# A_Unified_Framework_With_Multimodal_Fine-Tuning_for_Remote_Sensing_Semantic_Segmentation

# 全文摘要

## 全文概述

本文提出了一种基于多模态微调的统一框架，用于遥感语义分割任务。针对传统方法依赖单一模态数据导致的泛化能力不足问题，研究团队创新性地将 Segment Anything Model (SAM)与多模态融合技术结合，构建了 MFNet 网络。该框架通过引入多模态适配器(MMAdapter)和多模态低秩适配器(MMLoRA)两种微调机制，在保留 SAM 通用视觉知识的同时，有效融合光学图像与数字表面模型(DSM)等多源数据。核心创新点包括：1) 设计金字塔深度融合模块(DFM)，通过多尺度特征融合增强高阶语义表示；2) 验证 SAM 在 DSM 数据上的泛化能力，首次实现基础模型在非光学遥感数据中的应用；3) 提出可扩展的统一框架，支持任意模态扩展且无需修改 SAM 核心结构。实验表明，MFNet 在 ISPRS Vaihingen、Potsdam 和 MMHunan 三个基准数据集上均取得 SOTA 性能，其中 ViT-H 版本在 Vaihingen 数据集上达到 92.97%的 OA 和 85.03%的 mIoU。研究还揭示了 LoRA 参数压缩与信息丢失的权衡关系，为后续研究提供了重要参考。

## 术语解释

1. **SAM (Segment Anything Model)**: Meta AI 开发的通用分割模型，通过大规模自然图像预训练获得跨领域泛化能力，包含 ViT 图像编码器、提示编码器和掩码解码器三部分，是本文多模态框架的基础。

2. **MFNet (Multimodal Fine-Tuning Network)**: 本文提出的多模态语义分割网络，融合 SAM 图像编码器与 DFM 模块，通过 MMAdapter/MMLoRA 实现多源数据特征融合，支持光学图像与 DSM 等多模态输入。
3. **DFM (Deep Fusion Module)**: 基于金字塔结构的多尺度特征融合模块，通过 SE 融合单元整合 SAM 编码器输出的多层级特征，增强模型对遥感数据复杂特性的表征能力。

# 论文速读

## 论文方法

## 方法描述

该论文提出了一种统一的多模态微调框架，包括 MMAdapter 和 MMLoRA 两个模块，并应用于图像分割任务中。其中，MMAdapter 是一种双分支结构，通过共享权重处理多个输入通道的信息，并实现跨通道信息融合；而 MMLoRA 则是在标准 LoRA 的基础上扩展到多模态任务上，同样采用双分支结构，但可以在单个线性层内实现多个通道之间的交互。

## 方法改进

与传统的单模态微调策略相比，MMAdapter 和 MMLoRA 都采用了更加灵活的参数设置方式，可以针对不同的任务和数据集进行调整。此外，它们还可以有效地捕捉多模态信息中的关键特征，从而提高模型的性能。

## 解决的问题

该论文主要解决了在多模态图像分割任务中如何更好地利用不同来源的数据来提高模型性能的问题。通过引入 MMAdapter 和 MMLoRA 这两个模块，使得模型能够更好地学习多模态信息之间的关系，从而提高了模型的准确性和鲁棒性。

## 论文实验

本文主要介绍了针对多模态遥感数据的语义分割任务所设计的 MFNet 模型，并通过与现有的 15 种方法进行比较来验证其性能。在实验中，作者使用了三个不同的数据集（Vaihingen、Potsdam 和 MMHunan）来进行评估，每个数据集都包含了多个类别的遥感图像。作者使用了整体准确率（OA）、平均交并比（mF1）和平均交并面积（mIoU）等标准指标来衡量模型的性能。

在对 Vaihingen 数据集的实验中，MFNet 相比于其他方法取得了更好的结果，包括更高的 OA、mF1 和 mIoU 得分。具体来说，MFNet 使用 ViT-H 作为基础模

型时，在四个特定类别上表现更好，分别是建筑物、树木、低植被和硬质表面。此外，作者还进行了可视化分析，结果显示 MFNet 能够生成更清晰、更精确的边界，从而更好地分离地面物体。

在对 Potsdam 数据集的实验中，MFNet 同样取得了更好的结果，包括更高的 OA、mF1 和 mIoU 得分。此外，作者还观察到 MFNet 在识别树和低植被的同时存在一些挑战，这可能是由于它们具有相似的特征以及交错或重叠的分布导致的。因此，将更多的专业知识应用于这些难以区分的类别是一个有趣的未来方向。

在对 MMHunan 数据集的实验中，MFNet 的表现也很好，相比 MultiSenseSeg 取得了更高的 OA、mF1 和 mIoU 得分。然而，作者发现，在这个数据集中，较大的基础模型可能会更容易过拟合，因此选择适当的基础模型对于不同类型的遥感场景非常重要。

最后，作者还进行了几个额外的实验来验证 MFNet 的不同组成部分的有效性。例如，他们发现仅使用单模式数据而不进行任何微调机制会导致性能下降，而使用标准的 Adapter/LoRA 机制可以有效地提取遥感多模态特征。此外，作者还探讨了数据量对模型性能的影响，并发现在训练数据量达到一定比例后，进一步增加训练数据对下游性能的提升效果会逐渐减弱。

总的来说，本文通过与其他现有方法的比较，证明了 MFNet 在多模态遥感数据上的有效性和优越性。此外，作者还提供了一些有价值的见解，如如何选择适当的基础模型以及如何优化数据量以提高模型性能。

# 关键图表解读

# 关键图表解读

**图 1**：传统任务特定模型与 SAM 在遥感任务中的知识差异对比。传统模型受限于特定任务数据，而 SAM 通过大规模自然图像预训练获得通用视觉知识，但缺乏多模态遥感任务适配框架。

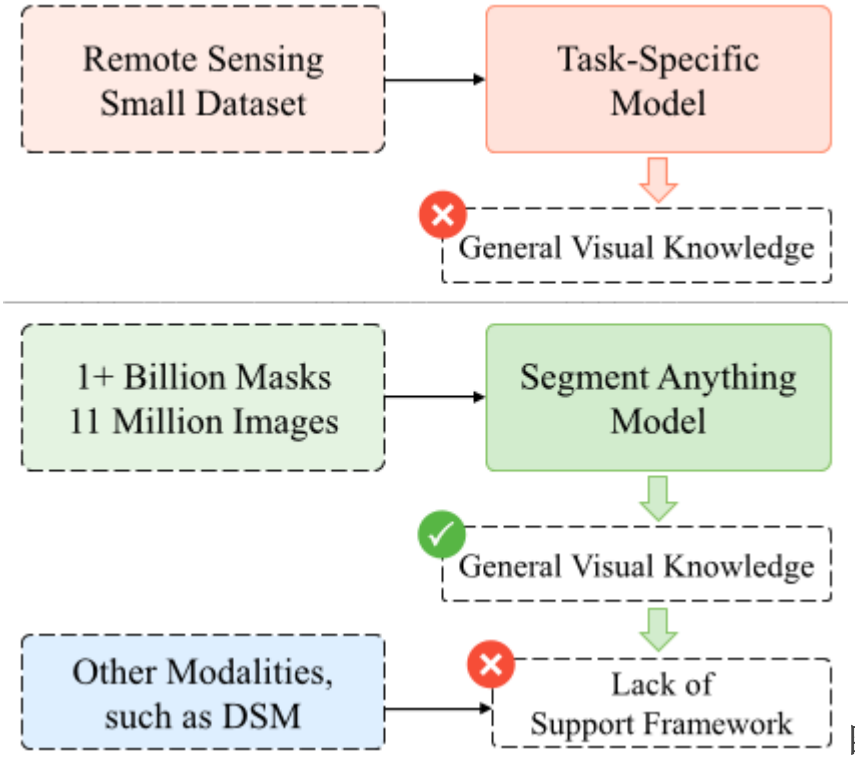
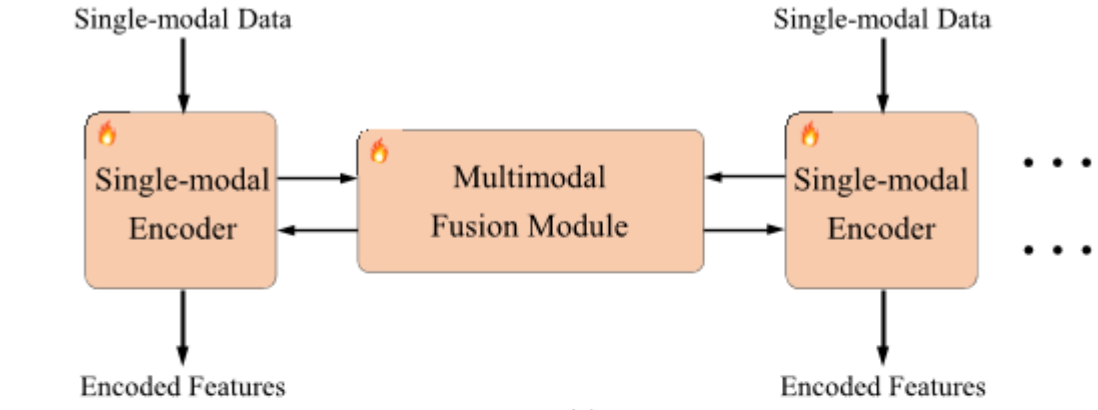



**图 2**：多模态融合方法对比与统一框架设计。传统方法为每种模态分配独立编码器并联合训练，而本文提出的统一框架通过冻结 SAM 编码器参数，仅优化多模态适配模块实现高效融合。



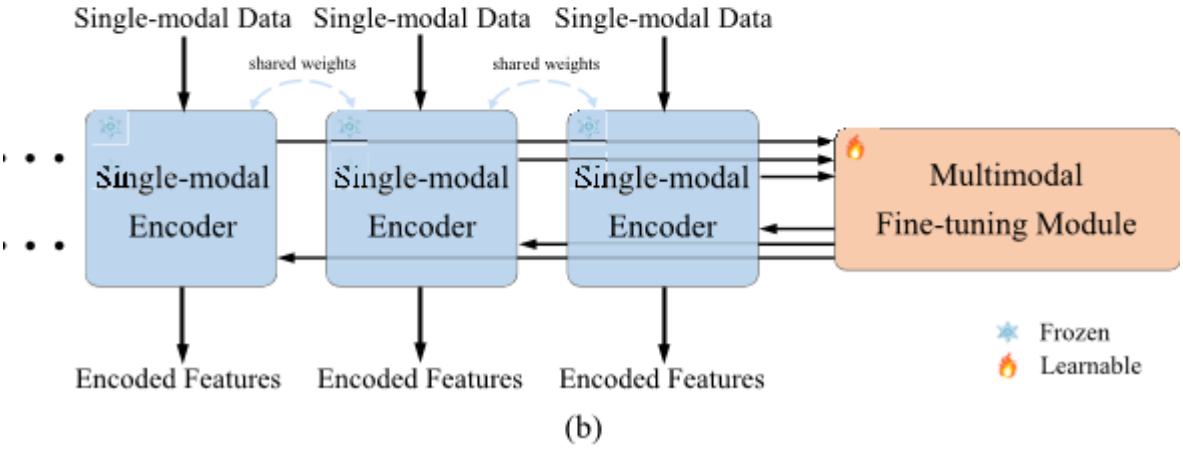
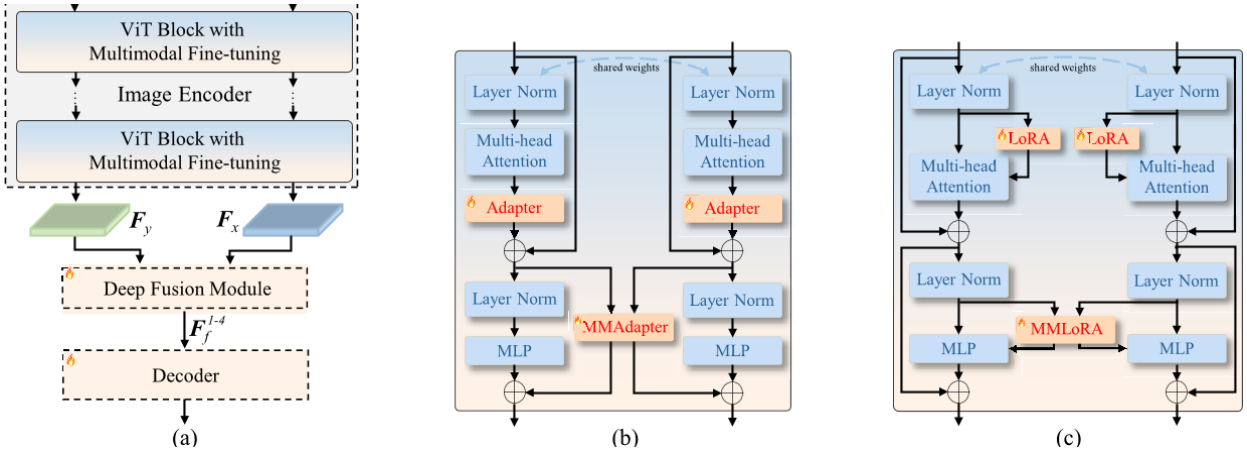



**图 5**：MFNet 架构详解。展示基于 MMAdapter/MMLoRA 的 SAM 编码器改造、金字塔深度融合模块（DFM）及通用解码器的完整网络结构，突出其模块化设计与参数高效性。




# 论文总结

# 关键图表解读

**图 1**：传统任务特定模型与 SAM 在遥感任务中的知识差异对比。传统模型受限于特定任务数据，而 SAM 通过大规模自然图像预训练获得通用视觉知识，但缺乏多模态遥感任务适配框架。

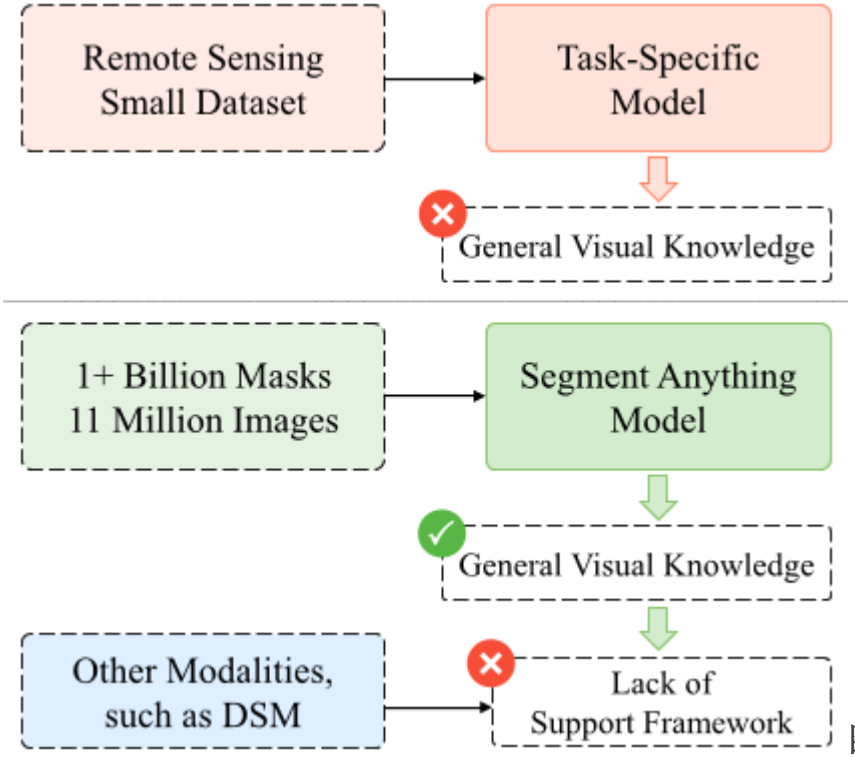



**图 2**：多模态融合方法对比与统一框架设计。传统方法为每种模态分配独立编码器并联合训练，而本文提出的统一框架通过冻结 SAM 编码器参数，仅优化多模态适配模块实现高效融合。

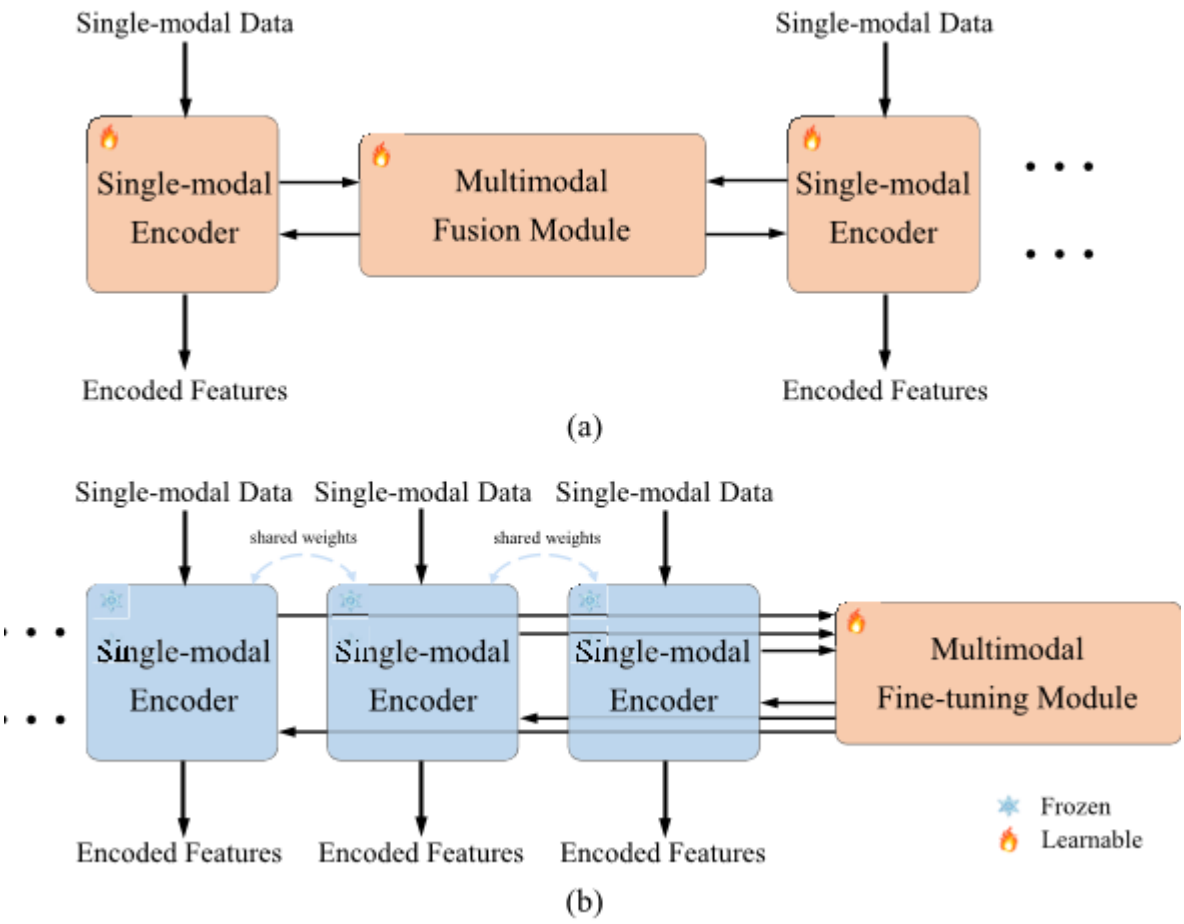



**图 5**：MFNet 架构详解。展示基于 MMAdapter/MMLoRA 的 SAM 编码器改造、金字塔深度融合模块（DFM）及通用解码器的完整网络结构，突出其模块化设计与参数高效性。