d\times r}$，$\boldsymbol{W}_{\text{up}} \in \mathbf{R}^{r\times k}$，$r \ll \min(d,k)$。

图5(a)为预训练模型的侧旁分支中引入低秩自适应结构。如图5(a)所示，训练时仅对低秩分解后的矩阵进行梯度更新，推理时将LoRA权重与原始权重自动合并。使用LoRA微调后，可学习网络参数由 $d\times k$ 变为 $r\times(d+k)$。如图5(b)所示，由于LoRA侧重于微调参数权重的查询和值矩阵，主要用于ViT骨干网络微调，在分割大模型(Kirillov等，2023)和扩散模型(Ho等，2020)等模型中应用广泛。

LongLoRA(Chen等，2024d)提出可转换的移位稀疏注意力机制，降低运算量并缩短训练时间。GLoRA(Chavan等，2023)引入门控机制动态扩展并逐层搜索各层适配器，模型灵活性和适应性得到大幅提升。自适应低秩自适应AdaLoRA(Zhang等，2023b)根据重要性程度为权重矩阵动态分配参数，利用奇异值分解获取权重矩阵进行微调。DyLoRA(Li等，2020b)训练出具有不同秩的LoRA模块以降低训练时间成本。除此之外，多项研究专注于通过诸如拉普拉斯近似(Yang等，2024a)或其余新颖的训练策略(Hayou等，2024；Huang等，2024)等多种方法提升LoRA的性能。Liu等人(2024)通过将模型权重分解为大小和方向来设计权重分解低秩自适应方法。

表1总结了3种主流PEFT方法的优缺点及代表性文献。总结来看，提示微调灵活性和适应性强，但需要根据特定任务调整；适配器微调轻量化且易于扩展，但可能存在推理时延问题；低秩自适应兼容性较强，但优化秩的值依赖搜索。

除上述3类主流方法外，研究人员还探索了PEFT策略的各种组合以增强微调效果，包括手动组合类别中结合低秩融合机制的适配器方法(Yin等，2023)、紧凑适配(Karimi等，2021)，以及自动组合类别中融合VPT、LoRA和Adapter等微调方法的神经元提示搜索(Zhang等，2022)、自动配置搜索方法(Zhou等，2024)和以门控机制激活不同微调方式的统一参数高效微调(Mao等，2022)等。



(a) 低秩自适应结构示意图　　(b) LoRA结构在Transformer中的应用方式

图5　低秩自适应微调方法示意图及其用法

Fig. 5　Schematic diagram of low-rank adaptation fine-tuning method and its application

((a) low-rank adaptation structure; (b) application paradigm of LoRA structure in Transformer)

2　基础模型回顾

遥感解译任务的应用离不开基础模型的选择与设计，在海量遥感图像样本上训练的大规模预训练模型能为不同领域的下游任务提供鲁棒的模型初始化参数。本节将重点对遥感领域应用中的基础模型进行阐述，主要包括视觉预训练模型、多模态预训练模型和分割大模型的基本组成及其使用。

2.1　Transformer结构和ViT

Transformer模型主要由编码器(encoder)和解码器(decoder)两部分组成，模型中的多头注意力机制

表 1　典型参数高效微调方法对比与总结

Table 1　Comparison and summary of typical parameter-efficient fine-tuning methods

类型	代表性方法	核心思想	优点	缺点
提示微调	Prefix-tuning(Li和Liang,2021)；P-Tuning v2(Liu等,2022)；VPT(Jia等,2022)；GateVPT(Yoo等,2023)；VQT(Tu等,2023)；E²VPT(Han等,2023)；UPT(Zang等,2022)	根据特定任务,在连续空间中优化生成适配下游任务的软提示,将其串接模型输入或隐藏状态中,可通过梯度反向传播优化。	不依赖人类手动设计,灵活性和适应性强,可根据任务和数据优化调整。	不同任务所需提示长度、深度、位置差异较大,需重新调整提示策略。
适配器微调	Sequential Adapter(Houlsby等,2019)；Residual Adapter(Lin等,2020)；Conv-Adapter(Chen等,2024c)；ViT Adapter(Chen等,2022a)；AdaptFormer(Chen等,2022b)；Adapter Fusion(Pfeiffer等,2021)	向预训练模型内部添加可学习网络模块即适配器,保持模型其他参数不变。	轻量化,参数量少,灵活性强,收敛速度快。	存在推理时延,数据量少时易过拟合,领域差异较大时效果不好。
低秩自适应微调	LoRA(Hu等,2021)；LongLoRA(Chen等,2024d)；GLoRA(Chavan等,2023)；AdaLoRA(Zhang等,2023b)；DyLoRA(Li等,2020b)；Mona(Yin等,2024)；FLoRA(Si等,2025)	针对自注意力模块中原始权重矩阵进行低秩分解,通过重参数化更新的增量权重降低计算量。	不会增加推理时间,训练门槛低,兼容性强。	优化秩的值需要大量搜索及策略设计。

可将输入分成多个头,在不同头进行独立注意力计算,再对结果进行拼接并投影到输出空间,有利于增强模型对复杂上下文关系的捕捉能力。为有效捕捉输入图像中的远程依赖关系,Dosovitskiy等人(2021)提出基于自注意力机制的视觉Transformer模型,引入视觉Transformer以完成图像识别任务。其将图像划分为小块,并使用Transformer的自注意力机制建模全局上下文信息。ViT在ImageNet分类任务上表现出色,为视觉领域引入了Transformer的思想(Wang等,2023)。ViT的关键组件包括补丁嵌入层、位置编码器、Transformer编码器以及分类头。Transformer中的自注意力机制允许每个标记关注所有其他标记,有助于网络捕捉全局上下文信息。注意力权重的计算方式为

$$Attention(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = softmax(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}})\boldsymbol{V} \tag{2}$$

式中,$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$表示输入矩阵,d_k表示键向量的维度,*softmax*表示激活函数。视觉Transformer凭借其强大的捕捉局部和全局模式的能力,在诸如土地覆盖分类、场景识别以及植被分析等遥感解译任务中展现出巨大应用前景(Scheibenreif等,2022)。

2.2　CLIP及其变体

对比图文预训练模型(contrastive language image pre-training,CLIP)(Radford等,2021)作为一种简单而有效的预训练模型,其联合了视觉编码器和文本编码器学习视觉表示,通过拉近图文表征解决了两个模态对齐的问题,具有极强的鲁棒性和泛化能力。图6展示了CLIP模型的结构示意图,其通常采用手工制作的提示词,例如“A photo of a [ ]”的形式生成文本嵌入表示。

CLIP的学习流程如下:1)通过预训练对比模型,将文本和图像送入对应编码器中提取特征向量,在多模态空间中计算文本和图像特征的相似度使其学习到相应图文特征;2)通过标签文本建立数据集分类器,将类别对应的描述转化为文本编码向量,描述的形式通常为“A photo of a [ ]”;3)对待分类图像进行编码,计算图像向量和文本向量的评分,根据得分判定待测图像对应类别。

在训练CLIP模型时,对比预训练在对齐图文对中起到重要作用。对于批次大小为N的标签文本和图像构建N^2个图像文本对,选定N个匹配的图像文本对为正样本对,余下N^2-N个未匹配的图像文本对为负样本对。训练时采用InfoNCE损失函数(He

等，2020）对网络进行优化，其损失函数可表达为

$$L=-\frac{1}{N}\sum_{i=1}^{N}\log\frac{\exp\left(\frac{\Phi(\boldsymbol{I}_i,\boldsymbol{T}_i)}{\tau}\right)}{\sum_{j=1}^{N}\exp\left(\frac{\Phi(\boldsymbol{I}_i,\boldsymbol{T}_j)}{\tau}\right)} \tag{3}$$

式中，$\Phi(\cdot)$表示两个向量之间的余弦相似度计算模块，$\boldsymbol{I}_i$和$\boldsymbol{T}_i$分别表示第i个图像和对应文本的嵌入向量，exp表示指数函数，τ是用于缩放余弦相似度值的温度参数。

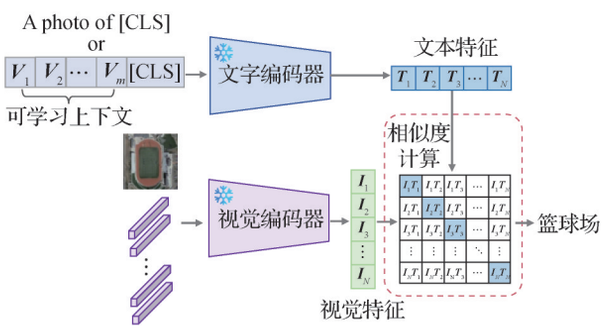


图6　CLIP模型结构（Radford等，2021）

Fig. 6　CLIP model structure (Radford et al., 2021)

作为预训练视觉语言模型（vision language model，VLM）的文本输入，提示词能从VLM内部中提取特定任务信息以引导下游任务。不同于直接以手工设计“A photo of a [CLS]”作为提示词的方法，研究人员以CLIP为原型针对提示进行优化，提出众多改进方法以提高模型在特定任务中的效率和准确性。上下文优化（context optimization，CoOp）（Zhou等，2022a）将可学习的连续向量作为提示，具体而言，可学习的M个上下文向量由与词嵌入具有相同维度的$\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_M$向量集合表示，$M$是决定上下文长度的超参数，通过反向传播机制对其进行优化。对于给定类别y，提示可表示为$\boldsymbol{t}_y=\{[\boldsymbol{v}_1],[\boldsymbol{v}_2],\cdots,[\boldsymbol{v}_M],[CLS_y]\}$。虽然CoOp方法在少样本分类问题上表现出显著改进，但易受到领域偏移问题影响，条件化上下文优化方法（conditional context optimization，CoCoOp）（Zhou等，2022b）提出以视觉特征为条件的提示学习策略，通过引入元网络生成M个条件标记。知识引导的上下文优化提示方法（Yao等，2023）通过最小化提示间差异增强泛化能力，而协同提示词方法（Roy和Etemad，2024）对训练模型进行约束以避免下游任务上的过拟合现象。概念提示学习方法（Zhang等，2024c）旨在通过建立视觉概念缓存来捕捉多级视觉特征并生成针对性提示。

2.3　分割大模型

Meta AI Research提出的SAM（segment anything model）（Kirillov等，2023）是视觉基础大模型，具有强大的零样本泛化能力。如图7所示，SAM采用基于视觉Transformer的预训练掩码自编码器（He等，2022b）将图像处理为中间特征，并将先验提示编码为嵌入式标记。然后，用掩码解码器的交叉注意力机制对图像特征和提示嵌入进行交互，最终生成掩码输出。

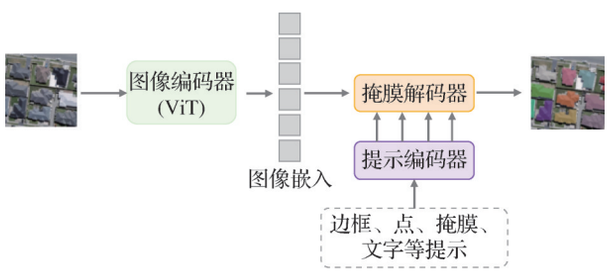


图7　SAM结构示意图（Kirillov等，2023）

Fig. 7　Schematic diagram of segment anything model (SAM) (Kirillov et al., 2023)

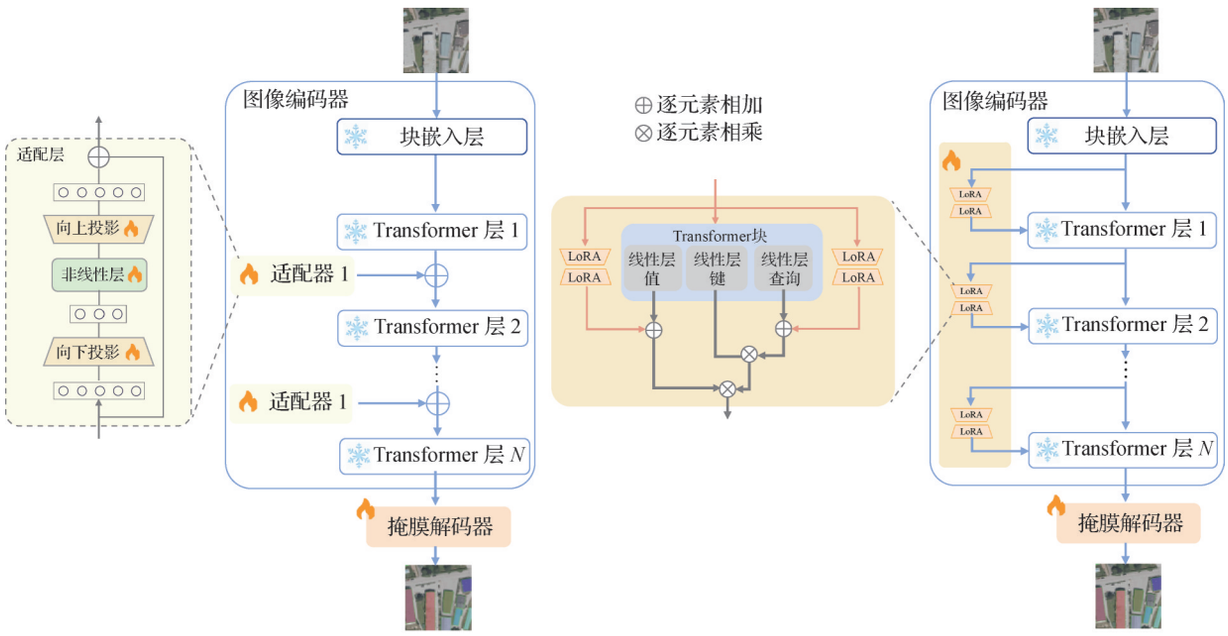
由于SAM基础模型是基于通用图像训练而成且不针对特定任务设计，为克服SAM在提取遥感图像特征方面的局限性，同时保留SAM原有结构，SAM编码器常与适配器及低秩自适应模块组合，得到适用于遥感图像下游解译任务的参数高效微调方法SAM-Adapter和SAM-LoRA，具体使用方式如图8所示。

图8（a）展示了基于适配器微调的SAM结构（SAM-adapter）的示意图。适配器模块利用基于输入特征产生的提示引入额外的特定任务知识，如图像的纹理或频率信息。具体而言，第i层的提示输入$\boldsymbol{P}_i$经过两层多层感知器（multi-layer perceptron，MLP）处理以及一层激活函数处理后，被添加到Transformer结构的输出中，该过程可表示为

$$\boldsymbol{F}_{i+1}=\boldsymbol{O}_i+M_2(\sigma(M_1(\boldsymbol{P}_i))) \tag{4}$$

式中，$\boldsymbol{O}_i$表示第i层的输出特征，$\boldsymbol{F}_{i+1}$表示第$i+1$层的特征输入，M_1和M_2表示两个多层感知器，σ表示高斯误差线性单元激活函数。

图8（b）展示了基于低秩自适应的SAM结构（SAM-LoRA）的示意图。类似于SAM-Adapter，SAM-LoRA在两个连续的Transformer结构中引入额外层。不同之处在于，LoRA模块通过低秩全连接层实现跳



(a) 使用 Adapter 微调的 SAM 结构　　(b) 使用 LoRA 微调的 SAM 结构

图 8　SAM 与两种典型参数高效微调方法结合使用的示意图

Fig. 8　Model structure of SAM with adapter and LoRA modules

((a) SAM structure with adapter module; (b) SAM structure with LoRA module)

跃连接，而不会引入额外信息。LoRA 模块的架构及位置如图 8(b)左侧所示，LoRA 层通常应用于 Transformer 结构中的查询和值投影层。微调过程中，冻结所有 Transformer 的原始权重，仅对新添加低秩自适应模块中的全连接层进行微调。

随着 SAM 分割大模型的发展和模型应用部署的需求，衍生出多个基于 SAM 的轻量化大模型，旨在保持较高分割性能的同时降低计算成本和内存占用(Sun 等，2025)。FastSAM(Zhao 等，2023b)将分割任务分成全实例分割和提示引导的掩膜选择两个串行阶段，能较好地分割规则形状物体。MobileSAM(Zhang 等，2023a)对图像编码器和掩膜编码器进行解耦式优化，将 SAM 中图像编码器 ViT-Huge 蒸馏到轻量化的 ViT-Tiny 上，其推理速度是 SAM 的 60 倍且分割性能与 SAM 相当。EfficientSAM(Xiong 等，2024)利用掩码图像预训练 SAMI(SAM image)模型，然后依此构建高效 SAM，其在多个视觉任务上性能均比其余掩码预训练方法增益明显。

3　PEFT 方法的遥感图像解译应用

当前遥感大模型发展如火如荼，利用卷积神经网络(convolutional neural network，CNN)、循环神经网络、Transformer 结构以及注意力机制等成熟的深度学习技术，通过对海量遥感数据训练，能够实现对不同下游任务的高效应用。但现有的参数高效微调方法往往仅基于预训练模型的共享知识来适应下游任务，而忽略了遥感图像的独有特性。不同于自然光学图像，由于观测场景、传感器参数以及成像气候条件等差异，遥感影像呈现出多时相、多传感器、多分辨率以及多要素的特点，使得遥感大模型的构建与下游应用存在诸多挑战。

一方面，高质量大规模样本集构建较难。随着空天遥感数据信息获取与处理技术的不断发展，海量遥感图像样本数据为遥感下游业务提供了坚实的数据支撑。但这些数据的数量和质量仍难以匹敌计算机视觉领域的上亿图像数据集，训练对于下游任务具有鲁棒特征表达的模型所需高质量数据随任务需求提升不断增加，而传统人工标注耗时耗力且依赖专家知识，面临成本高效率低的难题。如何自动化构建高质量动态样本库，甚至实现单模态到多模态数据集的构建面临巨大挑战。尽管目前衍生出一系列诸如 RingMo(Sun 等，2023)、SkyScript(Wang 等，2024d)、RSICap(Hu 等，2023a)等多模态遥感数据集，但这些数据集通常针对专用任务设计，向多类下游任务迁移的能力有待进一步提高。

另一方面，当前遥感领域基础模型的构建和应用通常遵循“预训练 + 微调”的范式，预训练模型可直接采用开源视觉语言大模型或根据自监督学习在海量高质量多模态遥感数据上进行预训练构建，随后需要利用全参数微调来迁移预训练模型的视觉理解能力，从而在下游任务中取得良好的性能。上述微调机制存在两大弊端，一是次优化性能。当预训练数据和测试数据分布差异较大时，由于知识偏差难以保证较好的模型效果。二是资源消耗大。全参数微调需要为各下游任务存储预训练基础模型副本，再针对特定任务和应用场景定制化训练并更新模型参数，其规模等同于预训练基础模型。由于这些参数独立运行不可共享，导致在实际应用中会耗费大量资源。例如，遥感基础模型中常用的Swin-L模型参数体量为195 M，比常用的ResNet50（residual network）约大8.3倍，具体应用时训练内存消耗和推理时间分别大约大4.1倍和2.3倍。其次，随着下游任务的扩展和部署环境的持续变化，参数存储呈线性增长，会极度耗费存储空间。

回归到PEFT方法在特殊领域的应用中，从下游任务方面看，现有视觉高效参数微调方法主要关注图像分类任务，而在诸多强调多尺度特性的密集预测任务（如语义分割和目标检测）中少有涉及，如何将高效微调方法运用于基础模型并适配多个下游任务也至关重要。其次，从特征利用方面看，目前大多数PEFT方法以单独细化特征的方式进行微调，而忽略了特征之间的空间关联，对于上下文信息丰富的遥感图像而言，更需要加强目标空间上下文特征的感知能力。最后，从训练效率方面看，由于在遥感场景中并非总能获取足够量带标签的训练数据，因此在数据稀缺的应用场景下如何高效利用少量数据进行微调并使模型快速收敛也值得深入研究。

因此，针对遥感图像解译应用开展参数高效微调研究是人工智能和遥感对地观测交叉领域的研究重点。本节将分别阐述PEFT方法在不同解译任务上的研究进展。

3.1 场景识别

场景识别是指为图像中的特定场景赋予语义类别，根据图像内部地物的空间关系和分布模式识别出如城市、森林、水体和农业区域等预定义类别。该任务对于土地利用与覆盖图绘制、环境监测和城市规划等至关重要。针对场景识别任务的提示微调方法多集中于探索模型在小样本或开放场景下的能力，主要用于检验模型的域泛化能力，表2总结了不同应用条件下场景识别中PEFT方法的代表性工作。

为有效避免潜在的灾难性遗忘问题，Yang等人（2024b）在骨干网络层之间引入用于遥感场景分类的软自适应提示机制（soft adaption for rs scene classification，StARs）以控制提示标记的信息流转，该机制对浅层和深层提示采用不同的处理方式。现有方法通常对所有图像施加关于特定任务的固定提示，忽略了不同图像的各异性且微调后模型性能难以达到最优。考虑到遥感图像中地物分布复杂，不同类别场景可能出现相似特征，同类场景可能存在显著差异，Fang等人（2023）提出实例感知的视觉提示（instance-aware visual prompting，IVP）生成方法，其中实例级提示生成模块Meta-Net能有效聚合输入图像的上下文信息，通过动态提示生成校准预训练模型的特征表达。

在适配器微调应用方面，尹文昕等人（2024）提出多尺度融合适配器微调方法（multi-fusion adapter，MuFA），其通过融合不同下采样倍率的瓶颈模块实现纹理特征和语义特征的融合，从而有效提升地物分类精度。不同于适配器串联嵌入模型的形式，并联形式嵌入的独立分支能保留模型原始特征，且由于不增加层数减少了微调参数量。为了保留目标的关键特征信息且加快模型训练速度，该方法还设计了基于不同下采样率的多尺度融合策略对不同尺度特征进行加权融合。

Zhu等人（2024）提出的元学习视觉微调方法（meta visual prompt tuning，MVP）首次将提示微调和元学习机制引入少样本条件下遥感图像场景识别。在元学习训练和微调阶段，仅对新加入的提示词参数进行更新，减少了模型参数存储需求并减轻过拟合风险。此外，为缓解不同成像条件下模型泛化能力弱的问题，该方法还提出基于ViT模型的图像块随机重组策略。为适应真实开放世界的应用需求，遥感图像分类模型需要具备不断吸收新知识且保留优化旧知识的能力，即避免灾难性遗忘现象。为避免模型优化时特定知识间的互相干扰并缓解灾难性遗忘，Zhao等人（2023a）将场景识别相关知识解耦为共享知识和特定知识。具体而言，首先构建多阶段共享的提示池，然后采用键—值查询策略根据输入

表 2　面向场景识别的参数高效微调方法总结

Table 2　Summary of typical parameter-efficient fine-tuning methods toward scene recognition

方法	类型	应用场景	贡献	优点	缺点
StARs（Yang等，2024b）	Prompt微调（单模态）	一般	引入软适应机制逐层交互优化提示词信息流，防止灾难性遗忘	参数更新量少，知识保留能力强，防止灾难性遗忘	提示词生成灵活性不足，依赖固定初始化，缺乏动态适配能力
IVP（Fang等，2023）	Prompt微调（单模态）	一般	充分考虑输入图像上下文信息，提出实例感知的动态视觉提示生成方法	参数效率高，解决图像背景复杂、类内差异大的问题	模型性能提升依赖于输入序列长度，逐层生成提示导致输入序列长度增加，训练时间较长
尹文昕等人（2024）	Adapter微调	一般	融合不同下采样倍率的瓶颈模块实现纹理和语义特征融合	能保留模型原始特征，不增加层数	多尺度计算可能增加模型复杂度
MVP（Zhu等，2024）	Prompt微调（单模态）	少样本	基于ViT设计随机块重组数据增强方法，利用提示微调和元学习机制将预训练模型自适应到下游任务	元学习提升提示参数初始化效率，存储需求低	对预训练ViT模型依赖性强
Continual（Zhao等，2023a）	Prompt微调（单模态）	类增量泛化	利用前缀微调思想并对场景识别相关知识进行解耦式提示学习	知识解耦针对性强，动态提示选择机制能减少对历史数据的依赖	提示词查询机制复杂，提示池可扩展性有待提升，对提示长度和数量较为敏感，需要精细调参
APPLeNet（Singha等，2023）	Prompt微调（多模态）	类增量泛化	对视觉风格和图像原始内容进行解耦引入CLIP模型，提出基于视觉注意力参数化的提示学习网络	针对遥感图像小目标和多尺度特性进行优化，泛化能力强	多模态提示协同能力有限，未充分融入遥感领域知识
EPT（Lan等，2025）	Prompt微调（多模态）	类增量泛化	提出多层级多粒度提示设计增强语义表达能力，通过轻量化网络引入船舶领域先验知识	跨模态提示调优能提升基类到新类的泛化能力，多粒度提示能增强文本提示的语义表达	需要额外标注多粒度标签以辅助提示设计，数据准备成本高
MaPLe（Bi等，2023）	Prompt微调（多模态）	少样本	利用MaPLe方法同时对视觉和语言两个分支进行耦合学习，使模型更好地理解和关联不同模态的信息	参数效率较高，跨模态耦合提升领域适应性	未根本解决文本提示和遥感语义信息之间的差异鸿沟
FreqDiMFT（Zhang等，2024b）	Prompt微调（多模态）	类增量泛化	提出频率分布集成模块和自适应特征精炼模块	频域特征利用能有效解决遥感图像高类内多样性和低类间差异性问题	多模态提示微调依赖GPT生成专家描述，依赖大规模多模态预训练模型
ProDFSL（Ji等，2024）	Prompt微调（多模态）	少样本	结合ChatGPT生成语义描述，构建跨模态知识生成模块，增强原型表征	动态生成软提示，跨模态原型优化可增强类间区分度	多模态交互计算复杂度较高，依赖ChatGPT生成语义描述的质量

图像表征查找与类别相关的若干提示。然后，这些提示与输入图像嵌入式特征一同输入场景识别模型完成类别确定。

虽然上述部分方法采用动态且自适应的提示优化方法，由于未充分对齐跨模态特征存在视觉和语义的匹配偏差，其本质仍是仅调整文本提示向量或视觉提示的单模态提示方法。多模态提示如MaPLe方法（Khattak等，2023）能同时调整视觉和文本提示

参数，通过联合优化增强跨模态对齐和协同，使模型更好地理解和关联不同模态的信息并显著提升泛化能力。Singha等人（2023）提出视觉注意力参数化提示学习网络（visual attention parameterized prompts learning network，APPLeNet）对视觉风格和图像原始内容进行解耦，并通过视觉特征动态生成文本提示。具体而言，首先提取视觉内容特征和图像风格特征并用基于注意力机制的注入模块对两者进行结合，其中冻结的图像编码器负责提取多尺度视觉内容特征，域特定的风格特征由批归一化统计量表示。然后用轻量化网络处理上述特征，实现视觉信息对文本提示的动态调整。最后用图文监督对比损失和内容冗余惩罚损失共同优化模型，为避免提示信息引起的模型参数冗余，还提出反相关归一化器以提升特征判别能力。该方法中文本提示的生成直接依赖于视觉特征，无需显式调整视觉分支参数，属于隐式跨模态交互方法。Lan等人（2025）提出联合优化视觉与文本模态的方法，通过轻量化网络融合遥感船只先验信息并调整视觉特征，然后设计分层级多粒度文本提示并结合视觉特征生成动态上下文信息。该方法仅将两个模态提示间接对齐而并未显式共享参数，MaPLe方法分别在视觉和文本编码器中插入可学习提示，通过线性层共享参数对跨模态提示进行协同优化，属于显式多模态交互方法。受MaPLe方法启发，为克服手工制作提示在遇到特定领域任务或数据时性能欠佳的问题，Bi等人（2023）在少样本条件下遥感场景分类任务中首次探索了基于多模态预训练模型的微调范式，验证了多模态提示耦合的有效性。

Zhang等人（2024b）以融入遥感领域知识的VLM模型为原型，提出基于频率分布的多模态微调策略（frequency distribution-based multi-modal fine-tuning，FreqDiMFT），以缓解小样本场景中泛化能力局限问题。具体而言，在视觉分支中运用快速傅里叶变换构建频率分布整合模块，从频域角度充分捕捉基于多尺度频率的深层语义特征，再通过嵌入局部—全局频率分布信息解决遥感图像中类间相似度高和类内多样的问题。为突出对特定类别物体的理解，同时消除与目标无关背景的干扰，在Transformer模块中设计自适应特征精炼模块以有效过滤冗余特征。在文本提示层为减轻领域漂移，采用将基本模板与遥感领域专业知识相结合的输入格式，以生成更具辨别性的类别原型。不局限于Bi等人（2023）所提方法对双分支提示耦合关系的关注，该方法还充分结合遥感领域知识对视觉和文本多模态提示进行调优，有效突出深层语义特征并过滤冗余特征。类似地，融合多模态知识进行少样本分类任务的工作还有Ji等人（2024）提出的ProDFSL方法。该方法首先用ChatGPT为每类样本生成特定文字描述，然后引入跨模态语义知识生成模块，以生成更新的动态软提示，并用浅层Transformer模块对序列间依赖关系进行建模。

综合上述方法可见，场景识别多采用基于提示的微调方法进行改进，利用单模态提示时存在提示生成灵活度低、动态适配能力差的问题，但参数效率较高；利用多模态提示能通过联合优化增强跨模态对齐和协同，显著提升泛化能力，更适用于少样本和类增量泛化场景下的应用。

3.2 语义分割

语义分割旨在为图像中各个像素赋予一个语义类别，如建筑物、林地和湖泊等，地物分类属于语义分割任务。该任务提供的高分辨像素级信息对于环境监测、灾害管理和农业治理等应用而言不可或缺。SAM架构的灵活性和通用性使其在图像分割领域具有强大的性能，然而将其直接应用于遥感领域下游任务可能存在域偏移导致的局限性，研究人员针对如何结合数据特点和应用需求对SAM进行高效微调展开了系列研究，代表性方法如表3所示。

Gao等人（2024）首次将提示学习用于仅有源域上预训练的离线模型和目标域数据的无源域自适应分割场景。首先提出注意力引导的提示生成策略，通过对不同层和不同位置赋予不同权重生成动态提示。其次，提出基于相似距离的特征层领域对齐策略，以减少目标域特征与源域原型间的差异。

遥感领域对SAM进行定制化改进的工作主要集中在实例分割下游任务应用。RoadSAM是Feng等人（2024）提出的用于高分辨遥感图像道路提取的SAM适配器框架，该方法引入了3种变体形式（顺序式、平行式、混合式），使适配器能灵活嵌入Transformer模块的不同位置。为保留图像的重要细节和结构信息，还设计显式视觉提示模块EVP，从补丁嵌入特征和高频成分特征中学习显式提示信息。Zhang等人（2024a）提出的RSAM-Seg无需手动干预以提供提示信息，该方法在SAM编码器的ViT模块

表 3　面向语义分割的参数高效微调方法总结

Table 3　Summary of typical parameter-efficient fine-tuning methods toward semantic segmentation

方法	类型	贡献	优点	缺点
Gao等人(2024)	Prompt微调	提出“注意力引导提示调优”策略用于无源域自适应应用	能动态调整不同层和位置的提示权重	计算复杂度略高,训练时间较长
RoadSAM(Feng等,2024)	Adapter微调	提出3种适配器变体,并结合高频分量信息增强边缘提取能力	引入遥感先验知识,参数效率高且计算资源占用少	显式视觉提示依赖高频分量提取,对图像预处理过程敏感
RSAM-Seg(Zhang等,2024a)	Adapter微调	为结合遥感图像特性在SAM编码器结构中以混合式形式插入适配缩放模块和适配特征模块	无需手工设计提示,所提框架可用于辅助标注	同上,编码器构建依赖高频分量特征提取
MC-SAM Seg(Zheng等,2025)	Adapter微调	SAM编码器中引入多认知视觉适配器(Mona)	多尺度目标复杂细节捕捉能力强	适配器设计复杂,训练收敛较慢
SAM-RSIS(Luo等,2024)	Adapter微调	提出包含边界框生成阶段和掩膜生成阶段的两阶段SAM分割框架SAM-RSIS	多尺度目标分割能力提升,有助于小目标分割	模型训练流程复杂,分割结果依赖检测器精度
Pu等人(2024)	Adapter微调	设计了端到端的SAM轻量化适配结构用于机载SAR语义分割	轻量化掩码解码器的使用使模型相对于原始SAM的类别可感知能力强	模型精度仍依赖低频语义信息提取
Xue等人(2024)	LoRA微调	提出双编码器架构,并改进SAM掩码解码器,提出UperNet头进行多尺度特征融合	大量节省训练参数	双编码器结构消耗内存
Wang等人(2024c)	LoRA微调	提出显式解耦的前景—语义适配器网络(DFSA-Net),用于广义少样本语义分割GFS-Seg	背景干扰抑制能力强,通过低秩矩阵适配新类别,并能保留基础类别性能	双分支网络结构设计复杂,增加计算成本

中以混合式形式插入适配缩放模块Adapter-Scale,并在ViT模块间插入适配特征模块Adapter-Feature。与RoadSAM相似,该模块旨在结合高频率图像信息和图像嵌入特征以生成包含图像信息的提示。

为适应遥感图像中目标的多尺度特性,Zheng等人(2025)利用多认知视觉适配器构建SAM-Mona编码器,所提分割框架MC-SAM Seg通过微调SAM-Mona编码器以及特征聚合器提取高质量特征。为适应遥感视觉任务对复杂场景下多尺度目标特征提取、目标特性利用以及目标密集排列形式的要求,该方法集成了多认知视觉适配器,用多分支深度可分离卷积增强模型对多尺度和细粒度特征的提取。Luo等人(2024)提出的SAM分割框架SAM-RSIS包含边界框生成阶段和掩膜生成阶段。受ViT-Adapter启发,在第1阶段利用多尺度适配器微调ViT骨干网络并训练目标检测器;在第2阶段利用边界框信息作为提示,并结合适配器的多尺度特征,对掩码解码器进行微调。Pu等人(2024)设计了类别感知的SAM适配器结构并用于机载合成孔径雷达(synthetic aperture radar, SAR)图像的地物分类任务。

还有部分工作采用低秩自适应对SAM模型进行参数高效微调以完成下游任务的适配,大多数方法将LoRA应用于ViT编码器中每个Transformer块的查询和值投影层。Xue等人(2024)结合SAM编码器和LoRA方法用于航空图像的特征提取和微调。为充分利用多尺度上下文信息,还采用交叉注意力模块有效结合ViT的全局语义和CNN的局部细节特征。受UperNet头部能够处理多尺度特征的启发,使用UperNet头替换SAM掩码解码器进行多尺度特征融合并生成分割掩码,一定程度上消除提示嵌入导致的混淆问题。

现有的广义少样本语义分割方法在应用于航空影像时面临背景类内差异大、基础类别性能下降以及新类别过拟合等问题,Wang等人(2024c)提出用于广义少样本语义分割GFS-Seg的解耦式前景—语

义适配器网络(disentangled foreground-semantic adapter network, DFSA-Net)。为减少背景特征带来的干扰,DFSA-Net采用前景—语义解码器,将语义分割分解为前景聚合和多类别细化两个过程以加强前景判别能力。为缓解微调过程中基础类别性能下降以及新类别过拟合的情况,在微调阶段提出解耦式低秩适配器,旨在保留基础参数的同时确保能高效适应新类别。最后,引入推理集成策略对基础解码器和新解码器的预测进行合并,获得最终输出结果。

综上可见,语义分割应用中的PEFT方法主要集中在对SAM编码器进行适配器或LoRA的改进,针对遥感图像中目标尺度多样化和所处场景复杂的问题,通过在编码器中引入先验知识或加入多尺度结构以提升模型对小目标的感知能力。

3.3 变化检测

变化检测通过对比分析不同时间拍摄的影像识别和监测地物要素发生的变化。通过精准地检测变化,为环境管理、城市规划、灾害应急以及气候变化研究提供了重要先验信息。同语义分割任务,在变化检测任务中SAM同样应用广泛,表4总结了变化检测任务中PEFT方法的代表性工作。LoRA的引入可赋予SAM参数高效微调的能力,减轻自然图像和遥感图像之间的域差异,使得模型更好地适应特定任务或数据集,同时减少计算资源消耗,因此SAM与Adapter或LoRA的结合在遥感图像的实例分割、变化检测等方面应用广泛。典型方式是将Adapter或LoRA微调引入SAM编码器中的Transformer模块进行处理,配合提升模型的全局信息融合能力。

方乐缘等人(2024)提出基于时差提示SAM(temporal difference prompted SAM, TDPS)的遥感图像变化检测方法,具体而言,在编码器Transformer层的自注意力矩阵$\boldsymbol{Q}$,$\boldsymbol{K}$,$\boldsymbol{V}$的线性投影层中引入LoRA微调。其次,还设计了时相差异提示生成器,通过对双时相图像的特征差异与可学习的查询嵌入共同优化,得到SAM模型解码器所需提示向量。针对大量像素级标签缺失条件下的变化检测问题,Wang等人(2024b)还提出基于EfficientSAM的弱监督变化检测方法ESAM-CD。首先,构建以图像级标签为强监督信息的分类模型以生成高质量的多尺度类激活图,然后提出多尺度CAM(class activation map)融合模块精修变化目标的边界。接着以双时相图像和伪标签作为网络输入,将LoRA嵌入EfficientSAM中各Transformer层中进行微调。不同于TDPS的是,为减少EfficientSAM对于提示的依赖,该方法提出基于通用卷积层的无提示解码器预测变化图。相似地,Chen等人(2024b)采用LoRA微调减轻语义空间上的领域偏移问题,提出“时光穿梭”激活机制(time travelling pixels, TTP)以增强双时相图像特征建模能力,并设计了轻量高效的多级预测头,对密集的变化语义信息进行解码。

除了上述用低秩自适应对SAM模型进行高效微调外,在变化检测任务中也常用适配器对SAM进行微调。Mei等人(2024)基于MobileSAM提出的SCD-SAM方法使用语义适配器以聚合关于变化对象的语义导向信息。该方法还设计渐进式特征聚合双向解码器,分别融合二元变化特征和语义变化特征,从而整合上下文信息并减轻不同尺度间的语义差异。Ding等人(2024)采用FastSAM视觉编码器提取视觉表示,提出卷积适配器聚合以任务导向的变化信息。此外为利用SAM提取的固有语义表征,结合任务无关的语义学习分支以提升对潜在语义特征变化的建模能力。

3.4 目标检测

目标检测是指发现并定位图像中存在的多个目标。不同于语义分割任务,目标检测关注于识别特定物体并确定其在图像中的位置,目标检测任务中进行高效参数微调的代表性研究如表5所示。

针对少样本条件下的遥感图像增量目标检测任务,Lu等人(2024)提出基于双频率提示的少样本增量学习框架(few-shot incremental object detection via dual-frequency prompt, FSIOD-DFP),用双频率提示生成器以同时减轻模型的过拟合和灾难性遗忘问题。在生成提示时对图像的频率成分进行解耦使其适应于先前任务上训练的基类模型。接着引入自归一化损失引导经提示修正的图像有效利用基础模型知识,此外还提出任务解耦的检测头解决新旧任务不同导致的背景偏移问题。

Pu和Xu(2025)将基于低秩自适应的参数高效微调方法用于星载实时目标检测模型的部署,减少向地面传输数据的需求并极大节省了通信资源。针对骨干网络以及旋转RCNN检测头中的全连接层进行LoRA微调,而对特征金字塔网络、区域建议网络及检测头末端的轻量化分类和回归全连接层模块采

表 4　面向变化检测的参数高效微调方法总结

Table 4　Comparison and summary of typical parameter-efficient fine-tuning methods toward change detection

方法	基准架构	类型	贡献	优点	缺点
TDPS(方乐缘 等,2024)	原始SAM	LoRA微调	通过LoRA微调和SAM编码器中的提示生成器产生双时相差异特征提示向量,直接转化为提示输入	无需人工设计提示,LoRA微调可有效弥合数据差异,模型以端到端形式优化	提示生成模块会增加推理时间,计算代价和实时性有提升空间
ESAM-CD (Wang等,2024b)	EfficientSAM	LoRA微调	提出多尺度CAM融合策略和无提示解码器	满足低资源需求,无需具体提示作为输入,可用无提示解码器预测变化图,减少对人工提示的依赖,微调成本极低	多尺度CAM融合预处理耗时;依赖骨干网络,产生的像素级伪标签噪声可能影响精度
TTP(Chen等,2024b)	原始SAM	LoRA微调	设计“时光穿梭”激活机制和轻量化多级解码预测头	双时相建模能力强,能实现多尺度特征融合,无额外模块引入	多级解码头的上/下采样操作可能增加内存占用
SCD-SAM (Mei等,2024)	MoblieSAM	Adapter微调	设计深度特征交互模块以对全局—局部特征互补,渐进解码细化边界	有效整合上下文信息,弥合不同尺度语义差异,参数量少	双编码器结构可能引入冗余计算,导致训练时间较长
SAM-CD (Ding等,2024)	FastSAM	Adapter微调	引入与任务无关的语义学习分支以模拟语义潜在特征	编码器能融合多尺度特征,结合时序约束损失提升语义一致性,训练效率高	FastSAM轻量化结构可能牺牲部分精度,任务无关分支略微增加模型复杂度

用全参数微调。此外,为实现可训练参数量和检测器性能的权衡,该方法还对LoRA的权重矩阵提出低秩逼近策略,结合奇异值分解方法分析参数矩阵的固有秩和低秩逼近误差以选择最优秩参数。

Hu等人(2024a)首次探索了SAM在跨域目标检测方面的应用,提出基于SAM适配器微调的SAR图像跨域目标检测框架。首先在SAM编码器中引入适配器模块,然后在微调阶段冻结SAM原始参数,并用标记的源域SAR图像数据更新适配器模块的参数。除上述应用之外,Zhao和Zhang(2024)将参

表 5　面向目标检测的参数高效微调方法总结

Table 5　Summary of typical parameter-efficient fine-tuning methods toward object detection

解译任务	主要方法	贡献	类型	优点	缺点
通用目标检测	FSIOD-DFP (Lu等,2024)	提出双频率提示生成器并对频率成分解耦	提示微调	仅需引入少量参数,能避免灾难性遗忘	精度提升依赖提示生成器设计,多任务串联设计导致存储开销大
有向目标检测	Pu和Xu(2025)	提出首个用于星载有向目标检测的高效参数微调模型	LoRA微调	大量节省通信资源,可实现轻量化检测	模型精度依赖秩参数的选择
跨域目标检测	Hu等人(2024a)	提出基于SAM适配器微调的SAR图像跨域目标检测框架,用源域数据更新适配器参数以适应于目标域数据	Adapter微调	无需任何目标域数据参与	结构单一,可进一步结合SAR图像特性改进
建筑物灾害监测	Zhao和Zhang (2024)	提出多阶段融合适配器模块	Adapter微调	模型辨别细微特征差异能力强	模型结构复杂,参数量多
油污监测	Wu等人(2024)	将YOLOv8生成的边框作为提示输入SAM适配器,然后融合生成掩码和类别信息	Adapter微调	能有效解决像素级分类冲突问题	掩码结果依赖提示输入,流程稍复杂

数高效适配器微调用于建筑物灾害监测，将Adapter注入到每个Transformer模块的末端形成多阶段融合适配器模块，从而增强模型对细微特征的辨别能力。Wu等人（2024）还将基于适配器的SAM用于SAR油污检测任务中，其思路类似Luo等人（2024）提出的SAM-RSIS。首先用YOLOv8（you only look once）检测器获取目标类别和边框，然后将边框坐标作为提示输入到SAM适配器中以检索类别无关的掩码，最后用有序掩码融合模块整合掩码和类别信息，有效缓解了SAM中的像素类别冲突现象。

3.5 基础模型的多任务应用

视觉语言大模型应用的终极目标是适配并提升特定下游任务上的性能表现，而参数高效微调能将大模型的能力快速迁移到下游应用场景，前几小节讨论的是单任务的下游应用，本小节聚焦于研究适用于多任务处理的基础大模型微调方法。

为使PEFT方法满足密集预测型应用的任务需求，Dong等人（2024）提出参数高效联合微调框架UPetu，主要包含高效量化适配器模块（efficient quantization adapter module，EQAM）和上下文感知提示模块（context-aware prompt module，CPAM）。EQAM用于在微调中学习特定任务的特征表示，其创新结合了量化和适配器优化，在前馈线性层中引入高效量化以获取效率和精度的平衡。EQAM的设计有利于细粒度特征信息和特定任务知识的交互。为充分利用丰富的上下文特征，设计的CPAM可动态生成依赖输入的多尺度条件提示以增强上下文建模能力。上述两个模块的协同集成增强了基础模型的表示学习能力和泛化迁移能力。

Hu等人（2024b）提出适配器高效微调架构AiRs以适配遥感下游应用任务，所提框架在Transformer中的多头注意力层和多层感知器后分别加入空间上下文适配模块（spatial context adapter，SCA）和语义适配器模块（semantic response adapter，SRA）。SCA能有效弥补预训练特征空间感知能力的不足。SCA包含大卷积核的空间可分离卷积，以更小代价高效聚合多头注意力层的空间信息。SRA用反向瓶颈操作算子精修语义特征，弥补之前层对遥感语义信息把握的不足。该方法还研究了适配器的集成策略对网络高效训练的影响，从前向传播和反向传播的角度分析了顺序策略和平行策略引入适配器的优势与缺陷，并提出基于残差结构的高效集成策略。

提高微调的效率是模型得以广泛应用的关键，侧边式调节（sidetuning，ST）通过绕开卷积神经网络梯度流的方式减少训练内存占用，然而这种方法很难实现与完全微调相当的性能。为兼顾微调训练效率和模型性能，Hu等人（2024c）提出高效训练适配框架（training-efficient adapting，TEA），通过在基础模型上拼接以GhostNetv2为原型的轻量级网络对提取特征进行精修，弥补了跳过特征提取器进行反向传播造成的梯度流损失，同时保证网络的训练效率和网络参数效率。

上述模型均针对背景复杂、场景范围大、目标尺度多的解译任务特点展开研究，采用冻结主干网络且引入可训练模块的高效参数微调范式展开设计，主要面向检测、分割和分类等有密集预测需求的任务，旨在有效降低全参数量微调的计算或存储成本。

4 模型性能评估分析

4.1 常用数据集

表6以解译任务为类别总结了目前常见的可见光及SAR遥感图像解译常用数据集。

为综合评估所比较的方法，采用3个最常用的遥感场景分类数据集，即加州大学发布的土地利用遥感数据集（University of California at merced land-use dataset，UCM）、西北工业大学发布的场景分类数据集（Northwestern Polytechnical University remote sensing image scene classification，NWPU-RESISC45）以及多来源航空影像数据集（aerial image dataset，AID）进行性能对比。

为公平对比各类方法的性能，训练时所有图像尺寸调整为224×224像素，UCM按照8∶2划分训练集和测试集，NWPU-RESISC45按照2∶8划分，AID按照5∶5划分。检测任务中DOTA、HRSC2016和DIOR/DIOR-R数据集的输入尺寸依次为1 024×1 024像素、1 024×1 024像素和800×800像素，UCAS-AOD和NWPU-VHR10数据集输入尺寸为1 000×600像素。语义分割任务中典型数据集iSAID通常将图像裁剪为896×896像素作为输入尺寸，Vaihingen、Potsdam和LoveDA数据集输入尺寸为512×512像素。对于典型变化检测数据集LEVIR-CD和WHU-CD，输入尺寸通常设置为256×256像素。

表6　常用遥感图像解译数据集汇总

Table 6　Summary of commonly used remote sensing image interpretation datasets

解译任务	数据集	尺寸/像素	类别	图像数量/幅	文献
场景识别	NWPU-RESISC45	256 × 256	45	31 500	Cheng等人(2017)
	UC-Merced	256 × 256	21	2 100	Yang和Newsam(2010)
	AID	600 × 600	30	10 000	Xia等人(2017)
	FGSC-23	–	23	4 052	Zhang等人(2020)
	FGSCR-42	50 × 50 ～ 1 5 00 × 1 500	42	9 320	Di等人(2021)
	EuroSAT	64 × 64	10	27 000	Helber等人(2019)
	PatternNet	256 × 256	38	30 400	Li等人(2018a)
	RSICD	224 × 224	30	10 000	Lu等人(2018)
	MLRSNet	256 × 256	46	109 161	Qi等人(2020)
目标检测	HRSC2016	1 024 × 1 024	31	1 061	Liu等人(2017)
	NWPU-VHR10	533 × 597 ～ 1 728 × 1 028	10	800	Cheng等人(2014)
	DOTA-v1.0	800 × 800 ～ 4 000 × 4 000	15	2 806	Xia等人(2018)
	DIOR/DIOR-R	800 × 800	20	23 463	Li等人(2020a)；Cheng等人(2022)
	FAIR1M	1 000~10 000	37	15 000	Sun等人(2022)
	SSDD	500 × 500	1	1 160	Li等人(2017)
	HRSID	800 × 800	1	5 604	Wei等人(2020)
	UCAS-AOD	1 280 × 659、1 372 × 941	2	2 420	Zhu等人(2015)
地物分类/语义分割	Vaihingen/Potsdam	800～3 000、6 000 × 6 000	6	33/38	International Society for Photogrammetry and Remote Sensing (2020)
	LoveDA	1 024 × 1 024	7	5 987	Wang等人(2022c)
	URUR	5 120 × 5 120	8	3 008	Ji等人(2023)
	DeepGlobe RE	1 024 × 1 024	6	6 226	Demir等人(2018)
	WHU aerial building	512 × 512、1 024 × 1 024	1	8 188/5 640	Ji等人(2019)
	WHU-Mix	300 × 300、512 × 512、650 × 650	1	64 000	Wei等人(2023)
	Inria	5 000 × 5 000	1	2 880	Maggiori等人(2017)
	38-cloud	5 000 × 5 000	1	826	Mohajerani和Saeedi (2019)
	Sentinel-2	224 × 224	1	1 966	欧洲航天局(2015)
	iSAID	800 × 800～4 000 × 13 000	15	2 806	Waqas等人(2019)
变化检测	LEVIR-CD	1 024 × 1 024	6	637	Chen和Shi(2020)
	WHU-CD	15 354 × 32 507	1	7 620	Ji等人(2019)
	Second-CD	512 × 512	6	2 968	Yang等人(2022)
	Landsat-CD	416 × 416	–	2 385	Yuan等人(2022)
	S2Looking	1 024 × 1 024	1	5 000	Shen等人(2021
	CLCD	512 × 512	4	600	Yang和Huang(2021)

注：“–”表示未提供相应数据。

4.2 评测指标

在着重于检验模型泛化能力的通用化零样本场景识别任务中，为同时评估基础类别和新增类别的识别准确率，通常采用基类和新类识别准确率的调和平均（harmonic mean average，HMA）评估模型整体性能。在一般场景中用OA（overall accuracy）或ACC（accuracy）表示模型识别准确率。在类增量式连续学习任务中通常用平均精度ACC和反向传输BWT（backward transfer，BWT）两项指标以评估所提方法的性能。评测指标BWT能反映持续学习方法中对旧类目标知识的保留能力。

在地物分类或变化检测任务中，通常用查准率（precision，Pre）、查全率（recall，Rec）、交并比（intersection over-union，IoU）、总体准确率（OA）和F1分数来评估变化检测算法的性能。在语义分割任务中，还可选择不同IoU阈值下的掩膜精度AP_{mask}和边框检测精度AP_{box}作为评估指标，如AP_{mask}@0.5表示阈值为0.5时的掩膜检测精度。值得注意的是，由于ISPRS Vaihingen/Potsdam数据集中在计算定量指标时通常忽略杂波类别，因此平均交并比（mean IoU，mIoU）仅针对5类目标进行计算。在目标检测中通常以平均精度均值（mean average precision，mAP）或准确率和召回率作为性能评价指标。

4.3 实验结果分析

本节首先区分是否为多任务应用，然后在单任务应用中依据不同微调类型在以下各小节中对现有算法性能进行评估，并从可训练网络参数、模型精度、内存消耗以及推理延时等角度进行分析，便于研究人员清晰理解参数高效微调引导的遥感图像解译任务的发展脉络和性能比较，从而推动该领域的进一步发展。由于篇幅限制，本节表格中实验结果仅罗列代表性文献的部分实验结果，详细结果可查阅对应文献。

4.3.1 Prompt微调系列方法精度分析

表7总结并归纳了多种经典PEFT方法的实验结果，以对比系列方法在遥感场景分类任务上的性能。考虑到遥感图像解译中以提示词作微调的方法多在场景识别下游任务上进行增量学习、持续学习或少样本学习等条件下的泛化应用，表7中区分不同场景并主要选用UCM、AID、RESISC-45数据集进行性能比较。表7中ViT-T表示的骨干网络为ViT-tiny，ViT-S表示的骨干网络为ViT-small，ViT-B/32和ViT-B/16均以Base模型为骨干网络，图像块大小分别设置为32×32像素和16×16像素。Way表示某任务中涉及的类别数量，Shot表示每个类别中用于训练的样本数量，halfway表示对于基类和新类各选取总类别数的一半。

表7中所示文献结果表明，即使使用专用大规模视觉基础模型ViTAEv2-S，以ViT-base为骨干的LVM-StARS方法（Yang等，2024b）相比以ViTAEv2-S为骨干的IVP在AID数据集上准确率高出1.34%，平均准确率能达到98%左右。究其原因，改进的深层浅层提示方法在层与层间引入软自适应机制，能有效控制提示标记的信息流转，相比于仅使用硬提示准确率提升2%左右。与完全微调相比，LVM-StARS仅需要更新0.1%～0.5%的训练参数，总体准确率能提高1.71%～3.94%。

Zhu等人（2024）提出的MVP方法在元训练阶段引入提示微调获得的性能增益要优于元微调阶段引入提示，原因在于前者能帮助模型获得更好的初始化值以便于向新任务泛化。5-shot条件下在每个任务中选取5类时分别用ViT-tiny，ViT-small作为骨干网络进行测试，MVP方法与全参数微调相比在所有任务类别上的平均精度依次增加6.64%和3.46%，再次证明了提示微调方法的泛化能力较强。Zhao等人（2023a）提出的Pre-T方法在连续学习任务中表现出色，较好克服了灾难性遗忘现象，而使用Pro-T时指标BWT在RESISC-45和AID上分别下降4.93%和9.62%，性能较为逊色。在多尺度高分辨率遥感图像数据集RESISC-45和AID上，提示长度和提示数量越小反而越能克服灾难性遗忘，这是因为多尺度遥感图像中特征分布形状和目标空间组成关系较为复杂，引入较多提示参数易引起知识混淆干扰。

表7中所示文献结果表明，与使用CLIP进行零样本学习方法相比，APPLeNet在新类上泛化能力更强，在PatternNet、RSICD、RESISC45以及MLRSNet等数据集上的平均精度优势明显，其中已知类别精度高出33.47%，未知类别精度高出4.04%，整体HMA值高出15.93%。将APPLeNet与上下文提示优化系列方法进行比较，从RESISC45上的基类、新类精度和HMA值上来看，APPLeNet性能分别比CoOp和CoCoOp高出2.2%、4.7%、4.2%以及1.5%、3.3%、2.9%。与基于提示对齐的梯度引导方法相比，APPLeNet在新类上的精度提升更为明显，

表 7　提示微调方法在不同应用场景中的性能对比

Table 7　Comparison of the performance of prompt tuning methods under different application

基本场景	研究工作	条件	骨干网络	数据集	指标:得分/%
一般场景	IVP(Fang等,2023)	–	Swin-T	UCM	ACC:99.76
				AID	ACC:97.11
				RESISC45	ACC:95.27
			ViTAEv2-S	UCM	ACC:99.67
				AID	ACC:96.69
				RESISC45	ACC:95.00
	LVM-StARS(Yang等,2024b)		ViT-base	AID	ACC:98.03
				RESISC45	ACC:95.53
类增量学习	Continual(Pre-T)(Zhao等,2023a)	–	ViT-B/16	UCM	ACC:92.33
					BWT:−1.43
				AID	ACC:81.10
					BWT:−9.58
				RESISC45	ACC:72.90
					BWT:−11.56
	Continual(Pro-T)(Zhao等,2023a)	–	ViT-B/16	UCM	ACC:92.33
					BWT:−3.50
				AID	ACC:78.22
					BWT:−14.51
				RESISC45	ACC:64.61
					BWT:−21.18
	FSIOD-DFP(Lu等,2024)	5-shot	ResNet-101	DIOR	mAP (Base/Novel/All):71.59/31.34/61.53
				NWPU VHR-10	mAP (Base/Novel/All): 90.56/74.95/85.87
				FAIR1M-Airplane	mAP (Base/Novel/All):55.37/40.18/51.23
小样本学习	MVP(Zhu等,2024)	5way 5-shot	ViT-T	UCM	ACC:89.3
				WHU	ACC:93.4
				RESISC45	ACC:81.5
				AID	ACC:84.7
			ViT-S	UCM	ACC:84.1
				WHU	ACC:89.6
				RESISC45	ACC:77.2
				AID	ACC:79.6
	MaPLe(Bi等,2023)	16-shot	ViT-B/16	UCM	ACC:95.7
				AID	ACC:92.2
	ProDFSL(Ji等,2024)	5-shot	ResNet12	RESISC45	ACC:94.12
				AID	ACC:91.04
			Conv-256	RESISC45	ACC:91.27
				AID	ACC:87.00
通用化零样本学习	FreqDiMFT (Zhang等,2024b)	60way 16-shot	ViT-B/32	RSSDIVCS	HMA:69.46
				LSRS	HMA:37.22
	EPT(Lan等,2025)	4-shot	ViT-B/16	EuroSAT	HMA:74.72
				FGSCM-52	HMA:22.80
				FGSCR-42	HMA:29.75
	APPLeNet(Singha等,2023)	16-shot (halfway)	ViT-B/16	PatternNet	HMA:77.55
				RSICD	HMA:74.16
				RESISC45	HMA:72.73
				MLRSNet	HMA:65.66
域自适应学习	Source-free DA(Gao等,2024)	–	ViT-Huge	Vaihinge-Potsdam	mIoU:46.62
				Rural-Urban	mIoU:42.83

注:“–”表示无。

整体HMA值高出2.23%。由此可见，解耦视觉风格和图像原始内容并进行注意力关联，基于遥感图像特性动态生成提示有利于增强泛化能力。

利用多模态提示交互处理的方法中，Bi等人(2023)将耦合式多模态对齐提示微调用于小样本场景识别任务，MaPLe相比CLIP-Adapter在UCM和AID数据集上精度提升12%和6.2%，相比单模态上下文优化提示方法CoOp精度提升1.7%和0.6%。Ji等人提出的ProDFSL在小样本学习任务中能达到优于迁移学习、元学习、度量学习和传导式学习等方法的效果。使用最普通的提示学习能改善少样本分类任务的精度，在1-shot条件下精度增加0.6%，使用“This is a photo of [Class_name]”此类提示在1-shot条件下分类精度提升1.1%左右，然而仅将类别名作为提示由于未引入遥感领域知识性能提升甚微。整合跨模态语义知识生成模块和视觉软提示后，1-shot条件下精度提升4.8%，再次证明了跨模态联合嵌入提示能提升特征丰富性、增强语义连贯性以及提高领域自适应能力。

4.3.2 Adapter微调系列方法精度分析

表8中所示文献结果表明，尹文昕等人(2024)提出的多尺度融合适配模块MuFA相比于采用串联连接和单尺度结构的Adapter以及采用并联连接和单尺度结构的AdaptFormer在UCM和RESISC-45数据集上还能分别提升0.72%、0.47%和1.84%、0.22%，其中并联式适配结构既保持了模型性能又降低了资源开销，多尺度融合有利于保留关键的特征信息。

针对道路提取任务设计的RoadSAM采用显式视觉提示EVP和频率Adapter微调方法，在URUR和DeepGlobe RE数据集上均是混合适配器效果最好，使用并行式和顺序式适配器性能相当，但并行结构灵活性更强，便于后续优化和改进。加入基于嵌入式向量的提示策略后，在URUR和DeepGlobe RE上的F1和IoU值提升6.6%、7.82%和9.24%、12.23%，再次证明冻结的补丁嵌入式特征和高频分量提示能提升分割性能。同时在训练参数量方面，所提方法仅需更新约10%的模型参数就即可实现大幅性能提升。RSAM-Seg同样整合了图像的高频信息，在DeepGlobe RE数据集上相比以中心点作为负样本类提示的方法整体精度和IoU值分别高出5.3%和33.4%左右。RoadSAM和RSAM-Seg原理相似，但RSAM-Seg分割的IoU值高出10.83%。

SAM-RSIS在捕捉精细的目标轮廓细节和实现高质量实例分割方面表现突出，掩膜精度AP75相比其余先进的实例分割方法及SAM变体方法均达到最优。在WHU aerial building上的实验结果表明，与SAM变体方法中最优的SAM-Seg相比，mAP和AP75还高出4%和4.5%，而Grounded-SAM由于没有针对遥感领域数据进行调优，同以点作为提示的SAM相似，分割性能均较差。MC-SAM Seg方法在引入Mona适配结构后，相比最好的RSPrompter(Chen等，2024a)在WHU aerial building数据集上的AP_{mask}上高出2.0%，在交并比阈值为0.75的条件下AP_{box}高出6.5%，这表明通过引入特定领域知识的改进适配器微调能增强分割模型的图像理解能力并提升模型上限。该方法还验证了所提模块在HRSID上的有效性。MC-SAM Seg在用于背景复杂的建筑物、密集排列的房屋以及含噪且尺度分布多样的目标时，相较于其他方法能生成更准确、更完整且更清晰的实例分割结果。综合WHU aerial building数据集在阈值区间0.5～0.95上的掩膜生成平均精度来看，SAM-RSIS比MC-SAM Seg高出1.4%，平均精度可达72.6%。

表7和表8中所示文献结果表明，针对场景识别使用改进视觉提示或嵌入适配器进行参数高效微调，在RESISC-45上总体准确率达到94%～95%左右，与全参数微调性能相当。UCM上使用PEFT方法总体准确率均在99%左右，且采用改进的提示微调方法还能将准确率提高0.6%～0.7%左右。

4.3.3 LoRA微调系列方法精度分析

表8和表9中所示文献结果表明，SAM-CD方法中采用FastSAM编码器相比于普通SAM能大幅降低参数量和浮点运算量，且在LEVIR-CD数据集上的F1值和mIoU高达95.5%和91.68%。虽然EfficientSAM本身分割性能更强，但FastSAM搭配适配器的微调效果比ESAM-CD在F1值和mIoU上分别高出13.88%和22.73%。SAM-CD，TDPS和TTP在LEVIR-CD上的整体准确率均达到99%以上且精度相近，SAM-CD相比TDPS和TTP在F1值和IoU上高出3.2%，5.98%和3.4%，6.08%。ESAM-CD方法利用EfficientSAM完成半监督变化检测任务，在WHU-CD和LEVIR-CD数据集上的F1值比性能最优的SOTA(state-of-the-art)弱监督方法高出5.10%和

表 8　适配器微调方法在不同解译任务中的方法对比

Table 8　Comparison of the performance of adapter tuning methods under different interpretation task

解译任务	方法	出版期刊/平台	网络架构	数据集	指标:得分/%
场景识别	MuFA(尹文昕 等,2024)	电子与信息学报	ViT-base	UCM	ACC:98.57
				RESISC45	ACC:91.29
			Swin-Tiny	UCM	ACC:99.52
				RESISC45	ACC:94.10
地物分类	RoadSAM(Feng等,2024)	GRSL	原始SAM	URUR	F1:73.29
					IoU:57.84
				DeepGlobe RE	F1:81.65
					IoU:68.99
	MC-SAM Seg(Zheng等,2025)	JSTAR	SAM-Mona	WHU aerial building	AP_{mask}:71.2
					AP_{mask}@0.5:91.5
				HRSID	AP_{mask}:66.4
					AP_{mask}@0.5:97.0
	SAM-RSIS(Luo等,2024)	TGRS	原始SAM(ViT-Huge)	WHU aerial building	AP_{mask}/AP_{mask}@0.5:72.6/91.0
				WHU-Mix	AP_{mask}/AP_{mask}@0.5:52.4/80.2
				NWPU VHR-10	AP_{mask}/AP_{mask}@0.5:64.1/88.5
	RSAM-Seg(Zhang等,2024a)	ArXiv	原始SAM	DeepGlobe RE	F1 score:75.48
					mIoU:79.82
					OA:97.85
变化检测	SCD-SAM(Mei等,2024)	TGRS	MoblieSAM(ViT-Tiny)	Second-CD	F1-score:70.51
					mIoU:77.75
					OA:90.16
				Landsat-CD	F1-score:81.16
					mIoU:83.04
					OA:93.30
				LEVIR-CD	F1-score:95.50
					mIoU:91.68
					OA:99.14
				WHU-CD	F1-score:97.58
					mIoU:95.36
					OA:93.30
目标检测	Cross-domain DA(Hu等,2024a)	PIERS	原始SAM	HRSID	Prec:61.0

4.63%,证实了用LoRA微调EfficientSAM以及使用通用卷积层作解码器的潜力。

表9中所示文献结果表明,将LoRA用于语义分割这种密集预测任务中效果较好,Xue等人(2024)所提方法在ISPRS Vaihingen和Potsdam数据集上mIoU可达82.18%和85.69%,以ViT-H为骨干网络条件下与不使用LoRA相比F1值高出2.42。在训练参数量方面与其余不采用SAM结构的普通分割方法相当,且以ViT-B,ViT-L和ViT-H为骨干网络时,可训练参数量占模型总参数量比例分别为42.54%,20.04%和12.50%。DFSA-Net中引入的解耦式低秩适配器模块能有效保留基类知识,在5-shot条件下基类和新类mIoU分别提升7.27%和1.52%。

4.3.4　综合微调方法的多任务应用精度分析

表10总结并归纳了最新综合微调方法UPetu、AiRs、TEA及多种经典PEFT方法在典型解译应用上的对比实验结果。针对TEA方法采用以中间尺度即GhostNetV2 1.3×为侧边适配模块的基准模型进行结果比较。UPetu中使用GeoSense上预训练的ConvNeXt-B作为遥感基础模型,AiRs、TEA和其余对比PEFT方法均使用ImageNet-21k上以Swin-B为特征提取器训练所得的预训练模型。在FAIR1M上进行旋转框目标检测实验时,所有方法默认将RingMo作为预训练模型。UPetu分别针对地物分类

表9　低秩自适应方法在不同解译任务中的方法对比

Table 9　Comparison of the performance of LoRA methods under different interpretation task

解译任务	研究工作	出版期刊/平台	网络架构	数据集	指标:得分
变化检测	TDPS(方乐缘 等, 2024)	信号处理	原始SAM	LEVIR-CD/ WHU-CD	F1-score:91.1%/94.0% IoU:83.7%/88.8% OA:99.1%/99.5%
	TTP(Chen等,2024b)	Arxiv	原始SAM	LEVIR-CD	F1-score:92.1% IoU:85.6% OA:99.2%
	ESAM-CD(Wang等, 2024b)	TRGS	EfficientSAM	LEVIR-CD/ WHU-CD	F1-score:81.62%/73.01% IoU:68.95%/57.49%
地物分类	SAM+LoRA(Xue等, 2024)	GRSL	ViT-H (EfficientNet-B4辅助编码)	Vaihingen/Potsdam	mIoU:82.18%/85.69% mF1:85.83%/86.24% 参数量:90.11 M GFLOP:248.95
	DFSA-Net(Wang等, 2024c)	TGRS	ResNet101	NWPU(base & new) iSAID(base & new)	mIoU:63.55%/41.84% mIoU:69.18%/18.85%
目标检测	LoRA-Det(Pu和Xu, 2025)	TGRS	Swin Transformer	HRSC2016 DOTAv1.0 DIOR-R	mAP/参数量:71.9%/12.30 M mAP/参数量:76.27%/12.30 M mAP/参数量:64.2%/14.50 M

任务和变化检测任务在Potsdam和LEVIR-CD数据集上实验了Full-tuning, Bias-tuning, Head-tuning, LoRA, Adapter(具体有Conv-Adapter和AdaptFormer两种)等共7种微调方法。在上述两个数据集上进行适配器微调时所得结果采用的是Conv-Adapter方法。在AiRs和TEA文献中,head-tuning指将与下游任务相关的组件设为可训练,冻结所有预训练特征提取部分,因此可训练参数量可忽略不计,表中没有记录数值。

图9对几种典型PEFT方法的参数量和占比进行可视化展示,其中参数量表示所提参数高效微调方法的实际训练参数量,百分比表示可训练参数量相对于全参数微调时参数量的比例。

表10和图9中所示文献结果表明,针对场景识别任务使用改进视觉提示、嵌入适配器或低秩自适应模块进行参数高效微调,RESISC-45上总体准确率与全参数微调性能相当,能达到95%左右。在UCM和AID上使用经典的适配器和低秩自适应方法的效果与全参数微调相近,总体准确率均在99%和97%左右,且训练参数量减少的程度相当均在4%左右。使用UPetu微调在UCM、AID、NWPU-RESISC45数据集上能取得媲美全参数微调方法的

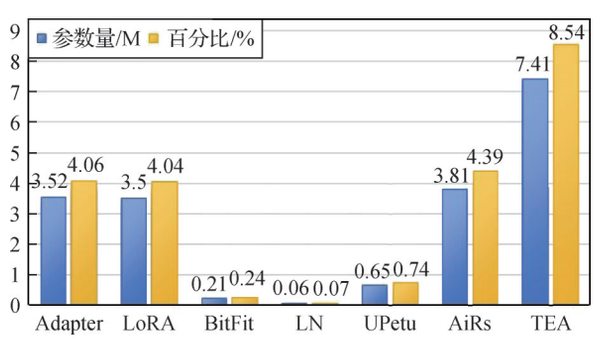


图9　典型微调方法参数量需求比较

Fig. 9　Comparison of parameter amount among typical PEFT methods

精度,且在AID上还超过全参数微调1%,可训练参数量只占全参数微调参数量的0.74%,大幅提升了模型参数的利用效率且降低了训练和推理过程中的计算成本。

由于UPetu文献中与TEA和AiRs采用不同的基础模型,因此无法直接比较各个下游任务中使用不同微调方法的训练参数量消耗,这也是UPetu方法在地物分类任务中参数量占比与其他方法的值差距较大的原因所在。UPetu的可训练参数量与Full-tuning, Head-tuning, LoRa, Adapter等PEFT方法相当,在地物分类任务中相对于全参数量微调方法可

表 10　综合微调方法在多任务应用中的性能对比

Table 10　Comparison of the performance of comprehensive fine-tuning methods under multi-task application

/%

任务	数据集	传统方法		PEFT方法						
		Full-Tuning	Head-Tuning	Adapter	LoRA	BitFit	LN	UPetu	AiRs	TEA
场景识别	UCM	99.05	89.05	98.86	99.14	95.62	95.71	**98.52**	99.33	98.76
	AID	97.29	80.66	96.78	96.99	93.54	94.15	**96.29**	97.51	97.04
	RESISC45	95.50	78.84	95.05	94.45	90.48	94.10	**93.79**	95.38	94.95
地物分类	Vaihingen	82.26	77.66	82.19	81.86	80.52	80.28	–	82.19	82.08
	Potsdam	84.37	78.61	82.11	82.48	81.30	–	**83.17**	–	–
	LoveDA	53.10	48.69	53.13	53.07	51.20	51.02	–	53.59	52.60
	iSAID	68.70	60.82	67.65	66.66	64.36	64.03	–	68.25	67.45
水平框检测	NWPU-VHR10	94.88	91.89	94.30	95.72	88.98	93.98	–	96.29	96.18
	UCAS-AOD	95.94	95.24	96.36	96.43	96.58	96.29	–	96.85	96.87
	DIOR	76.40	67.76	76.40	75.75	71.08	69.94	–	76.99	76.20
旋转框检测	FAIR1M	25.10	16.11	24.30	21.23	20.89	21.09	–	25.22	24.56
变化检测	LEVIR-CD	93.12	84.45	87.46	86.95	–	–	**88.5**	–	–

注：加粗字体表示文献 UPetu 中所得结果，“–”表示无对应参考结果。

训练参数量占 1/4。而由于采用不同的基础模型，AiRs 方法中记录的上述 PEFT 方法可训练参数量占比在 0.07%～4.39% 之间。AiRs 仅需训练全部参数量的 4.39%，在 NWPU-VHR10、UCAS-AOD、DIOR 和 FAIR1M 等检测数据集上即可达到分别超越全参数微调方法 1.41%、0.91%、0.59%、0.12% 的检测精度。虽然 AiRs 在可训练参数量方面利用效率高且性能可观，但其在节省内存占有量和推理时间方面不如 TEA。

在变化检测任务中，UPetu 方法仅用 0.88% 的参数量即可达到超过其余 PEFT 方法的性能，但离全参数微调精度还有一定差距。在亟需兼顾训练参数量存储和模型性能的实际应用中，AiRs 方法可能是最佳选择，其证实了适配器的有效集成可以使 PEFT 方法性能大幅提升甚至超越全参数微调方法，在地物分类、目标检测等密集预测任务中也能发挥效能。

5　研究展望

本文根据解译任务和典型参数高效微调方法总结了 PEFT 方法在遥感下游领域的应用，阐述并比较了各类方法的特点和主要贡献。目前来看，面向遥感图像解译的参数高效微调研究仍处于起步阶段，该领域仍存在进一步探索的潜力。鉴于前述研究进展，对该领域的研究展望如下：

1）面向生成任务和多模态数据的高效微调研究。一方面，当前大多数参数高效微调方法是面向理解型任务量身定制的，如场景分类、语义分割、变化检测和目标检测等任务，然而在面向生成式任务中的探索应用较少，如时序图像生成、图像超分辨重建和图像修复等任务。已有研究人员为生成式稳定扩散模型提出相应的参数高效微调方法，在遥感领域已有文献 DiffusionSat（Khanna 等，2023）使用自监督学习生成逼真的遥感影像以解决多种生成式任务。另一方面，多模态数据相较于单模态数据更符合人类感知与认知，具备多模态理解、高效交互感知与逻辑推理等优势，可迁移到视觉定位、图像问答以及跨模态检索等多类下游任务。当前遥感领域也涌现出部分遥感多模态数据集和遥感多模态大模型（Hu 等，2023a；Sun 等，2023；Guo 等，2024；Wang 等，2024d；Li 等，2025），但在多模态领域针对视觉问答检索等任务进行参数高效微调方法的研究工作较为

稀缺。由于大型多模态模型相较于单模态模型通常需要更多的计算资源和内存资源，研究参数高效微调方法有助于促进跨模态知识的对齐，从而使下游多模态任务得到显著改进。综合上述两方面可见，结合适配器、低秩自适应和多种提示技术进行微调，有利于遥感生成式基础模型赋能结合文本、图像和音视频等多模态内容的下游应用，探索参数高效微调方法在多模态遥感领域生成式任务中的应用极具前景。

2）参数高效微调方法的可解释性研究。参数高效微调方法在应用于自然语言处理领域时可将提示解释为较为直观的描述，而在计算机视觉领域，特别是视觉提示常作为无序的基于标记的提示进行学习，难以转化为易于理解的形式，因此模型的预测可信度较低、可解释性能力较弱。当前工作在进行遥感图像解译时未充分有效利用物理、地理和专家等知识，缺乏可解释性。可解释性遥感大模型的研究成果相对较少但应用前景广泛，为提高遥感大模型的决策认知能力，需要融合先验信息或多源知识对模型进行引导，从数据、模型以及训练与推理等层面适当突破知识获取、表征与融合，数据、知识与模型相互耦合的机制构建，人机交互协同等技术，揭示模型处理多模态信息的内在机制，帮助大模型不断调整和优化决策，提升模型的鲁棒性和信任度，研制透明、可靠且高性能的参数高效微调多任务下游应用系统。

3）遥感领域的高效微调方法部署应用研究。对于大多数模型压缩算法需要进行微调训练以恢复模型精度，但对于大模型而言，全参数微调和训练成本高昂，将耗费大量计算资源并难以将模型部署于资源受限的机载星载边缘端设备。如何在参数高效微调算法中融合模型压缩，并兼顾模型精度和推理速度提升，以加强模型对边缘端设备的适应性和鲁棒性也是重要挑战。当前已有文献专注于研究参数高效微调剪枝、参数高效微调量化及内存高效的参数高效微调策略，旨在提升模型性能和计算效率的同时最大限度降低内存消耗，但相关研究在遥感领域较为匮乏。为此，设计高效计算的大型基础模型结构，突破遥感领域模型剪枝、量化与推理加速、下游任务高效适配等技术，实现遥感大模型的低成本训练、低内存消耗、高效实时推理和轻量化部署应用是未来发展的方向。

6 结 语

结合大模型和参数高效微调技术的遥感图像解译下游应用是当前的研究热点与发展方向。本文首先回顾了当前主流的参数高效微调方法并对基础模型的技术实现进行阐述，并提出PEFT运用于遥感领域时存在的几大方面问题。然后，根据最常用的参数高效微调方法，将遥感领域参数高效微调方法分为三大类的同时，分别从不同解译任务角度归纳总结了对应方法的精髓及异同点。其次，为进一步补充遥感图像解译任务的测评基准、评价指标和实验结果，本文从基本方法、实验数据、下游任务性能三大方面阐述并总结了30多种遥感领域的PEFT方法，开展的场景识别、语义分割、目标检测、变化检测等几大典型应用展现了PEFT方法在应用于遥感领域时的高精度、低参数量、强泛化能力特点。最后，基于对现存方法的归纳，讨论了其未来研究方向与发展挑战。

尽管视觉领域的参数高效微调方法在近年来得到了广泛关注，并取得了显著的研究进展，但为满足实际应用中对模型小容量内存占用以及大规模数据处理的需求，亟须进一步深入研究。本文希望通过对参数高效微调方法与遥感图像解译这一交叉应用进行综述，使得更多学者关注如何将遥感领域大模型结合参数高效微调技术赋能下游应用，进而适应未来遥感模型低成本训练、高效实时推理和轻量化部署的发展需求。

参考文献（References）

Bi H X, Gao Z W, Liu K, Song Q and Wang X T. 2023. Boosting few-shot remote sensing image scene classification with language-guided multimodal prompt tuning//Proceedings of 2023 International Conference on New Trends in Computational Intelligence. Qingdao, China: IEEE: 293-297 [DOI: 10.1109/NTCI60157.2023.10403750]

Chavan A, Liu Z, Gupta D, Xing E and Shen Z Q. 2023. One-for-all: generalized LoRA for parameter-efficient fine-tuning [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2306.07967.pdf

Chen H and Shi Z W. 2020. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. Remote Sensing, 12(10): #1662 [DOI: 10.3390/rs12101662]

Chen H, Tao R, Zhang H, Wang Y D, Li X, Ye W, et al. 2024c. Conv-

adapter: exploring parameter efficient transfer learning for ConvNets [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2208.07463.pdf

Chen K Y, Liu C Y, Chen H, Zhang H T, Li W Y, Zou Z X, et al. 2024a. RSPrompter: learning to prompt for remote sensing instance segmentation based on visual foundation model. IEEE Transactions on Geoscience and Remote Sensing, 62: #4701117 [DOI: 10.1109/TGRS.2024.3356074]

Chen K Y, Liu C Y, Li W Y, Liu Z L, Chen H, Zhang H T, et al. 2024b. Time travelling pixels: bitemporal features integration with foundation model for remote sensing image change detection//2024 IEEE International Geoscience and Remote Sensing Symposium. Athens, Greece: IEEE: 8581-8584 [DOI: 10.1109/IGARSS53475.2024.10640593]

Chen S F, Ge C J, Tong Z, Wang J L, Song Y B, Wang J, et al. 2022b. AdaptFormer: adapting vision transformers for scalable visual recognition [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2205.13535.pdf

Chen Y K, Qian S J, Tang H T, Lai X, Liu Z J, Han S, et al. 2024d. LongLoRA: efficient fine-tuning of long-context large language models [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2309.12307.pdf

Chen Z, Duan Y C, Wang W H, He J J, Lu T, Dai J F, et al. 2022a. Vision transformer adapter for dense predictions [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2205.08534.pdf

Cheng G, Han J W and Lu X Q. 2017. Remote sensing image scene classification: benchmark and state of the art. Proceedings of the IEEE, 105(10): 1865-1883 [DOI: 10.1109/JPROC.2017.2675998]

Cheng G, Han J W, Zhou P C and Guo L. 2014. Multi-class geospatial object detection and geographic image classification based on collection of part detectors. ISPRS Journal of Photogrammetry and Remote Sensing, 98: 119-132 [DOI: 10.1016/j.isprsjprs.2014.10.002]

Cheng G, Wang J B, Li K, Xie X X, Lang C B, Yao Y Q, et al. 2022. Anchor-free oriented proposal generator for object detection. IEEE Transactions on Geoscience and Remote Sensing, 60: #5625411 [DOI: 10.1109/TGRS.2022.3183022]

Chronopoulou A, Peters M E, Fraser A and Dodge J. 2023. AdapterSoup: Weight averaging to improve generalization of pretrained language models//Findings of the Association for Computational Linguistics: EACL 2023. Dubrovnik, Croatia: ACL: 2054-2063 [DOI: 10.18653/v1/2023.findings-eacl.153]

Demir I, Koperski K, Lindenbaum D, Pang G, Huang J, Basu S, et al. 2018. DeepGlobe 2018: a challenge to parse the earth through satellite images//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Salt Lake City, USA: IEEE: 172-181 [DOI: 10.1109/CVPRW.2018.00031]wei

Di Y H, Jiang Z G and Zhang H P. 2021. A public dataset for fine-grained ship classification in optical remote sensing images. Remote Sensing, 13(4): #747 [DOI: 10.3390/rs13040747]

Ding L, Zhu K, Peng D F, Tang H, Yang K W and Bruzzone L. 2024. Adapting segment anything model for change detection in VHR remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 62: #5611711 [DOI: 10.1109/TGRS.2024.3368168]

Dong Z, Gu Y F and Liu T Z. 2024. UPetu: a unified parameter-efficient fine-tuning framework for remote sensing foundation model. IEEE Transactions on Geoscience and Remote Sensing, 62: #5616613 [DOI: 10.1109/TGRS.2024.3382734]

Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X H, Unterthiner T, et al. 2021. An image is worth 16 × 16 words: transformers for image recognition at scale [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2010.11929.pdf

Edalati A, Tahaei M, Kobyzev I, Nia V P, Clark J J and Rezagholizadeh M. 2022. KronA: parameter efficient tuning with Kronecker adapter [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2212.10650.pdf

Fang L Y, Kuang Y, Liu Q, Yang Y and Yue J. 2023. Rethinking remote sensing pretrained model: instance-aware visual prompting for remote sensing scene classification. IEEE Transactions on Geoscience and Remote Sensing, 61: #5626713 [DOI: 10.1109/TGRS.2023.3336283]

Fang L Y, Kuang Y, Liu Q and Yue J. 2024. Temporal difference prompted SAM for remote sensing change detection. Journal of Signal Processing, 40(3): 417-427 (方乐缘, 旷洋, 刘强, 岳俊. 2024. 基于时差提示SAM的遥感变化检测. 信号处理, 40(3): 417-427) [DOI: 10.16798/j.issn.1003-0530.2024.03.001]

Feng W Q, Guan F L, Sun C H and Xu W. 2024. Road-SAM: adapting the segment anything model to road extraction from large very-high-resolution optical remote sensing images. IEEE Geoscience and Remote Sensing Letters, 21: #6012605 [DOI: 10.1109/LGRS.2024.3430900]

Gao K L, You X, Li K, Chen L Y, Lei J and Zuo X B. 2024. Attention prompt-driven source-free adaptation for remote sensing images semantic segmentation. IEEE Geoscience and Remote Sensing Letters, 21: #6012105 [DOI: 10.1109/LGRS.2024.3422805]

Guo X, Lao J W, Dang B, Zhang Y Y, Yu L, Ru L X, et al. 2024. SkySense: a multi-modal remote sensing foundation model towards universal interpretation for earth observation imagery//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: 27662-27673 [DOI: 10.1109/CVPR52733.2024.02613]

Han C, Wang Q F, Cui Y M, Cao Z W, Wang W G, Qi S Y, et al. 2023. E2VPT: an effective and efficient approach for visual prompt tuning//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 17445-17456 [DOI: 10.1109/ICCV51070.2023.01604]

Hayou S, Ghosh N and Yu B. 2024. LoRA+: efficient low rank adapta-

tion of large models [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2402.12354.pdf

He J X, Zhou C T, Ma X Z, Berg-Kirkpatrick T and Neubig G. 2022a. Towards a unified view of parameter-efficient transfer learning [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2110.04366.pdf

He K M, Chen X L, Xie S N, Li Y H, Dollár P and Girshick R. 2022b. Masked autoencoders are scalable vision learners//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 15979-15988 [DOI: 10.1109/CVPR52688.2022.01553]

He K M, Fan H Q, Wu Y X, Xie S N and Girshick R. 2020. Momentum contrast for unsupervised visual representation learning//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 9726-9735 [DOI: 10.1109/CVPR42600.2020.00975]

Helber P, Bischke B, Dengel A and Borth D. 2019. EuroSAT: a novel dataset and deep learning benchmark for land use and land cover classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12(7): 2217-2226 [DOI: 10.1109/JSTARS.2019.2918242]

Ho J, Jain A and Abbeel P. 2020. Denoising diffusion probabilistic models [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2006.11239.pdf

Houlsby N, Giurgiu A, Jastrz bski S, Morrone B, De Laroussilhe Q, Gesmundo A, et al. 2019. Parameter-efficient transfer learning for NLP [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/1902.00751.pdf

Hu E, Shen Y L, Wallis P, Allen-Zhu Z Y, Li Y Z, Wang S A, et al. 2021. LoRA: low-rank adaptation of large language models [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2106.09685.pdf

Hu K O, Wan H J, Ma F and Zhang F. 2024a. Cross-domain SAR object detection by efficiently fine-tuning SAM//2024 Photonics and Electromagnetics Research Symposium. Chengdu, China: IEEE: 1-7 [DOI: 10.1109/PIERS62282.2024.10618373]

Hu L Y, Lu W X, Yu H F, Yin D S, Sun X and Fu K. 2024c. TEA: a training-efficient adapting framework for tuning foundation models in remote sensing. IEEE Transactions on Geoscience and Remote Sensing, 62: #5648118 [DOI: 10.1109/TGRS.2024.3488731]

Hu L Y, Yu H F, Lu W X, Yin D S, Sun X and Fu K. 2024b. AiRs: adapter in remote sensing for parameter-efficient transfer learning. IEEE Transactions on Geoscience and Remote Sensing, 62: #5605218 [DOI: 10.1109/TGRS.2024.3351889]

Hu Y, Yuan J L, Wen C C, Lu X N and Li X. 2023a. RSGPT: a remote sensing vision language model and benchmark [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2307.15266.pdf

Huang C S, Liu Q, Lin B Y, Pang T Y, Du C and Lin M. 2024. LoraHub: efficient cross-task generalization via dynamic LoRA composition [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2307.13269.pdf

International Society for Photogrammetry and Remote Sensing. 2024. 2D semantic labeling-Vaihingen data [EB/OL]. [2025-01-11]. https://www.isprs.org/resources/datasets/benchmarks/UrbanSemLab/2d-sem-label-vaihingen.aspx

Ji D Y, Zhao F, Lu H T, Tao M Y and Ye J P. 2023. Ultra-high resolution segmentation with ultra-rich context: a novel benchmark//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 23621-23630 [DOI: 10.1109/CVPR52729.2023.02262]

Ji H, Gao Z, Ren J C, Wang X A, Gao T Y, Sun W B, et al. 2024. Prompting-to-distill semantic knowledge for few-shot learning. IEEE Geoscience and Remote Sensing Letters, 21: #6011605 [DOI: 10.1109/LGRS.2024.3414505]

Ji S P, Wei S Q and Lu M. 2019. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. IEEE Transactions on Geoscience and Remote Sensing, 57(1): 574-586 [DOI: 10.1109/TGRS.2018.2858817]

Jia M L, Tang L M, Chen B C, Cardie C, Belongie S, Hariharan B, et al. 2022. Visual prompt tuning//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 709-727 [DOI: 10.1007/978-3-031-19827-4_41]

Karimi Mahabadi R, Henderson J and Ruder S. 2021. Compacter: efficient low-rank hypercomplex adapter layers [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2106.04647.pdf

Khanna S, Liu P, Zhou L, Meng C, Rombach R, Burke M, et al. 2023. Diffusionsat: a generative foundation model for satellite imagery [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2312.03606.pdf

Khattak M U, Rasheed H, Maaz M, Khan S and Khan F S. 2023. MaPLe: multi-modal prompt learning//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 19113-19122 [DOI: 10.1109/CVPR52729.2023.01832]

Kirillov A, Mintun E, Ravi N, Mao H Z, Rolland C, Gustafson L, et al. 2023. Segment anything//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 3992-4003 [DOI: 10.1109/ICCV51070.2023.00371]

Lan L, Wang F X, Zheng X T, Wang Z M and Liu X W. 2025. Efficient prompt tuning of large vision-language model for fine-grained ship classification. IEEE Transactions on Geoscience and Remote Sensing, 63: #5600810 [DOI: 10.1109/TGRS.2024.3509721]

Lester B, Al-Rfou R and Constant N. 2021. The power of scale for parameter-efficient prompt tuning//Proceedings of 2021 Conference on Empirical Methods in Natural Language Processing. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics: 3045-3059 [DOI: 10.18653/v1/2021.emnlp-main.243]

Li C, Farkhoor H, Liu R and Yosinski J. 2018a. Measuring the intrinsic dimension of objective landscapes [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/1804.08838.pdf

Li H Z, Eills J G, Zhang L and Chang S F. 2018b. PatternNet: visual

pattern mining with deep neural network//Proceedings of 2018 ACM on International Conference on Multimedia Retrieval. Yokohama, Japan: IEEE: 291-299 [DOI: 10.1145/3206025.3206039]

Li J W, Qu C W and Shao J Q. 2017. Ship detection in SAR images based on an improved faster R-CNN//Proceedings of 2017 SAR in Big Data Era: Models, Methods and Applications. Beijing, China: IEEE: 1-6 [DOI: 10.1109/BIGSARDATA.2017.8124934]

Li K, Wan G, Cheng G, Meng L Q and Han J W. 2020a. Object detection in optical remote sensing images: a survey and a new benchmark. ISPRS Journal of Photogrammetry and Remote Sensing, 159: 296-307 [DOI: 10.1016/j.isprsjprs.2019.11.023]

Li X L and Liang P. 2021. Prefix-Tuning: optimizing continuous prompts for generation//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Online: Association for Computational Linguistics: 4582-4597 [DOI: 10.18653/v1/2021.acl-long.353]

Li Y H, Yang J and Wang J L. 2020b. DyLoRa: towards energy efficient dynamic LoRa transmission control//Proceedings of 2020 IEEE Conference on Computer Communications. Toronto, Canada: IEEE: 2312-2320 [DOI: 10.1109/INFOCOM41043.2020.9155407]

Li Y S, Wang L L, Wang T Z, Yang X, Luo J W, Wang Q, et al. 2025. STAR: a first-ever dataset and a large-scale benchmark for scene graph generation in large-size satellite imagery. IEEE Transactions on Pattern Analysis and Machine Intelligence, 47(3): 1832-1849 [DOI: 10.1109/TPAMI.2024.3508072]

Lin Z J, Madotto A and Fung P. 2020. Exploring versatile generative language model via parameter-efficient transfer learning [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2004.03829.pdf

Liu S Y, Wang C Y, Yin H X, Molchanov P, Wang Y C F, Cheng K T, et al. 2024. DoRA: weight-decomposed low-rank adaptation [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2402.09353.pdf

Liu X, Ji K X, Fu Y C, Tam W L, Du Z X, Yang Z L, et al. 2022. P-Tuning v2: prompt tuning can be comparable to fine-tuning universally across scales and tasks [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2110.07602.pdf

Liu Z K, Yuan L, Weng L B and Yang Y P. 2017. A high resolution optical satellite image dataset for ship recognition and some new baselines//Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods. Porto, Portugal: SciTePress: 324-331

Lu X N, Diao W H, Li J X, Zhang Y D, Wang P J, Sun X, et al. 2024. Few-shot incremental object detection in aerial imagery via dual-frequency prompt. IEEE Transactions on Geoscience and Remote Sensing, 62: #5624017 [DOI: 10.1109/TGRS.2024.3400595]

Lu X Q, Wang B Q, Zheng X T and Li X L. 2018. Exploring models and data for remote sensing image caption generation. IEEE Transactions on Geoscience and Remote Sensing, 56(4): 2183-2195 [DOI: 10.1109/TGRS.2017.2776321]

Luo M Y, Zhang T, Wei S Q and Ji S P. 2024. SAM-RSIS: progressively adapting SAM with box prompting to remote sensing image instance segmentation. IEEE Transactions on Geoscience and Remote Sensing, 62: #4413814 [DOI: 10.1109/TGRS.2024.3460085]

Maggiori E, Tarabalka Y, Charpiat G and Alliez P. 2017. Can semantic labeling methods generalize to any city? The Inria aerial image labeling benchmark//2017 IEEE International Geoscience and Remote Sensing Symposium. Fort Worth, USA: IEEE: 3226-3229 [DOI: 10.1109/IGARSS.2017.8127684]

Mao Y N, Mathias L, Hou R, Almahairi A, Ma H, Han J W, et al. 2022. UniPELT: a unified framework for parameter-efficient language model tuning//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Dublin, Ireland: ACL: 6253-6264 [DOI: 10.18653/v1/2022.acl-long.433]

Mei L Y, Ye Z Y, Xu C, Wang H Z, Wang Y, Lei C, et al. 2024. SCD-SAM: adapting segment anything model for semantic change detection in remote sensing imagery. IEEE Transactions on Geoscience and Remote Sensing, 62: #5626713 [DOI: 10.1109/TGRS.2024.3407884]

Mohajerani S and Saeedi P. 2019. Cloud-net: an end-to-end cloud detection algorithm for Landsat 8 imagery//2019 IEEE International Geoscience and Remote Sensing Symposium. Yokohama, Japan: IEEE: 1029-1032 [DOI: 10.1109/IGARSS.2019.8898776]

Nie X, Ni B L, Chang J L, Meng G F, Huo C L, Xiang S M, et al. 2024. Pro-tuning: unified prompt tuning for vision tasks. IEEE Transactions on Circuits and Systems for Video Technology, 34(6): 4653-4667 [DOI: 10.1109/TCSVT.2023.3327605]

Pfeiffer J, Kamath A, Rücklé A, Cho K and Gurevych I. 2021. AdapterFusion: non-destructive task composition for transfer learning//Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume. Online: ACL: 487-503 [DOI: 10.18653/v1/2021.eacl-main.39]

Pu X Y, Jia H C, Zheng L H, Wang F and Xu F. 2024. ClassWise-SAM-adapter: parameter efficient fine-tuning adapts segment anything to SAR domain for semantic segmentation [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2401.02326.pdf

Pu X Y and Xu F. 2025. Low-rank adaption on transformer-based oriented object detector for satellite onboard processing of remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 63: #5202213 [DOI: 10.1109/TGRS.2024.3524578]

Qi X M, Zhu P P, Wang Y B, Zhang L Q, Peng J H, Wu M F, et al. 2020. MLRSNet: a multi-label high spatial resolution remote sensing dataset for semantic scene understanding. ISPRS Journal of Photogrammetry and Remote Sensing, 169: 337-350 [DOI: 10.1016/j.isprsjprs.2020.09.020]

Radford A, Kim J W, Hallacy C, Ramesh A, Goh G, Agarwal S, et al.

2021. Learning transferable visual models from natural language supervision//Proceedings of the 38th International Conference on Machine Learning. Virtually: PMLR: 8748-8763

Roy S and Etemad A. 2024. Consistency-guided prompt learning for vision-language models [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2306.01195.pdf

Scheibenreif L, Hanna J, Mommert M and Borth D. 2022. Self-supervised vision transformers for land-cover segmentation and classification//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. New Orleans, USA: IEEE: 1421-1430 [DOI: 10.1109/CVPRW56347.2022.00148]

Shen L, Lu Y, Chen H, Wei H, Xie D H, Yue J B, et al. 2021. S2Looking: a satellite side-looking dataset for building change detection. Remote Sensing, 13 (24): #5094 [DOI: 10.3390/rs13245094]

Si C J, Wang X H, Yang X, Xu Z Q, Li Q Y, Dai J F, et al. 2025. Maintaining structural integrity in parameter spaces for parameter efficient fine-tuning [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2405.14739.pdf

Singha M, Jha A, Solanki B, Bose S and Banerjee B. 2023. APPLeNet: visual attention parameterized prompt learning for few-shot remote sensing image generalization using CLIP//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Vancouver, Canada: IEEE: 2024-2034 [DOI: 10.1109/CVPRW59228.2023.00196]

Su Y S, Wang X Z, Qin Y J, Chan C M, Lin Y K, Wang H D, et al. 2022. On transferability of prompt tuning for natural language processing//Proceedings of 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Seattle, United States: Association for Computational Linguistics: 3949-3969 [DOI: 10.18653/v1/2022.naacl-main.290]

Sun X, Wang P J, Lu W X, Zhu Z C, Lu X N, He Q B, et al. 2023. RingMo: a remote sensing foundation model with masked image modeling. IEEE Transactions on Geoscience and Remote Sensing, 61: #5612822 [DOI: 10.1109/TGRS.2022.3194732]

Sun X, Wang P J, Yan Z Y, Xu F, Wang R P, Diao W H, et al. 2022. FAIR1M: a benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery. ISPRS Journal of Photogrammetry and Remote Sensing, 184: 116-130 [DOI: 10.1016/j.isprsjprs.2021.12.004]

Sun X R, Liu J, Shen H T, Zhu X F and Hu P. 2025. On efficient variants of segment anything model: a survey [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2410.04960.pdf

Tu C H, Mai Z and Chao W L. 2023. Visual query tuning: towards effective usage of intermediate representations for parameter and memory efficient transfer learning//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 7725-7735 [DOI: 10.1109/CVPR52729.2023.00746]

Wang D, Zhang Q M, Xu Y F, Zhang J, Du B, Tao D C, et al. 2023. Advancing plain vision transformer toward remote sensing foundation model. IEEE Transactions on Geoscience and Remote Sensing, 61: #5607315 [DOI: 10.1109/TGRS.2022.3222818]

Wang H X, Chang J L, Zhai Y H, Luo X, Sun J N, Lin Z C, et al. 2024a. LION: implicit vision prompt tuning//Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI: 5372-5380 [DOI: 10.1609/aaai.v38i6.28345]

Wang J J, Zheng Z, Ma A L, Lu X Y and Zhong Y F. 2022c. LoveDA: a remote sensing land-cover dataset for domain adaptive semantic segmentation [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2110.08733.pdf

Wang M M, Zhou L, Zhang K Y, Li X H, Hao M and Ye Y X. 2024b. ESAM-CD: fine-tuned efficient SAM network with LoRA for weakly supervised remote sensing image change detection. IEEE Transactions on Geoscience and Remote Sensing, 62: #4708616 [DOI: 10.1109/TGRS.2024.3470808]

Wang Q X, Yin J H, Jiang H X, Feng J Q and Zhang G Y. 2024c. Disentangled foreground-semantic adapter network for generalized aerial image few-shot semantic segmentation. IEEE Transactions on Geoscience and Remote Sensing, 62: #5641114 [DOI: 10.1109/TGRS.2024.3455427]

Wang Y, Mukherjee S, Liu X, Gao J, Awadallah A H and Gao J. 2022a. Adamix: mixture-of-adapter for parameter-efficient tuning of large language models [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2205.12410.pdf

Wang Z C, Prabha R, Huang T Y, Wu J J and Rajagopal R. 2024d. SkyScript: a large and semantically diverse vision-language dataset for remote sensing//Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI: 5805-5813 [DOI: 10.1609/aaai.v38i6.28393]

Waqas Zamir S, Arora A, Gupta A, Khan S, Sun G L, Shahbaz Khan F, et al. 2019. iSAID: a large-scale dataset for instance segmentation in aerial images [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/1905.12886.pdf

Wei S J, Zeng X F, Qu Q Z, Wang M, Su H and Shi J. 2020. HRSID: a high-resolution SAR images dataset for ship detection and instance segmentation. IEEE Access, 8: 120234-120254 [DOI: 10.1109/ACCESS.2020.3005861]

Wei S Q, Zhang T, Ji S P, Luo M Y and Gong J Y. 2023. BuildMapper: a fully learnable framework for vectorized building contour extraction. ISPRS Journal of Photogrammetry and Remote Sensing, 197: 87-104 [DOI: 10.1016/j.isprsjprs.2023.01.015]

Wu W H, Wong M S, Yu X Y, Shi G Q, Tung Kwok C Y and Zou K. 2024. Compositional oil spill detection based on object detector and adapted segment anything model from SAR images. IEEE Geoscience and Remote Sensing Letters, 21: #4007505 [DOI: 10.1109/LGRS.2024.3382970]

Xia G S, Bai X, Ding J, Zhu Z, Belongie S, Luo J B, et al. 2018. DOTA: a large-scale dataset for object detection in aerial images// Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 3974-3983 [DOI: 10.1109/CVPR.2018.00418]

Xia G S, Hu J W, Hu F, Shi B G, Bai X, Zhong Y F, et al. 2017. AID: a benchmark data set for performance evaluation of aerial scene classification. IEEE Transactions on Geoscience and Remote Sensing, 55(7): 3965-3981 [DOI: 10.1109/TGRS.2017.2685945]

Xing J L, Liu J P, Wang J, Sun L L, Chen X, Gu X X, et al. 2024. A survey of efficient fine-tuning methods for vision-language models — prompt and adapter. Computers and Graphics, 119: #103885 [DOI: 10.1016/j.cag.2024.01.012]

Xiong Y Y, Varadarajan B, Wu L M, Xiang X Y, Xiao F Y, Zhu C C, et al. 2024. EfficientSAM: leveraged masked image pretraining for efficient segment anything//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 16111-16121 [DOI: 10.1109/CVPR52733.2024.01525]

Xue B W, Cheng H, Yang Q Q, Wang Y and He X N. 2024. Adapting segment anything model to aerial land cover classification with low-rank adaptation. IEEE Geoscience and Remote Sensing Letters, 21: #2502605 [DOI: 10.1109/LGRS.2024.3357777]

Yang A X, Robeyns M, Wang X and Aitchison L. 2024a. Bayesian low-rank adaptation for large language models [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2308.13111.pdf

Yang B H, Chen Y S and Ghamisi P. 2024b. LVM-StARS: large vision model soft adaption for remote sensing scene classification. IEEE Geoscience and Remote Sensing Letters, 21: #6013905 [DOI: 10.1109/LGRS.2024.3432069]

Yang J and Huang X. 2021. The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019. Earth System Science Data, 13(8): 3907-3925 [DOI: 10.5194/essd-13-3907-2021]

Yang K P, Xia G S, Liu Z C, Du B, Yang W, Pelillo M, et al. 2022. Asymmetric Siamese networks for semantic change detection in aerial images. IEEE Transactions on Geoscience and Remote Sensing, 60: #5609818 [DOI: 10.1109/TGRS.2021.3113912]

Yang Y and Newsam S. 2010. Bag-of-visual-words and spatial extensions for land-use classification//Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems. San Jose, USA: ACM: 270-279 [DOI: 10.1145/1869790.1869829]

Yao H T, Zhang R and Xu C S. 2023. Visual-language prompt tuning with knowledge-guided context optimization//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 6757-6767 [DOI: 10.1109/CVPR52729.2023.00653]

Yin D S, Hu L Y, Li B, Zhang Y Q and Yang X. 2024. 5% > 100% breaking performance shackles of full fine-tuning on visual recognition tasks [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2408.08345.pdf

Yin D S, Yang Y R, Wang Z C, Yu H F, Wei K W and Sun X. 2023. 1% vs 100%: parameter-efficient low rank adapter for dense predictions//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 20116-20126 [DOI: 10.1109/CVPR52729.2023.01926]

Yin W X, Yu H C, Diao W H, Sun X and Fu K. 2024. Parameter efficient fine-tuning of vision transformers for remote sensing scene understanding. Journal of Electronics and Information Technology, 46(9): 3731-3738 (尹文昕, 于海琛, 刁文辉, 孙显, 付琨. 2024. 遥感场景理解中视觉Transformer的参数高效微调. 电子与信息学报, 46(9): 3731-3738) [DOI: 10.11999/JEIT240218]

Yoo S, Kim E, Jung D, Lee J and Yoon S. 2023. Improving visual prompt tuning for self-supervised vision transformers//Proceedings of the 40th International Conference on Machine Learning, Honolulu, USA: PMLR: 40075-40092

Yuan P L, Zhao Q Z, Zhao X B, Wang X W, Long X F and Zheng Y C. 2022. A transformer-based Siamese network and an open optical dataset for semantic change detection of remote sensing images. International Journal of Digital Earth, 15(1): 1506-1525 [DOI: 10.1080/17538947.2022.2111470]

Zang Y H, Li W, Zhou K Y, Huang C and Loy C C. 2022. Unified vision and language prompt learning [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2210.07225.pdf

Zhang C N, Han D S, Qiao Y, Kim J U, Bae S H, Lee S, et al. 2023a. Faster segment anything: towards lightweight SAM for mobile applications [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2306.14289.pdf

Zhang J, Yang X B, Jiang R, Shao W and Zhang L. 2024a. RSAM-Seg: a sam-based approach with prior knowledge integration for remote sensing image semantic segmentation [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2402.19004.pdf

Zhang J J, Rao Y T, Huang X S, Li G Y, Zhou X and Zeng D. 2024b. Frequency-aware multi-modal fine-tuning for few-shot open-set remote sensing scene classification. IEEE Transactions on Multimedia, 26: 7823-7837 [DOI: 10.1109/TMM.2024.3372416]

Zhang Q R, Chen M S, Bukharin A, Karampatziakis N, He P C, Cheng Y, et al. 2023b. AdaLoRA: adaptive budget allocation for parameter-efficient fine-tuning [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2303.10512.pdf

Zhang S H and Pan Z G. 2025. Remote sensing large models: review and future prospects. Remote Sensing Technology and Application, 40(1): 1-13 (张帅豪, 潘志刚. 2025. 遥感大模型: 综述与未来设想. 遥感技术与应用, 40(1): 1-13) [DOI: 10.11873/j.issn.1004-0323.2025.1.0001]

Zhang X H, Lyu Y F, Yao L B, Xiong W and Fu C L. 2020. A new benchmark and an attribute-guided multilevel feature representation network for fine-grained ship classification in optical remote sensing images. IEEE Journal of Selected Topics in Applied Earth

Observations and Remote Sensing, 13: 1271-1285 [DOI: 10.1109/JSTARS.2020.2981686]

Zhang Y, Zhang C, Yu K, Tang Y S and He Z H. 2024c. Concept-guided prompt learning for generalization in vision-language models//Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI: 7377-7386 [DOI: 10.1609/aaai.v38i7.28568]

Zhang Y H, Zhou K Y and Liu Z W. 2022. Neural prompt search [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2206.04673.pdf

Zhang Y J, Li Y S, Dang B, Wu K, Guo X, Wang J, et al. 2024. Multimodal remote sensing large foundation models: current research status and future prospect. Acta Geodaetica et Cartographica Sinica, 53(10): 1942-1954 (张永军, 李彦胜, 党博, 武康, 郭昕, 王剑, 等. 2024. 多模态遥感基础大模型: 研究现状与未来展望. 测绘学报, 53(10): 1942-1954) [DOI: 10.11947/j.AGCS.2024.20240019]

Zhao F and Zhang C C. 2024. Parameter-efficient adaptation of foundation models for damaged building assessment//Proceedings of the 7th IEEE International Conference on Multimedia Information Processing and Retrieval. San Jose, USA: IEEE: 417-422 [DOI: 10.1109/MIPR62202.2024.00072]

Zhao L, Xu L R, Zhao L, Zhang X L, Wang Y H, Ye D Q, et al. 2023a. Continual learning for remote sensing image scene classification with prompt learning. IEEE Geoscience and Remote Sensing Letters, 20: #6012005 [DOI: 10.1109/LGRS.2023.3328981]

Zhao X, Ding W C, An Y Q, Du Y L, Yu T, Li M, et al. 2023b. Fast segment anything [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2306.12156.pdf

Zheng L H, Pu X Y, Zhang S and Xu F. 2025. Tuning a SAM-based model with multicognitive visual adapter to remote sensing instance segmentation. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 18: 2737-2748 [DOI: 10.1109/JSTARS.2024.3504409]

Zheng X T, Xiao X L, Chen X M, Lu W X, Liu X Y and Lu X Q. 2024. Advancements in cross-domain remote sensing scene interpretation. Journal of Image and Graphics, 29(6): 1730-1746 (郑向涛, 肖欣林, 陈秀妹, 卢宛萱, 刘小煜, 卢孝强. 2024. 跨域遥感场景解译研究进展. 中国图象图形学报, 29(6): 1730-1746) [DOI: 10.11834/jig.240009]

Zhou H, Wan X C, Vulić I and Korhonen A. 2024. AutoPEFT: automatic configuration search for parameter-efficient fine-tuning [EB/OL]. [2025-01-11]. https://arxiv.org/pdf/2301.12132.pdf

Zhou K Y, Yang J K, Loy C C and Liu Z W. 2022a. Learning to prompt for vision-language models. International Journal of Computer Vision, 130(9): 2337-2348 [DOI: 10.1007/s11263-022-01653-1]

Zhou K Y, Yang J K, Loy C C and Liu Z W. 2022b. Conditional prompt learning for vision-language models//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 16795-16804 [DOI: 10.1109/CVPR52688.2022.01631]

Zhu H G, Chen X G, Dai W Q, Fu K, Ye Q X and Jiao J B. 2015. Orientation robust object detection in aerial images using deep convolutional neural network//Proceedings of 2015 IEEE International Conference on Image Processing. Quebec City, Canada: IEEE: 3735-3739 [DOI: 10.1109/ICIP.2015.7351502]

Zhu J J, Li Y Y, Yang K, Guan N Y, Fan Z L, Qiu C P, et al. 2024. MVP: meta visual prompt tuning for few-shot remote sensing image scene classification. IEEE Transactions on Geoscience and Remote Sensing, 62: #5610413 [DOI: 10.1109/TGRS.2024.3359599]

Zhu J W, Lai S M, Chen X, Wang D and Lu H C. 2023. Visual prompt multi-modal tracking//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 9516-9526 [DOI: 10.1109/CVPR52729.2023.00918]

Zhu Y M, Feng J T, Zhao C Q, Wang M X and Li L. 2021. Counter-interference adapter for multilingual machine translation//Findings of the Association for Computational Linguistics: EMNLP 2021. Punta Cana, Dominican Republic: ACL: 2812-282 [DOI: 10.18653/v1/2021.findings-emnlp.240]

作者简介

陈诗琪,女,助理研究员,主要研究方向为遥感图像解译、SAR图像智能感知与处理。E-mail:chenshiqi12@nudt.edu.cn
朱荣强,通信作者,男,博士后,主要研究方向为智能信息处理、通信感知一体化设计。E-mail:zhurq91@163.com
杨学,男,助理教授,主要研究方向为基础视觉、多模态大模型和遥感影像解译。E-mail:yangxue-2019-sjtu@sjtu.edu.cn
廖宁,男,博士后,主要研究方向为多模态遥感大模型应用。E-mail:liaoning@sjtu.edu.cn
赵卫伟,男,研究员,主要研究方向为指挥控制系统及应用。E-mail:zhaozww@163.com