# CrossEarth-Gate: Fisher-Guided Adaptive Tuning Engine for Efficient Adaptation of Cross-Domain Remote Sensing Semantic Segmentation

Shilei Cao[1,4*], Ziyang Gong[3*], Hehai Lin[7], Yang Liu[2], Jiashun Cheng[5], Xiaoxing Hu[6], Haoyuan Liang[1,4], Guowen Li[1,4], Chengwei Qin[7], Hong Cheng[2], Xue Yang[3], Juepeng Zheng[1,4†], Haohuan Fu[4,8]

[1]Sun Yat-sen University, [2]The Chinese University of Hong Kong, [3]Shanghai Jiao Tong University, [4]National Supercomputing Center in Shenzhen, [5]The Hong Kong University of Science and Technology, [6]Beijing Institute of Technology, [7]The Hong Kong University of Science and Technology (Guangzhou), [8]Tsinghua University

## Abstract

*In Remote Sensing (RS), Parameter-Efficient Fine-Tuning (PEFT) has emerged as a key approach to activate the generalizable representation ability of foundation models for downstream tasks. However, existing specialized PEFT methods often fail when applied to large-scale Earth observation tasks, as they are unable to fully handle the multifaceted and unpredictable domain gaps (*e.g.*, spatial, semantic, and frequency shifts) inherent in RS data. To overcome this, we propose CrossEarth-Gate, which introduces two primary contributions. First, we establish a comprehensive RS module toolbox to address multifaceted domain gaps, comprising spatial, semantic, and frequency modules. Second, we develop a Fisher-guided adaptive selection mechanism that operates on this toolbox. This selection is guided by Fisher Information to quantify each module's importance by measuring its contribution to the task-specific gradient flow. It dynamically activates only the most critical modules at the appropriate layers, guiding the gradient flow to maximize adaptation effectiveness and efficiency. Comprehensive experiments validate the efficacy and generalizability of our method, where CrossEarth-Gate achieves state-of-the-art performance on 16 out of 18 cross-domain benchmarks for RS semantic segmentation.*

## 1. Introduction

With the rapid expansion of Geospatial Foundation Models (GFMs) in both capability and scale [9, 20, 48, 56], a central challenge emerges: how to efficiently activate their downstream potential. Parameter-Efficient Fine-Tuning (PEFT) [74, 81] methods aim to match or surpass full fine-tuning while updating only a small fraction of parameters, offering an attractive balance between performance and efficiency. However, when applied to large-scale and globally distributed Earth observation tasks, existing PEFT methods often suffer substantial performance degradation due to highly variable and unpredictable domain gaps [33, 40, 43, 68].

Domain gaps in Remote Sensing (RS) are multifaceted, arising from diverse sources like wavelength ranges, geographical landscapes, and climatic zones. These gaps manifest as a complex interplay of adaptation challenges: (1) **Spatial shifts**, representing the change in object structure and scale that require geometric integrity; (2) **Semantic shifts**, denoting the differences in class appearance and concepts; and (3) **Frequency shifts**, involving high-frequency spectral artifacts or textural noise from different features. However, as shown in Fig. 1 (a), existing PEFTs, whether general-purpose or RS-specific, typically specialize in one functional pathway. For example, LoRA [29] modifies the Multi-Head Self-Attention (MSA) layers to enhance spatial dependency modeling. AdaptFormer [5] adjusts the Multi-Layer Perception (MLP) layers to refine high-level semantic features. The Earth-Adapter [33] focuses on mitigating high-amplitude artifacts in the frequency domain.

While each method has its merits, this single-pathway design is inherently limited: it captures only one facet of RS domain shifts while leaving others unaddressed. As shown by the segmentation maps in Fig. 1(a), LoRA misclassifies large areas of water with high-frequency waves as forest, revealing its inability to process unfamiliar frequency artifacts. AdaptFormer fails to maintain spatial continuity of the road, shattering its structure and demonstrating a critical failure in spatial awareness. Earth-Adapter falters on semantic ambiguities and misidentifies coastal waves as for-

*Equal contribution.
†Corresponding author: zhengjp8@mail.sysu.edu.cn

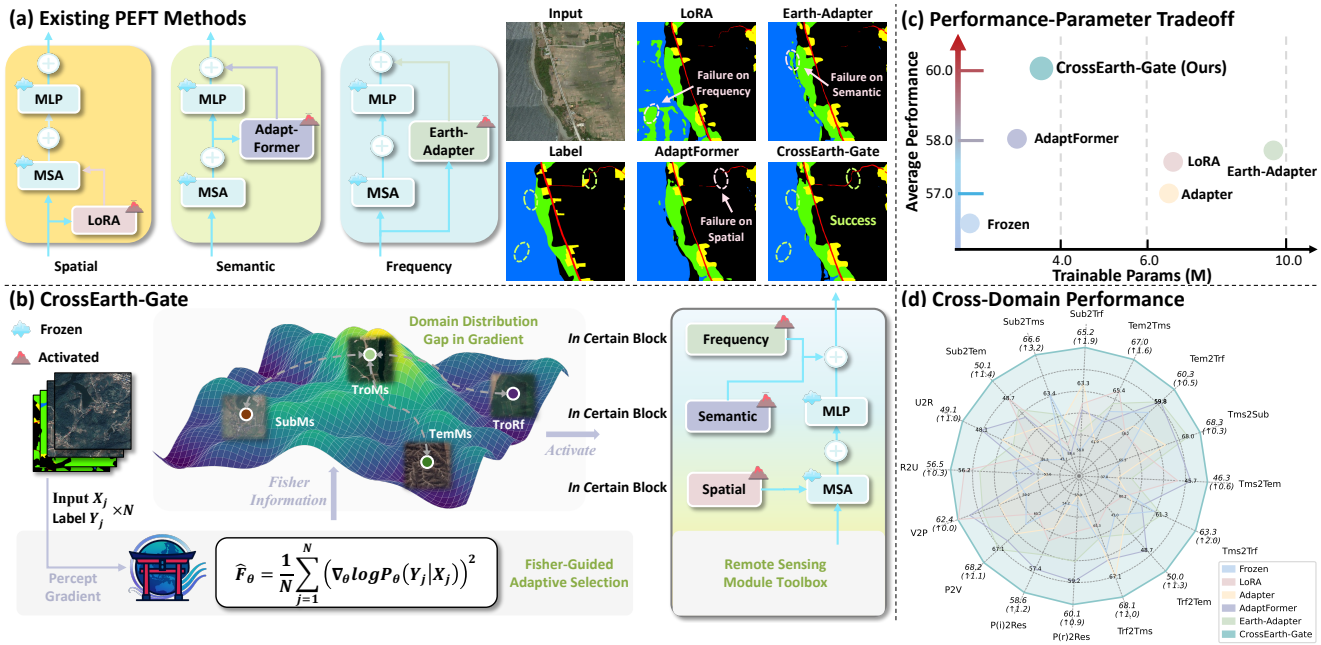


Figure 1. Overview of the CrossEarth-Gate and its comparative advantages. (a) Existing PEFTs typically focus on one specific functional pathway (*e.g.*, LoRA for spatial, Adaptformer for semantic, Earth-Adapter for frequency). The qualitative example of generalization across different climate zones shows each baseline failing on challenges outside its specialty, while our method succeeds. (b) Our proposed CrossEarth-Gate establishes a toolbox combining all three module types. Then, we utilize Fisher Information to guide the gradient flow to activate only the most critical modules at the most relevant blocks for a specific domain. (c) CrossEarth-Gate results in a superior Performance-Parameter Tradeoff. (d) CrossEarth-Gate achieves state-of-the-art results across 16 challenging DG and DA benchmarks.

est, committing a semantic error. These dilemmas motivate us to pursue a more dynamic mechanism capable of jointly addressing spatial, semantic, and frequency variations.

To address this, we propose **CrossEarth-Gate**, a Fisher-**G**uided **a**daptive **t**uning **e**ngine for **Cross**-domain PEFT in **Earth** observation. Our methods provide a dynamic and comprehensive solution to multifaceted domain gaps, which comprises two components. First, CrossEarth-Gate establishes a structured **RS module toolbox**, designed to inject trainable parameters into a different functional pathway of models, including spatial, semantic, and frequency modules using different PEFT methods. This provides a unified framework with the full capacity to address complex domain shifts. Another novel and critical design is the **Fisher-guided adaptive selection** that operates on this toolbox to dynamically choose the most critical modules at the appropriate layers. Specifically, Fisher Information guides the selection [15], which serves as a metric to evaluate the importance of each module in terms of its contribution to the current task-specific gradient flow. This mechanism periodically determines where to "cast its hook", directing the gradient flow of the model's parameters in a way that maximizes adaptation efficiency and effectiveness for domain-specific shifts. This process facilitates a multi-stage learning process, allowing for a robust and nuanced adaptation.

In addition, we rigorously evaluate CrossEarth-Gate across 18 Domain Generalization (DG) and Domain Adaptation (DA) benchmarks for RS semantic segmentation. It achieves State-Of-The-Art (SOTA) results on 16 of them, outperforming PEFT baselines by up to 3.2% mIoU while offering a better performance-efficiency tradeoff (See Fig. 1 (c) and (d)). The main contributions of this work are:

- We address the challenges posed by extreme heterogeneity and multifaceted domain shifts inherent in RS adaptation tasks with a unified, comprehensive framework.
- We propose CrossEarth-Gate, which introduces a structured RS module toolbox to comprehensively tackle the multifaceted RS domain shifts and a Fisher-guided adaptive selection mechanism to dynamically guide the gradient flow to maximize efficiency and effectiveness.
- CrossEarth-Gate outperforms existing specialized cross-domain methods, FMs, and PEFE approaches across 16 cross-domain benchmarks for RS semantic segmentation, validating its strong generalizability and effectiveness.

## 2. Related Works

### 2.1. Geospatial Foundation Models

The concept of Foundation Model (FM) originates from the NLP [72] domain, defined as "the base models trained on large-scale unlabeled data that can be adapted for a variety of downstream tasks" [3, 8, 12, 53, 54, 63, 65, 70]. Driven

by the success of self-supervised learning (SSL) techniques such as masked patch reconstruction, Vision Foundation Models (VFMs) have yielded significant achievements in Computer Vision (CV), serving as powerful backbones for general visual feature representation [6, 14, 23, 24, 35, 50].

Adapting existing VFM architectures and techniques to Earth observation data, typically represented as images, has given rise to Geospatial Foundation Models (GFMs) [9, 19, 20, 38, 39, 45, 48, 56, 57, 61, 67, 69, 83]. For example, SatMAE [9] introduces temporal and multi-spectral positional encodings for satellite imagery within the mask image modeling pre-training. To tackle domain shift, CrossEarth [19] utilizes Earth-style injection and multi-task training for domain generalization. However, as the representation capabilities and scale of these FMs increase, how to efficiently and effectively leverage their potential for diverse RS downstream tasks remains a key problem.

### 2.2. Cross-Domain PEFT in RS

PEFT has emerged as a promising solution to efficiently fine-tune FMs while using a minimum number of trainable parameters [21, 26, 29, 34, 81]. These methods can be broadly classified into four categories: Selective, Additive, Prompt, and Reparameterization approaches [81]. Selective PEFT focuses on only optimizing a specific subset of the model's parameters [2, 4, 60, 75, 80, 85, 86]. For instance, [60] selects sparse parameters based on Fisher information as a measure of parameter importance. Additive PEFT introduces extra trainable parameters into the frozen backbone [5, 26, 52, 52, 78]. Prompt PEFT incorporates learnable prompts into the input or the attention layers to adapt to a specific task. Reparameterization PEFT reformulates or decomposes existing model parameters so that only a subset requires adjustment during fine-tuning [29, 42]. Particularly, LoRA [29] decomposes the updated weight into two low-rank matrices, inspiring numerous subsequent methods [11, 22, 44, 82, 87]. Moreover, existing cross-domain studies begin to explore PEFT for Domain Adaptation (DA) [16, 18, 59, 76] and Domain Generalization (DG) [30, 71].

Recently, PEFT techniques have also gained increasing attention within the RS community, which commonly leverages VFMs or GFMs for various downstream tasks [13, 31, 32, 58, 62, 79, 84]. For instance, SLR [58] and DEFLECT [81] utilize low-rank matrices to adapt to different modalities of RS data, while Earth-Adapter [33] is designed to mitigate the domain shift resulting from frequency-domain artifacts. However, these PEFT methods typically focus on one aspect of the problem, such as spatial, semantic, and frequency adaptation, failing to address the complex, multifaceted nature of RS domain shifts. A method specialized for one aspect often fails when confronted with challenges from another. Consequently, a single static method fails to address these intertwined challenges comprehensively. To address this, we establish a structured toolbox combining three module types, targeting different functional pathways of the model. We then propose a Fisher-guided adaptive selection mechanism that dynamically identifies the most relevant modules, enabling a more comprehensive cross-domain adaptation.

## 3. Preliminary

**Optimization Objective.** Let us consider a cross-domain adaptation task with a training dataset $D = \{(\mathbf{X}, \mathbf{Y})\}$, where the $\mathbf{X} \in \mathbb{R}^{h \times w \times c}$ denotes the input RS images, comprising $h \times w$ pixels with $c$ bands, and $\mathbf{Y}$ represents corresponding label. For Domain Generalization (DG) tasks, $D$ represents the labeled source domain dataset, while Domain Adaptation (DA) additionally includes the unlabeled target domain input and its pseudo-labels. Given a pretrained FM with parameters $\alpha$, PEFT-related modules with parameters $\zeta$ ($|\zeta| \ll |\alpha|$), and the decoder head $\mathcal{H}_\phi$ parameterized by $\phi$, the objective of cross-domain PEFT is to enhance the model performance in the target domain by optimizing:

$$\arg\min_{\zeta,\phi} \mathbb{E}_{(\mathbf{X},\mathbf{Y})\in D} \mathcal{L}(\mathcal{H}_\phi(\mathcal{B}_{\alpha,\zeta}(\mathbf{X})), \mathbf{Y}), \tag{1}$$

where $\mathcal{B}_{\alpha,\zeta}$ signifies the backbone incorporating the PEFT modules and $\mathcal{L}$ is the loss function for the adaptation task.

**Transformer Block.** This work mainly focuses on Transformer-based [65] FMs, such as ViT [14]. The input $\mathbf{X}$ is first divided into $l$ patches and then encoded into $d$-dimensional embedding through a patch embedding layer, yielding the initial tokens $\mathbf{T}_1 \in \mathbb{R}^{l \times d}$. Subsequently, $\mathbf{T}_1$ is processed by a sequence of $I$ transformer blocks in the FM. We denote the input features of $i$-th block as $\mathbf{T}_i \in \mathbb{R}^{l \times d}$. Each block is typically composed of a Multi-head Self-Attention (MSA) module and a feed-forward Multi-Layer Perceptron (MLP) network, with residual connections and Layer Normalization (LN) applied after each module. The computation for the $i$-th block is defined as follows:

$$\mathbf{T}_i^{\text{attn}} = \text{MSA}(\mathbf{T}_i) + \mathbf{T}_i, \quad \mathbf{T}_i^{\text{attn}} \in \mathbb{R}^{l \times d}, \tag{2}$$

$$\mathbf{T}_{i+1} = \text{MLP}(\mathbf{T}_i^{\text{attn}}) + \mathbf{T}_i^{\text{attn}}, \quad \mathbf{T}_{i+1} \in \mathbb{R}^{l \times d}, \tag{3}$$

where we omit normalization for simplicity and $\mathbf{T}_{i+1}$ is the output of $i$-th block, which is passed as input to next block.

## 4. Methods

The Fig. 1 (b) presents an overview of the CrossEarth-Gate, which consists of a structured RS module toolbox and Fisher-guided adaptive selection mechanism.

### 4.1. Remote Sensing Module Toolbox

First, to equip the model with the comprehensive capabilities needed to tackle intertwined domain gaps, CrossEarth-Gate first establishes a structured and comprehensive **RS**

**module toolbox**. This toolbox integrates spatial, semantic, and frequency modules, each engineered into distinct functional pathways to provide specific aspects of the feature hierarchy for specific domain challenges. Instead of statically choosing one method for certain layers, we initially insert all these modules at every layer of the pre-trained backbone.

**Spatial Module.** The MSA sub-layer (Eq. (2)) explicitly models the relationships between tokens, allowing the model to build contextual understanding and capture spatial dependencies at various scales [25, 47]. In the RS application scenarios, adapting to change in object scale, spatial arrangement, or geographic layout necessitates a targeted tuning of this reasoning mechanism. Therefore, we employ the Spatial Module using Low-Rank Adaptation (LoRA) [29]. LoRA is predicated on the hypothesis that the weight update $\Delta\mathbf{W}$ for a pre-trained weight matrix $\mathbf{W}_0 \in \mathbb{R}^{d\times d}$ during adaptation possesses a low "intrinsic rank". This update can be parameterized by the product of two low-rank matrices, $\mathbf{A} \in \mathbb{R}^{d\times r}$ and $\mathbf{B} \in \mathbb{R}^{r\times d}$, where $r \ll d$ is the rank. The merged weight $\mathbf{W}$ is written as:

$$\mathbf{W} = \mathbf{W}_0 + \Delta\mathbf{W} = \mathbf{W}_0 + \mathbf{BA}, \quad \mathbf{W} \in \mathbb{R}^{d\times d}. \tag{4}$$

We strategically inject these trainable, low-rank matrices into the query ($\mathbf{W}_Q$) and value ($\mathbf{W}_V$) linear projection weights of each MSA module (Eq. (2)). This targeted intervention directly modulates the self-attention mechanism, influencing how the model weighs and aggregates spatial information. It allows the model to adapt its understanding of spatial context and object scales without altering the parameters in the original, frozen $\mathbf{W}_Q$ and $\mathbf{W}_V$ weights.

**Semantic Module.** The MLP sub-layer (Eq. (3)) transforms each token's representation independently, which is widely understood to be a key locus of factual and semantic knowledge within the model [17, 49, 77]. Moreover, generalizing concepts across different domains (e.g., "building" in rural and urban landscapes) logically requires modifying this stored knowledge. To adapt the model's high-level semantic knowledge, we introduce a Semantic Module based on the Adapter architecture [5, 26], placed in parallel with each MLP sub-layer. The adapter in $i$-th block comprises a down-projection layer with weight $\mathbf{W}_i^{down} \in \mathbb{R}^{d\times\hat{d}}$, a GELU activation, and an up-projection layer with weight $\mathbf{W}_i^{up} \in \mathbb{R}^{\hat{d}\times d}$, satisfying $\hat{d} \ll d$. The output of this module is added to the original MLP's output via a residual connection. Given the modified output $\hat{\mathbf{T}}_i^{\text{attn}}$ of the MSA integrating LoRA, the modified output of block $\hat{\mathbf{T}}_{i+1}$ becomes:

$$\hat{\mathbf{T}}_{i+1} = \text{MLP}(\hat{\mathbf{T}}_i^{\text{attn}}) + \text{Adapter}_i(\hat{\mathbf{T}}_i^{\text{attn}}) + \hat{\mathbf{T}}_i^{\text{attn}}, \tag{5}$$

$$\text{Adapter}_i(\hat{\mathbf{T}}_i^{\text{attn}}) = \text{GELU}(\hat{\mathbf{T}}_i^{\text{attn}} \cdot \mathbf{W}_i^{down}) \cdot \mathbf{W}_i^{up}. \tag{6}$$

This parallel structure allows the module to refine or adjust the semantic transformation performed by the frozen MLP without disrupting the original pre-trained knowledge flow.

**Frequency Module.** RS imagery is uniquely challenged by high-amplitude artifact influences, which are almost situated everywhere in the RS image. To address this, we introduce Frequency Modules based on the Earth-Adapter [33], which operates in the frequency domain. The module first utilizes the Fourier Transform to decompose the input features into low-frequency (structural) and high-frequency (detail/texture) components. These disentangled components are processed by distinct, lightweight adapter experts. A mixture-of-adapters router then learns to selectively process and recombine these frequency components, effectively mitigating artifact disturbances while preserving essential features. Given the modified output of the $i$-th block $\hat{\mathbf{T}}_{i+1}$, the input to next block $\tilde{\mathbf{T}}_{i+1}$ becomes:

$$\tilde{\mathbf{T}}_{i+1} = \hat{\mathbf{T}}_{i+1} + \text{Earth-Adapter}_i(\hat{\mathbf{T}}_{i+1}). \tag{7}$$

### 4.2. Fisher-Guided Adaptive Selection

Given the toolbox, the central challenge becomes determining which modules to activate for a specific task. Tuning all modules simultaneously is fundamentally inefficient, yet a static or heuristic selection would fail to efficiently leverage the full potential of the toolbox for diverse domain shifts, reverting to a sub-optimal strategy. Therefore, CrossEarth-Gate employs a principled, data-driven selection mechanism as the dynamic guide. We conceptualize the adaptation as a gradient flow: as the model fine-tunes, task-specific gradients flow through the network, and our different module types offer distinct pathways for this flow. Our goal is to periodically analyze this flow and "gate" it, directing it only to the modules that offer the highest impact for adaptation.

To quantify this impact in a theoretically-grounded way, we turn to the Fisher Information Matrix (FIM) [15]. A parameter's significance can be determined by evaluating the extent to which altering it influences the model's output. Given a model parameterized by $\theta \in \mathbb{R}^{|\theta|}$, we denote the output distribution of the model as $P_\theta(\mathbf{Y}|\mathbf{X})$ for input $\mathbf{X}$. Subsequently, we can assess how much a small parameter perturbation $\delta \in \mathbb{R}^{|\theta|}$ in the parameter changes the distribution using the Kullback-Leibler divergence [37] $\mathrm{D}_{KL}(P_\theta(\mathbf{Y}|\mathbf{X}) \parallel P_{\theta+\delta}(\mathbf{Y}|\mathbf{X}))$. As shown in [1, 51], when $\delta \to 0$, the following second-order approximation holds:

$$\mathbb{E}_{\mathbf{X}}\left[\mathrm{D}_{KL}(P_\theta(\mathbf{Y}|\mathbf{X}) \parallel P_{\theta+\delta}(\mathbf{Y}|\mathbf{X}))\right] = \delta^T F_\theta \delta + O(\delta^3), \tag{8}$$

where $F_\theta \in \mathbb{R}^{|\theta|\times|\theta|}$ is the FIM [15], defined as:

$$F_\theta = \mathbb{E}_{\mathbf{X}}\left[\mathbb{E}_{\mathbf{Y}\sim P_\theta(\mathbf{Y}|\mathbf{X})} \nabla_\theta \log P_\theta(\mathbf{Y}|\mathbf{X}) \nabla_\theta \log P_\theta(\mathbf{Y}|\mathbf{X})^T\right]. \tag{9}$$

Apparently, the FIM links $\delta$ to the resultant changes in the model's output distribution. However, the $|\theta| \times |\theta|$ size of $F_\theta$ renders its exact computation infeasible. Consequently, we adopt the empirical diagonal approximation of the FIM [36, 60], which is equivalently a vector in $\mathbb{R}^{|\theta|}$. When sampling $N$ data pairs $(\mathbf{X}_j, \mathbf{Y}_j)$ from the task dataset $D$, the diagonal FIM for $\theta$ can be empirically estimated as:

$$\hat{F}_\theta = \frac{1}{N} \sum_{j=1}^{N} (\nabla_\theta \log P_\theta(\mathbf{Y}_j|\mathbf{X}_j))^2, \tag{10}$$

where $\log P_\theta(\mathbf{Y}_j|\mathbf{X}_j)$ is realized as the negative task loss. This practical approximation relates the $\hat{F}_\theta$ to the square gradient of the parameter. We can conceptualize the adaptation as a flow, where the task-specific loss directs gradients back to the parameters. The gradient represents the direction and magnitude of this flow. A large $\hat{F}_\theta$ value signifies that this parameter is a "high-flow" channel, *i.e.*, a pathway where gradients are large and thus highly impactful on the model's output. This is precisely where we should "cast our hook" to achieve the maximum adaptation gain.

The FIM approximation in Eq. (10) provides an importance score for an individual parameter. To select entire modules, CrossEarth-Gate employs a dynamic gating mechanism that periodically re-evaluates module importance throughout the fine-tuning process. By using the empirical FIM as our "flow meter", we are able to dynamically "gate" the gradient flow, directing it only to the modules that offer the highest impact. Specifically, at every $N$ training iterations, the framework will temporarily activate all modules in the toolbox. The Fisher information is then computed by summing over a small batch of $M$ data samples. To derive a module-level importance score $S_i^z$ for the $z$-th type of module parameterized by $\zeta_i^z$ in the $i$-th block, we aggregate the scores of all its constituent parameters: $\hat{S}_i^z = \sum_{\zeta_i^z} \hat{F}_{\zeta_i^z}$. To make the scores of different types of modules comparable, we compute the category-normalized relative importance scores: $S_i^z = \frac{\hat{S}_i^z}{\sum_i^I \hat{S}_i^z}$, where $I$ represents the number of blocks. This ensures that the selection process is balanced, making all module types comparable and promoting a diverse, task-specific configuration. Subsequently, only Top-k modules with the highest scores are activated, and the gradient flow is gated to only these pathways for the next $N$ training iterations. CrossEarth-Gate repeats this process to dynamically adapt its tuning strategy to the most critical components for the task at hand. The hyperparameter configurations are listed in the Appendix.

# 5. Experiments

## 5.1. Settings

We conduct a comprehensive evaluation of CrossEarth-Gate on DG and DA benchmarks for RS semantic segmentation tasks. Additional dataset details, implementation specifics, and baseline configurations are provided in the Appendix.

**Datasets.** Our cross-domain analysis leverages a suite of diverse datasets. The CASID dataset specifically addresses domain shifts across four climate zones: Subtropical Monsoon (Sub), Temperate Monsoon (Tem), Tropical Monsoon (Tms), and Tropical Rainforest (Trf). The ISPRS Potsdam and Vaihingen contain aerial images collected over the cities of Potsdam and Vaihingen, with IR-R-G, R-G-B, and R-G-B-IR channels. RescueNet consists of aerial imagery focused on detecting buildings impacted by disasters to facilitate rescue operations. LoveDA presents a cross-scene dataset with RS images from both urban and rural areas. For DG experiments, we follow the protocol from [33]. We employ CASID to establish 12 out-of-domain generalization scenarios. We also evaluate on RescueNet with source domain as the RGB (P(r)2Res) and IR–R–G (P(I)2Res) channels of Potsdam. For DA experiments, we form four benchmarks: Potsdam to Vaihingen (P2V), Vaihingen to Potsdam (V2P), Rural to Urban (R2U), and Urban to Rural (U2R). These datasets provide a spanning multifaceted RS domain shifts, including unseen geographical regions, novel spectral bands, and diverse climatic conditions.

**Implementation Details.** All experiments are implemented based on the MMSegmentation [10] framework. We mainly employ Dinov2 [50] as the feature extraction backbone, paired with Mask2Former [7] as the segmentation head. For DA experiments, we utilize the DACS [64] self-training framework, whereas for DG experiments, we train the model via end-to-end supervised learning. The models are finetuned using AdamW [46] optimizer with a base learning rate of 1e-5 for the decoder and PEFT modules.

**Baselines.** We compare CrossEarth-Gate against existing PEFT methods, including VPT [34], SLR [58], Rein [71], LoRA [29], Adapter [26], Adaptformer [5], and Earth-Adapter [33]. We also select two conventional fine-tuning approaches: "Frozen", where all backbone parameters are fixed, and "Full-Tuning", where we fine-tune the entire model. Additionally, for the DG experiments, we include results from two cross-domain specialized models and the Full-Tuning of different GFMs, based on [19].

## 5.2. Cross-Domain Performance Comparison

**Generalization Across Climate Zones.** As presented in Tab. 1, Full-Tuning undergoes a catastrophic performance decline compared to the Frozen backbone, confirming that unconstrained fine-tuning severely overfits to the source domain. Specialized, static PEFTs present inconsistent performance, showing varying efficacy depending on the domain. For instance, the Earth-Adapter achieves the second-

Table 1. Performance comparison on CASID benchmarks across 12 DG experiments. We compare specialized DA models, FMs, PEFT methods, and CrossEarth-Gate. The **best** and second-best scores are indicated in bold and underlined, and the improvement of CrossEarth-Gate is shown in brackets. Sub: Subtropical Monsoon. Tem: Temperate Monsoon. Tms: Tropical Monsoon. Trf: Tropical Rainforest.

| Backbone | Method | Params (M) | Sub2Tem | Sub2Tms | Sub2Trf | Average | Tem2Sub | Tem2Tms | Tem2Trf | Average |
|---|---|---|---|---|---|---|---|---|---|---|
| MiT-B5 [73] | HRDA [28] | 81.4 | 34.6 | 63.0 | 56.2 | 51.3 | 62.7 | 63.3 | 44.9 | 57.0 |
| | DAFormer [27] | 81.4 | 36.0 | 63.5 | 59.0 | 52.8 | 59.8 | 63.8 | 51.7 | 58.4 |
| ViT-L [14] | MTP [66] | 304.2 | 39.4 | 60.0 | 61.6 | 53.7 | 66.2 | 64.7 | 48.3 | 59.7 |
| | SatMAE [9] | 304.2 | 33.1 | 52.9 | 48.7 | 44.9 | 51.8 | 43.7 | 36.2 | 43.9 |
| | ScaleMAE [56] | 304.2 | 39.2 | 60.7 | 56.4 | 52.1 | 62.5 | 57.3 | 45.9 | 55.2 |
| | RemoteCLIP [41] | 304.2 | 8.5 | 15.1 | 19.7 | 14.4 | 54.3 | 57.8 | 16.7 | 42.9 |
| DINOv2-L [50] | Frozen | 0.0 | 43.1 | 63.4 | 58.8 | 55.2 | 66.9 | 64.8 | 59.0 | 63.6 |
| | Full-Tuning | 304.2 | 38.6 | 62.3 | 58.3 | 53.0 | 58.7 | 55.8 | 43.8 | 52.8 |
| | CrossEarth [19] | 3.0 | 48.1 | 64.6 | 64.2 | 59.0 | 63.5 | 61.8 | 57.8 | 61.0 |
| | VPT [34] | 3.1 | 41.2 | 61.4 | 60.8 | 54.4 | 64.6 | 61.2 | 58.7 | 61.5 |
| | SLR [58] | 6.9 | 24.3 | 37.0 | 32.3 | 31.2 | 33.3 | 22.0 | 26.1 | 27.1 |
| | Rein [71] | 3.0 | 42.3 | 56.2 | 52.9 | 50.5 | 66.0 | 62.1 | 56.7 | 61.6 |
| | LoRA [29] | 6.4 | 48.7 | 60.5 | 62.3 | 57.2 | 68.2 | 65.4 | 58.2 | 64.0 |
| | Adapter [26] | 6.3 | 46.9 | 61.5 | 63.3 | 57.2 | 67.7 | 61.9 | 58.5 | 62.7 |
| | Adaptformer [5] | 3.2 | 47.9 | 58.4 | 62.1 | 56.1 | 68.2 | 64.4 | 59.8 | 64.1 |
| | Earth-Adapter [33] | 9.6 | 48.3 | 61.6 | 60.5 | 56.8 | **68.8** | 64.9 | 59.8 | 64.5 |
| | **CrossEarth-Gate** | 3.0-4.4 | **50.1 (+1.4)** | **66.6 (+2.0)** | **65.2 (+1.0)** | **60.6 (+1.6)** | 68.0 (-0.8) | **67.0 (+1.6)** | **60.3 (+0.5)** | **65.1 (+0.6)** |
| Backbone | Method | Params (M) | Tms2Sub | Tms2Tem | Tms2Trf | Average | Trf2Sub | Trf2Tem | Trf2Tms | Average |
| MiT-B5 [73] | HRDA [28] | 81.4 | 63.8 | 33.3 | 56.9 | 51.3 | 63.3 | 43.0 | 63.1 | 56.5 |
| | DAFormer [27] | 81.4 | 61.8 | 31.4 | 56.2 | 49.8 | 62.9 | 39.6 | 62.7 | 55.1 |
| ViT-L [14] | MTP [66] | 304.2 | 56.1 | 32.3 | 60.1 | 49.5 | 55.6 | 37.0 | 59.8 | 50.8 |
| | SatMAE [9] | 304.2 | 60.2 | 32.5 | 53.0 | 48.6 | 59.1 | 32.8 | 56.7 | 49.5 |
| | ScaleMAE [56] | 304.2 | 62.1 | 35.2 | 55.4 | 50.9 | 60.6 | 30.9 | 58.0 | 49.8 |
| | RemoteCLIP [41] | 304.2 | 22.5 | 12.3 | 22.5 | 19.1 | 32.1 | 26.0 | 23.6 | 27.2 |
| DINOv2-L [50] | Frozen | 0.0 | 65.3 | 37.0 | 60.9 | 54.4 | 68.0 | 43.0 | 66.7 | 59.2 |
| | Full-Tuning | 304.2 | 60.9 | 29.9 | 58.9 | 49.9 | 61.9 | 32.9 | 64.3 | 53.0 |
| | CrossEarth [19] | 3.0 | 69.1 | 40.5 | 60.7 | 56.8 | 67.9 | 42.3 | 64.4 | 58.2 |
| | VPT [34] | 3.1 | 64.8 | 33.9 | 56.8 | 51.8 | 67.3 | 38.0 | 66.9 | 57.4 |
| | SLR [58] | 6.9 | 39.6 | 23.6 | 38.7 | 34.0 | 42.2 | 24.2 | 34.3 | 33.5 |
| | Rein [71] | 3.0 | 58.2 | 29.0 | 53.4 | 46.9 | 62.6 | 33.8 | 61.1 | 52.5 |
| | LoRA [29] | 6.4 | 66.7 | 45.2 | 60.2 | 54.5 | 67.7 | 46.4 | 65.3 | 59.8 |
| | Adapter [26] | 6.3 | 67.8 | 38.5 | 60.4 | 55.5 | 68.9 | 47.7 | 67.1 | 61.2 |
| | Adaptformer [5] | 3.2 | 67.9 | 45.7 | 60.5 | 58.0 | **69.8** | 48.7 | 66.8 | 61.8 |
| | Earth-Adapter [33] | 9.6 | 68.0 | 39.2 | 61.3 | 56.2 | 68.4 | 47.8 | 65.7 | 60.6 |
| | **CrossEarth-Gate** | 3.0-4.4 | **68.3 (+0.3)** | **46.3 (+0.6)** | **63.3 (+2.0)** | **59.3 (+1.3)** | 69.0 (-0.8) | **50.0 (+1.3)** | **68.1 (+1.0)** | **62.4 (+0.6)** |

best average result (64.5%) when trained on the Tem domain, whereas the Adaptformer performs better (61.8%) when trained on the Trf domain. This variance suggests that different climate-zone transfers necessitate a different emphasis on spatial, semantic, or frequency adjustments, exposing the inherent limitations of the specialized PEFT strategy. In contrast, CrossEarth-Gate demonstrates consistent performance, achieving SOTA results in 10 of the 12 scenarios and securing the highest average mIoU across all four source domains, all while maintaining high parameter efficiency. This superior result highlights the critical advantage of our structured RS module toolbox, which comprehensively tackles the complex RS domain shifts. While our method is dominant, in two isolated cases (Tem2Sub and Trf2Sub), static methods like Earth-Adapter or Adaptformer achieve marginally better results. This suggests that for these specific shifts, their fixed module placement coincidentally aligns well with the required feature adaptation. However, these static, specialized methods fail to generalize to other domains. The comprehensive results underscore the robustness and superior generalization capability of the CrossEarth-Gate framework across different climate zones.

**Generalization to Disaster Scenarios.** Tab. 2 shows the results of the generalization task from a standard aerial dataset (Potsdam) to a post-disaster scenario (RescueNet), involving simultaneous shifts in geography, spectral bands, and target object appearance. CrossEarth-Gate again achieves the highest mIoU in both experiments, with an mIoU of 60.1% on P(r)2Res and 58.6% on P(i)2Res. Our method demonstrates strong generalization on ambiguous, large-scale classes like Impervious surfaces and Clutter, which are prevalent in post-disaster environments. However, for the Building class, the Frozen backbone outperforms our method in the P(r)2Res case. This is a noteworthy finding, suggesting that the DINOv2's pre-trained representations for highly structured objects are already extremely robust and generalizable. In this context, any adaptation may marginally disrupt these near-optimal features for the Building class. Despite these class-specific variances, the comprehensive SOTA mIoU scores validate the superior and more balanced generalization capability of CrossEarth-Gate in the complex cross-domain RS scenarios.

**Adaptation to Unlabeled Target Domains.** The DA results presented in Tab. 3, where the model has access to

Table 2. Performance comparison on P(r)2Res and P(i)2Res DG experiments. Surf: Impervious surfaces. Bldg: Building. Clut: Clutter.

| Backbone | Method | Params (M) | P(r)2Res (Classes) | | | | | mIoU (%) | P(i)2Res (Classes) | | | | | mIoU (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Surf | Bldg | Tree | Car | Clut | | Surf | Bldg | Tree | Car | Clut | |
| MiT-B5 [73] | HRDA [28] | 81.4 | 28.8 | 27.2 | 56.8 | 4.0 | 66.4 | 36.6 | 26.1 | 29.2 | 44.3 | 5.2 | 61.6 | 33.3 |
| | DAFormer [27] | 81.4 | 33.0 | 32.6 | 55.0 | 7.6 | 66.7 | 39.0 | 34.1 | 43.6 | 41.7 | 12.0 | 58.9 | 38.1 |
| ViT-L [14] | MTP [66] | 304.2 | 34.3 | 40.9 | 41.7 | 16.3 | 62.7 | 39.2 | 31.3 | 36.8 | 0.5 | 14.4 | 57.5 | 28.1 |
| | SatMAE [9] | 304.2 | 13.7 | 13.8 | 26.1 | 3.7 | 27.1 | 16.9 | 15.5 | 13.7 | 24.4 | 3.0 | 21.1 | 15.5 |
| | ScaleMAE [56] | 304.2 | 23.5 | 31.9 | 45.4 | 0.1 | 47.7 | 29.7 | 13.9 | 35.0 | 48.7 | 0.3 | 49.4 | 29.4 |
| | RemoteCLIP [41] | 304.2 | 2.2 | 8.2 | 2.6 | 0.0 | 0.0 | 2.6 | 1.3 | 10.1 | 27.5 | 0.2 | 0.8 | 8.0 |
| DINOv2-L [50] | Frozen | 0.0 | 45.0 | **61.7** | 64.1 | 33.8 | 66.3 | 54.2 | 53.0 | 61.4 | 56.5 | 31.1 | 69.1 | 54.2 |
| | Full-Tuning | 304.2 | 21.3 | 22.5 | 25.2 | 3.0 | 61.8 | 26.7 | 17.5 | 29.6 | 64.5 | 3.2 | 65.3 | 36.0 |
| | CrossEarth [19] | 3.0 | 43.6 | 59.7 | 51.0 | 16.4 | 63.2 | 46.8 | 41.9 | 60.6 | 51.3 | 10.6 | 62.9 | 45.5 |
| | VPT [34] | 3.1 | 50.3 | 56.6 | 58.8 | 38.1 | 72.0 | 55.1 | 52.8 | 61.0 | 58.8 | 35.6 | 71.1 | 55.9 |
| | SLR [58] | 6.9 | 52.1 | 59.3 | 61.0 | 21.7 | 72.8 | 53.4 | 54.4 | 57.4 | 55.5 | 19.3 | 71.3 | 51.6 |
| | Rein [71] | 3.0 | 48.8 | 59.9 | 51.4 | 23.0 | 69.4 | 50.5 | 21.1 | 42.0 | 56.3 | 19.9 | 59.7 | 39.8 |
| | LoRA [29] | 6.4 | 51.1 | 57.4 | 62.5 | 39.1 | 72.8 | 56.6 | 50.1 | 59.8 | **65.1** | 35.1 | 72.3 | 56.5 |
| | Adapter [26] | 6.3 | 56.7 | 59.6 | 53.1 | 30.3 | 69.8 | 53.9 | 56.1 | 59.3 | 63.5 | 30.7 | 73.3 | 56.1 |
| | Adaptformer [5] | 3.2 | 58.5 | 60.3 | 65.5 | 38.9 | 73.1 | 59.2 | 55.4 | **61.9** | 60.2 | **37.1** | 72.6 | 57.4 |
| | Earth-Adapter [33] | 9.6 | 56.5 | 61.1 | 63.4 | 39.6 | 72.4 | 58.6 | 54.2 | 60.8 | 61.8 | 35.1 | 72.8 | 57.0 |
| | **CrossEarth-Gate** | 3.3-4.0 | **60.8 (+2.3)** | 58.6 (-3.1) | **67.5 (+2.0)** | **40.5 (+0.9)** | **73.3 (+0.2)** | **60.1 (+0.9)** | **57.7 (+1.6)** | 61.5 (-0.4) | 64.6 (-0.5) | 34.6 (-2.5) | **74.7 (+1.4)** | **58.6 (+1.2)** |

Table 3. Performance comparison on DA benchmarks between representative DA models, PEFT methods, and CrossEarth-Gate.

| Methods | Params (M) | Domain Adaptation | | | | Avg. |
|---|---|---|---|---|---|---|
| | | P2V | V2P | R2U | U2R | |
| HRDA [28] | 81.4 | 67.6 | 58.6 | 53.2 | 35.3 | 53.7 |
| DAFormer [27] | 81.4 | 64.4 | 54.8 | 52.7 | 42.5 | 53.6 |
| Frozen | 0.0 | 66.1 | 59.2 | 54.9 | 45.5 | 56.4 |
| Full-Tuning | 304.2 | 62.4 | 59.6 | 42.6 | 35.8 | 50.1 |
| VPT [34] | 3.1 | 65.0 | 57.6 | 55.8 | 47.0 | 56.4 |
| SLR [58] | 6.9 | 14.4 | 16.5 | 23.8 | 9.6 | 16.1 |
| Rein [71] | 3.0 | 64.1 | 59.5 | 54.7 | 34.5 | 53.2 |
| LoRA [29] | 6.4 | 65.2 | **62.4** | 56.2 | 46.6 | 57.6 |
| Adapter [26] | 6.3 | 66.8 | 60.3 | 53.6 | 47.7 | 57.1 |
| Adaptformer [5] | 3.2 | 66.9 | 62.2 | 53.7 | 48.1 | 57.7 |
| Earth-Adapter [33] | 1.0-3.9 | 67.1 | 61.6 | 56.0 | 47.5 | 58.1 |
| **CrossEarth-Gate** | 1.7-3.9 | **68.2 (+1.1)** | **62.4 (+0.0)** | **56.5 (+0.3)** | **49.1 (+1.0)** | **59.1 (+1.0)** |

unlabeled target data, further validate our approach. Full-Tuning fails again, likely due to instability from noisy target-domain pseudo-labels. In contrast, PEFT methods, including LoRA, Adaptformer, and Earth-Adapter, all outperform the strong Frozen baseline, demonstrating their benefits in the DA context. However, CrossEarth-Gate achieves the best performance across four adaptation scenarios with the highest average mIoU of 59.1%. This consistent success highlights the efficacy of our Fisher-guided selection in the DA context. The gradient flow, informed by pseudo-labels from a specific target domain, allows our model to precisely identify and activate only the most critical modules needed for that particular adaptation. While our method tied with the LoRA on the V2P task, this does not indicate a limitation. Rather, it suggests that this specific V2P adaptation may be predominantly a spatial-scale challenge, which LoRA's static architecture coincidentally addresses well. However, CrossEarth-Gate's ability to win or tie across all diverse UDA scenarios validates its more robust and principled approach to dynamic adaptation.

### 5.3. Ablation Studies

**Backbone Generalizability.** As detailed in Tab. 4, we evaluate CrossEarth-Gate across five pre-trained VFMs and GFMs on the CASID benchmark, including DINOv2 series [50], SAM [35], SatMAE [9], and Scale-MAE [56]. Specifically, Full-Tuning is an unstable and generally poor strategy for DG. It performs significantly worse than the Frozen baseline on the DINOv2 series, a clear case of overfitting to the source domain, while incurring an enormous parameter cost. In contrast, CrossEarth-Gate consistently and significantly outperforms both the Frozen and Full-Tuning baselines across all five architectures. It achieves the highest average mIoU in every case, with improvements from 2.1 % to 7.6 % mIoU, while tuning a minimal number of parameters. This robust performance strongly validates the remarkable versatility and generalizability of the CrossEarth-Gate.

**Core Components Impact.** We further dissect the framework by ablating its key components, with results visualized in Fig. 3. When all modules from the toolbox are trained simultaneously without selection, it results in the most performance degradation. Its failure proves that naively increasing trainable parameters is detrimental, likely leading to conflicting gradient updates. This result empirically validates the necessity of CrossEarth-Gate to guide the gradient flow for utilizing adaptation resources effectively. Removing each specialized module type causes a comparable drop in mean performance. All three adaptation pathways are non-redundant and essential for achieving comprehensive generalization. For instance, the Sub and Tem domains suffer most from the removal of spatial and frequency modules, while Trf is most impacted by the loss of semantic modules, underscoring the varied nature of these domain challenges.

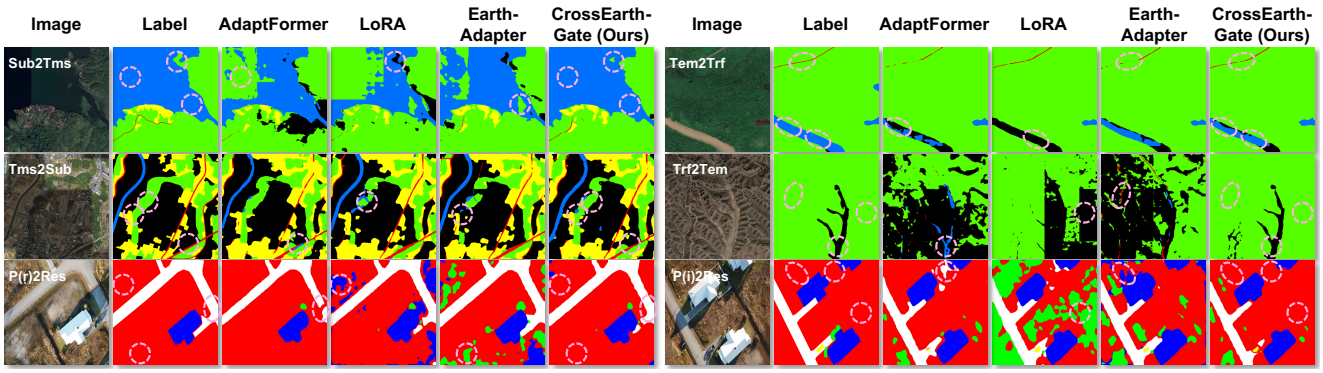


Figure 2. Visualizations of predicted segmentation maps of PEFT methods. In the CASID [43] dataset, red is the road class, yellow is the building class, blue is the water class, green is the forest class, and black is the background class. In the RescueNet [55] dataset, white is the impervious surface class, red is the clutter class, blue is the building class, green is the vegetation class, and yellow is the car class.

Table 4. Ablation studies of model backbone on CASID benchmarks. We only show the average performance of four climate domains as the source domain, respectively. We demonstrate the generalizability of CrossEarth-Gate across different backbones.

| Backbone | Method | Params (M) | Source Domain | | | | Mean |
|---|---|---|---|---|---|---|---|
| | | | Sub | Tem | Tms | Trf | |
| SatMAE (Large) [9] | Frozen | 0.0 | 24.4 | 21.5 | 24.8 | 26.3 | 24.3 |
| | Full-Tuning | 304.2 | 35.4 | 14.9 | 35.2 | 29.8 | 28.8 |
| | **CrossEarth-Gate** | 2.9-4.0 | **37.1** | **21.8** | **36.8** | **31.8** | **31.9** |
| Scale-MAE (Large) [56] | Frozen | 0.0 | 50.7 | 48.5 | 50.9 | 50.9 | 50.3 |
| | Full-Tuning | 304.2 | 52.8 | 57.6 | 51.9 | 52.6 | 53.7 |
| | **CrossEarth-Gate** | 2.9-3.6 | **56.3** | **58.1** | **54.9** | **53.8** | **55.8** |
| SAM (Huge) [35] | Frozen | 0.0 | 53.3 | 57.5 | 53.2 | 53.4 | 54.4 |
| | Full-Tuning | 631.2 | 54.7 | 58.2 | 51.7 | 54.6 | 54.8 |
| | **CrossEarth-Gate** | 4.0-4.9 | **56.1** | **61.9** | **56.3** | **59.3** | **58.4** |
| DINOv2 (Small) [50] | Frozen | 0.0 | 49.5 | 59.6 | 51.9 | 52.7 | 53.4 |
| | Full-Tuning | 22.1 | 49.4 | 49.3 | 48.7 | 50.2 | 49.4 |
| | **CrossEarth-Gate** | 0.7 | **54.6** | **62.0** | **52.7** | **55.8** | **56.2** |
| DINOv2 (Base) [50] | Frozen | 0.0 | 51.3 | 62.5 | 52.8 | 54.4 | 55.3 |
| | Full-Tuning | 86.6 | 50.2 | 53.7 | 50.2 | 49.4 | 50.9 |
| | **CrossEarth-Gate** | 1.4-1.8 | **55.6** | **63.6** | **54.7** | **57.4** | **57.8** |

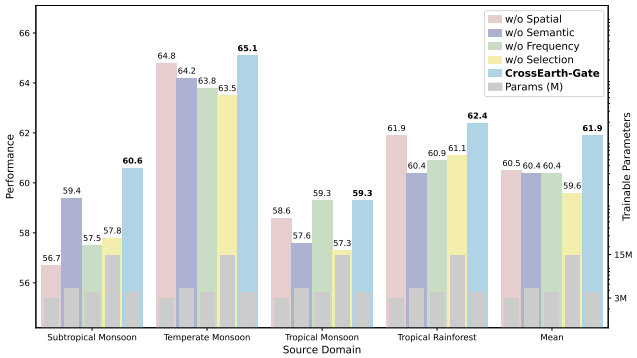


Figure 3. Ablation study of model component on CASID benchmarks. We compare the performance and trainable parameters of CrossEarth-Gate against versions with key components removed.

## 5.4. Qualitative Comparison

We also provide a qualitative comparison in Fig. 2, visualizing the predictions from our method, LoRA, AdaptFormer, and Earth-Adapter across six DG scenarios. The PEFT baselines, while effective for their intended challenge, fail outside their specialization. For example, the semantic-focused AdaptFormer fails to detect large areas of water with frequency artifacts in the Sub2Tms task. The spatial-related LoRA misclassifies the forest as water in the Tms2Sub task, resulting in semantic errors. The frequency-focused Earth-Adapter demonstrates spatial discontinuity of the road in the Tem2Trf task. Notably, CrossEarth-Gate avoids the catastrophic semantic errors, correctly captures the spatial extent of objects, and produces clean, accurate maps, mitigating the influence of artifacts. Our method consistently produces segmentation maps that are closer to the ground-truth labels across all six scenarios, demonstrating a more robust and superior generalization capability.

## 6. Conclusion

This paper addresses the limitations of current static, specialized PEFT methods in cross-domain RS adaptation. Existing approaches often fall short by focusing on a single facet of the complex spatial, semantic, and frequency shifts inherent to RS data. We introduce CrossEarth-Gate to tackle this limitation through two key innovations: a comprehensive RS module toolbox and a Fisher-guided adaptive selection mechanism. The toolbox integrates three module types to provide distinct functional pathways of models. Subsequently, by leveraging Fisher Information to quantify module importance, CrossEarth-Gate dynamically directs the gradient flow, activating only the most critical modules for the specific adaptation task. Comprehensive experiments demonstrate the effectiveness, efficiency, generalizability, and explainability of our method.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant T2125006 and Grant 42401415; in part by Shenzhen Science and Technology Program under Grant KCXFZ20240903093759004 and Grant KJZD20230923115106012; in part by Guangdong Science & Technology Program under Grant 2025B0101080001; in part by Tsinghua SIGS KA Cooperation Fund; and in part by the Research Grants Council of the Hong Kong Special Administrative Region, China (No. CUHK 14206625). Yang Liu is supported in part by the Postdoctoral Fellowship Scheme of The Chinese University of Hong Kong.

## References

[1] Kashif Abbass, Muhammad Zeeshan Qasim, Huaming Song, Muntasir Murshed, Haider Mahmood, and Ijaz Younis. A review of the global climate change impacts, adaptation, and sustainable mitigation measures. *Environmental science and pollution research*, 29(28):42539–42559, 2022. 4

[2] Alan Ansell, Edoardo Maria Ponti, Anna Korhonen, and Ivan Vulić. Composable sparse fine-tuning for cross-lingual transfer. *arXiv preprint arXiv:2110.07560*, 2021. 3

[3] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020. 2

[4] Shilei Cao, Hehai Lin, Jiashun Cheng, Yang Liu, Guowen Li, Xuehe Wang, Juepeng Zheng, Haoyuan Liang, Meng Jin, Chengwei Qin, Hong Cheng, and Haohuan Fu. Task-adaptive parameter-efficient fine-tuning for weather foundation models. In *The Fourteenth International Conference on Learning Representations*, 2026. 3

[5] Shoufa Chen, Chongjian Ge, Zhan Tong, Jiangliu Wang, Yibing Song, Jue Wang, and Ping Luo. Adaptformer: Adapting vision transformers for scalable visual recognition. *Advances in Neural Information Processing Systems*, 35:16664–16678, 2022. 1, 3, 4, 5, 6, 7

[6] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PmLR, 2020. 3

[7] Bowen Cheng, Ishan Misra, Alexander G Schwing, Alexander Kirillov, and Rohit Girdhar. Masked-attention mask transformer for universal image segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1290–1299, 2022. 5

[8] Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, et al. Palm: Scaling language modeling with pathways. *Journal of Machine Learning Research*, 24(240): 1–113, 2023. 2

[9] Yezhen Cong, Samar Khanna, Chenlin Meng, Patrick Liu, Erik Rozi, Yutong He, Marshall Burke, David Lobell, and Stefano Ermon. Satmae: Pre-training transformers for temporal and multi-spectral satellite imagery. *Advances in Neural Information Processing Systems*, 35:197–211, 2022. 1, 3, 6, 7, 8

[10] MMSegmentation Contributors. MMSegmentation: Openmmlab semantic segmentation toolbox and benchmark. https://github.com/open-mmlab/mmsegmentation, 2020. 5

[11] Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. Qlora: Efficient finetuning of quantized llms. *Advances in neural information processing systems*, 36: 10088–10115, 2023. 3

[12] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pages 4171–4186, 2019. 2

[13] Zhe Dong, Yanfeng Gu, and Tianzhu Liu. Upetu: A unified parameter-efficient fine-tuning framework for remote sensing foundation model. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–13, 2024. 3

[14] Alexey Dosovitskiy. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 3, 6, 7

[15] Ronald A Fisher. On the mathematical foundations of theoretical statistics. *Philosophical transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character*, 222(594-604):309–368, 1922. 2, 4

[16] Yunhe Gao, Xingjian Shi, Yi Zhu, Hao Wang, Zhiqiang Tang, Xiong Zhou, Mu Li, and Dimitris N Metaxas. Visual prompt tuning for test-time domain adaptation. *arXiv preprint arXiv:2210.04831*, 2022. 3

[17] Mor Geva, Roei Schuster, Jonathan Berant, and Omer Levy. Transformer feed-forward layers are key-value memories. *arXiv preprint arXiv:2012.14913*, 2020. 4

[18] Ziyang Gong, Fuhao Li, Yupeng Deng, Deblina Bhattacharjee, Xianzheng Ma, Xiangwei Zhu, and Zhenming Ji. Coda: Instructive chain-of-domain adaptation with severity-aware visual prompt tuning. In *European Conference on Computer Vision*, pages 130–148. Springer, 2024. 3

[19] Ziyang Gong, Zhixiang Wei, Di Wang, Xianzheng Ma, Hongruixuan Chen, Yuru Jia, Yupeng Deng, Zhenming Ji, Xiangwei Zhu, Naoto Yokoya, et al. Crossearth: Geospatial vision foundation model for domain generalizable remote sensing semantic segmentation. *arXiv preprint arXiv:2410.22629*, 2024. 3, 5, 6, 7

[20] Xin Guo, Jiangwei Lao, Bo Dang, Yingying Zhang, Lei Yu, Lixiang Ru, Liheng Zhong, Ziyuan Huang, Kang Wu, Dingxiang Hu, et al. Skysense: A multi-modal remote sensing foundation model towards universal interpretation for earth observation imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 27672–27683, 2024. 1, 3

[21] Zeyu Han, Chao Gao, Jinyang Liu, Jeff Zhang, and Sai Qian Zhang. Parameter-efficient fine-tuning for large models: A

comprehensive survey. *arXiv preprint arXiv:2403.14608*, 2024. 3

[22] Soufiane Hayou, Nikhil Ghosh, and Bin Yu. Lora+: Efficient low rank adaptation of large models. *arXiv preprint arXiv:2402.12354*, 2024. 3

[23] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9729–9738, 2020. 3

[24] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16000–16009, 2022. 3

[25] Ye Hong, Yatao Zhang, Konrad Schindler, and Martin Raubal. Context-aware multi-head self-attentional neural network model for next location prediction. *Transportation Research Part C: Emerging Technologies*, 156:104315, 2023. 4

[26] Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin De Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. Parameter-efficient transfer learning for nlp. In *International conference on machine learning*, pages 2790–2799. PMLR, 2019. 3, 4, 5, 6, 7

[27] Lukas Hoyer, Dengxin Dai, and Luc Van Gool. Daformer: Improving network architectures and training strategies for domain-adaptive semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9924–9935, 2022. 6, 7

[28] Lukas Hoyer, Dengxin Dai, and Luc Van Gool. Hrda: Context-aware high-resolution domain-adaptive semantic segmentation. In *European conference on computer vision*, pages 372–391. Springer, 2022. 6, 7

[29] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. Lora: Low-rank adaptation of large language models. *The International Conference on Learning Representations*, 1(2): 3, 2022. 1, 3, 4, 5, 6, 7

[30] Jiajun Hu, Jian Zhang, Lei Qi, Yinghuan Shi, and Yang Gao. Learn to preserve and diversify: Parameter-efficient group with orthogonal regularization for domain generalization. In *European Conference on Computer Vision*, pages 198–216. Springer, 2024. 3

[31] Leiyi Hu, Wanxuan Lu, Hongfeng Yu, Dongshuo Yin, Xian Sun, and Kun Fu. Tea: A training-efficient adapting framework for tuning foundation models in remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 2024. 3

[32] Leiyi Hu, Hongfeng Yu, Wanxuan Lu, Dongshuo Yin, Xian Sun, and Kun Fu. Airs: Adapter in remote sensing for parameter-efficient transfer learning. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–18, 2024. 3

[33] Xiaoxing Hu, Ziyang Gong, Yupei Wang, Yuru Jia, Gen Luo, and Xue Yang. Earth-adapter: Bridge the geospatial domain gaps with mixture of frequency adaptation. *arXiv preprint arXiv:2504.06220*, 2025. 1, 3, 4, 5, 6, 7

[34] Menglin Jia, Luming Tang, Bor-Chun Chen, Claire Cardie, Serge Belongie, Bharath Hariharan, and Ser-Nam Lim. Visual prompt tuning. In *European conference on computer vision*, pages 709–727. Springer, 2022. 3, 5, 6, 7

[35] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4015–4026, 2023. 3, 7, 8

[36] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017. 5

[37] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1): 79–86, 1951. 4

[38] Guowen Li, Xintong Liu, Yang Liu, Mengxuan Chen, Shilei Cao, Xuehe Wang, Juepeng Zheng, Jinxiao Zhang, Haoyuan Liang, Lixian Zhang, Jiuke Wang, Meng Jin, Hong Cheng, and Haohuan Fu. Tianquan-s2s: A subseasonal-to-seasonal global weather model via incorporate climatology state. In *The Fourteenth International Conference on Learning Representations*, 2026. 3

[39] Weijie Li, Wei Yang, Yuenan Hou, Li Liu, Yongxiang Liu, and Xiang Li. Saratr-x: Towards building a foundation model for sar target recognition. *IEEE Transactions on Image Processing*, 2025. 3

[40] Yu Liang, Shilei Cao, Juepeng Zheng, Xiucheng Zhang, Jianxi Huang, and Haohuan Fu. Low saturation confidence distribution-based test-time adaptation for cross-domain remote sensing image classification. *International Journal of Applied Earth Observation and Geoinformation*, 139: 104463, 2025. 1

[41] Fan Liu, Delong Chen, Zhangqingyun Guan, Xiaocong Zhou, Jiale Zhu, Qiaolin Ye, Liyong Fu, and Jun Zhou. Remoteclip: A vision language foundation model for remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–16, 2024. 6, 7

[42] Peiyu Liu, Ze-Feng Gao, Wayne Xin Zhao, Zhi-Yuan Xie, Zhong-Yi Lu, and Ji-Rong Wen. Enabling lightweight fine-tuning for pre-trained language model compression based on matrix product operators. *arXiv preprint arXiv:2106.02205*, 2021. 3

[43] Songlin Liu, Linwei Chen, Li Zhang, Jun Hu, and Ying Fu. A large-scale climate-aware satellite image dataset for domain adaptive land-cover semantic segmentation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 205:98–114, 2023. 1, 8

[44] Shih-Yang Liu, Chien-Yi Wang, Hongxu Yin, Pavlo Molchanov, Yu-Chiang Frank Wang, Kwang-Ting Cheng, and Min-Hung Chen. Dora: Weight-decomposed low-rank adaptation. In *Forty-first International Conference on Machine Learning*, 2024. 3

[45] Yang Liu, Zinan Zheng, Jiashun Cheng, Fugee Tsung, Deli Zhao, Yu Rong, and Jia Li. Cirt: Global subseasonal-to-

seasonal forecasting with geometry-inspired transformer. In *The Thirteenth International Conference on Learning Representations*, 2025. 3

[46] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. 5

[47] Haoyan Luo and Lucia Specia. From understanding to utilization: A survey on explainability for large language models. *arXiv preprint arXiv:2401.12874*, 2024. 4

[48] Oscar Manas, Alexandre Lacoste, Xavier Giró-i Nieto, David Vazquez, and Pau Rodriguez. Seasonal contrast: Unsupervised pre-training from uncurated remote sensing data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9414–9423, 2021. 1, 3

[49] Kevin Meng, David Bau, Alex Andonian, and Yonatan Belinkov. Locating and editing factual associations in gpt. *Advances in neural information processing systems*, 35:17359–17372, 2022. 4

[50] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023. 3, 5, 6, 7, 8

[51] Razvan Pascanu and Yoshua Bengio. Revisiting natural gradient for deep networks. *arXiv preprint arXiv:1301.3584*, 2013. 4

[52] Jonas Pfeiffer, Aishwarya Kamath, Andreas Rücklé, Kyunghyun Cho, and Iryna Gurevych. Adapterfusion: Non-destructive task composition for transfer learning. *arXiv preprint arXiv:2005.00247*, 2020. 3

[53] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PmLR, 2021. 2

[54] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, 21(140):1–67, 2020. 2

[55] Maryam Rahnemoonfar, Tashnim Chowdhury, and Robin Murphy. Rescuenet: A high resolution uav semantic segmentation dataset for natural disaster damage assessment. *Scientific data*, 10(1):913, 2023. 8

[56] Colorado J Reed, Ritwik Gupta, Shufan Li, Sarah Brockman, Christopher Funk, Brian Clipp, Kurt Keutzer, Salvatore Candido, Matt Uyttendaele, and Trevor Darrell. Scale-mae: A scale-aware masked autoencoder for multiscale geospatial representation learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4088–4099, 2023. 1, 3, 6, 7, 8

[57] Linus Scheibenreif, Michael Mommert, and Damian Borth. Masked vision transformers for hyperspectral image classification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2166–2176, 2023. 3

[58] Linus Scheibenreif, Michael Mommert, and Damian Borth. Parameter efficient self-supervised geospatial domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 27841–27851, 2024. 3, 5, 6, 7

[59] Jiachen Sun, Mark Ibrahim, Melissa Hall, Ivan Evtimov, Z Morley Mao, Cristian Canton Ferrer, and Caner Hazirbas. Vpa: Fully test-time visual prompt adaptation. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 5796–5806, 2023. 3

[60] Yi-Lin Sung, Varun Nair, and Colin A Raffel. Training neural networks with fixed sparse masks. *Advances in Neural Information Processing Systems*, 34:24193–24205, 2021. 3, 5

[61] Maofeng Tang, Andrei Cozma, Konstantinos Georgiou, and Hairong Qi. Cross-scale mae: A tale of multiscale exploitation in remote sensing. *Advances in Neural Information Processing Systems*, 36:20054–20066, 2023. 3

[62] Romain Thoreau, Valerio Marsocci, and Dawa Derksen. Parameter-efficient adaptation of geospatial foundation models through embedding deflection. *arXiv preprint arXiv:2503.09493*, 2025. 3

[63] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023. 2

[64] Wilhelm Tranheden, Viktor Olsson, Juliano Pinto, and Lennart Svensson. Dacs: Domain adaptation via cross-domain mixed sampling. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 1379–1389, 2021. 5

[65] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 2, 3

[66] Di Wang, Jing Zhang, Minqiang Xu, Lin Liu, Dongsheng Wang, Erzhong Gao, Chengxi Han, Haonan Guo, Bo Du, Dacheng Tao, et al. Mtp: Advancing remote sensing foundation model via multitask pretraining. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17:11632–11654, 2024. 6, 7

[67] Di Wang, Meiqi Hu, Yao Jin, Yuchun Miao, Jiaqi Yang, Yichu Xu, Xiaolei Qin, Jiaqi Ma, Lingyu Sun, Chenxing Li, et al. Hypersigma: Hyperspectral intelligence comprehension foundation model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025. 3

[68] Junjue Wang, Zhuo Zheng, Ailong Ma, Xiaoyan Lu, and Yanfei Zhong. Loveda: A remote sensing land-cover dataset for domain adaptive semantic segmentation. *arXiv preprint arXiv:2110.08733*, 2021. 1

[69] Yi Wang, Hugo Hernández Hernández, Conrad M Albrecht, and Xiao Xiang Zhu. Feature guided masked autoencoder for self-supervised learning in remote sensing. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024. 3

[70] Xiu-Shen Wei, Yu-Yan Xu, Chen-Lin Zhang, Gui-Song Xia, and Yu-Xin Peng. Cat: A coarse-to-fine attention tree for

semantic change detection. *Visual Intelligence*, 1(1):3, 2023. 2

[71] Zhixiang Wei, Lin Chen, Yi Jin, Xiaoxiao Ma, Tianle Liu, Pengyang Ling, Ben Wang, Huaian Chen, and Jinjin Zheng. Stronger fewer & superior: Harnessing vision foundation models for domain generalized semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 28619–28630, 2024. 3, 5, 6, 7

[72] Walter F Wiggins and Ali S Tejani. On the opportunities and risks of foundation models for natural language processing in radiology. *Radiology: Artificial Intelligence*, 4(4):e220119, 2022. 2

[73] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in neural information processing systems*, 34: 12077–12090, 2021. 6, 7

[74] Yi Xin, Siqi Luo, Haodi Zhou, Junlong Du, Xiaohong Liu, Yue Fan, Qing Li, and Yuntao Du. Parameter-efficient fine-tuning for pre-trained vision models: A survey. *arXiv e-prints*, pages arXiv–2402, 2024. 1

[75] Runxin Xu, Fuli Luo, Zhiyuan Zhang, Chuanqi Tan, Baobao Chang, Songfang Huang, and Fei Huang. Raise a child in large language model: Towards effective and generalizable fine-tuning. *arXiv preprint arXiv:2109.05687*, 2021. 3

[76] Senqiao Yang, Jiarui Wu, Jiaming Liu, Xiaoqi Li, Qizhe Zhang, Mingjie Pan, Yulu Gan, Zehui Chen, and Shanghang Zhang. Exploring sparse visual prompt for domain adaptive dense prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 16334–16342, 2024. 3

[77] Yunzhi Yao, Ningyu Zhang, Zekun Xi, Mengru Wang, Ziwen Xu, Shumin Deng, and Huajun Chen. Knowledge circuits in pretrained transformers. *Advances in Neural Information Processing Systems*, 37:118571–118602, 2024. 4

[78] Dongshuo Yin, Leiyi Hu, Bin Li, Youqun Zhang, and Xue Yang. 5%¿ 100%: Breaking performance shackles of full fine-tuning on visual recognition tasks. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 20071–20081, 2025. 3

[79] Dongshuo Yin, Ting-Feng Zhao, Deng-Ping Fan, Shutao Li, Bo Du, Xian Sun, and Shi-Min Hu. Remote sensing tuning: A survey. *Computational Visual Media*, 2025. 3

[80] Elad Ben Zaken, Shauli Ravfogel, and Yoav Goldberg. Bitfit: Simple parameter-efficient fine-tuning for transformer-based masked language-models. *arXiv preprint arXiv:2106.10199*, 2021. 3

[81] Dan Zhang, Tao Feng, Lilong Xue, Yuandong Wang, Yuxiao Dong, and Jie Tang. Parameter-efficient fine-tuning for foundation models. *arXiv preprint arXiv:2501.13787*, 2025. 1, 3

[82] Longteng Zhang, Lin Zhang, Shaohuai Shi, Xiaowen Chu, and Bo Li. Lora-fa: Memory-efficient low-rank adaptation for large language models fine-tuning. *arXiv preprint arXiv:2308.03303*, 2023. 3

[83] Lixian Zhang, Yi Zhao, Runmin Dong, Jinxiao Zhang, Shuai Yuan, Shilei Cao, Mengxuan Chen, Juepeng Zheng, Weijia Li, Wayne Zhang, et al. A 2-mae: A spatial-temporal-spectral unified remote sensing pre-training method based on anchor-aware masked autoencoder. *IEEE Transactions on Geoscience and Remote Sensing*, 2026. 3

[84] Yuxiang Zhang, Wei Li, Mengmeng Zhang, Jiawei Han, Ran Tao, and Shunlin Liang. Spectralx: Parameter-efficient domain generalization for spectral remote sensing foundation models. *arXiv preprint arXiv:2508.01731*, 2025. 3

[85] Bingchen Zhao, Haoqin Tu, Chen Wei, Jieru Mei, and Cihang Xie. Tuning layernorm in attention: Towards efficient multi-modal llm finetuning. *arXiv preprint arXiv:2312.11420*, 2023. 3

[86] Henry Hengyuan Zhao, Pichao Wang, Yuyang Zhao, Hao Luo, Fan Wang, and Mike Zheng Shou. Sct: A simple baseline for parameter-efficient fine-tuning via salient channels. *International Journal of Computer Vision*, 132(3):731–749, 2024. 3

[87] Zihan Zhong, Zhiqiang Tang, Tong He, Haoyang Fang, and Chun Yuan. Convolution meets loRA: Parameter efficient finetuning for segment anything model. In *The Twelfth International Conference on Learning Representations*, 2024. 3