Parameter-Efficient Adaptation of Geospatial Foundation Models through Embedding Deflection

Romain Thoreau[1,*] Valerio Marsocci[2,*] Dawa Derksen[1]

[1]Data Campus - CNES, the French Space Agency [2]Φ-lab - ESA, the European Space Agency
[*] Equal contribution

ICCV OCT 19-23, 2025 HONOLULU HAWAII

Introduction

Geospatial foundation models (GFMs), have become ubiquitous, as the availability of satellite data have skyrocketed.

GFMs need to be finetuned, stored, and deployed, which motivates **parameter-efficient finetuning** (PEFT).

In this work, *we adapt GFMs pretrained on RGB satellite images to multispectral images* for environmental applications, by integrating **inductive biases**.

Preliminaries

Let us consider a downstream task, for which we have a labeled dataset of multispectral images. To solve the task, we consider a **GFM**, built on a Vision Transformer (**ViT**) and **pretrained on RGB** satellite images.

Data processing in ViTs can be considered in three stages: i) dimensionality change (from the image $X \in \mathbb{R}^{C \times H \times W}$ to the patch embeddings $\boldsymbol{x} \in \mathbb{R}^{n \times d}$, ii) data transport, and iii) task-specific predictions.

ViTs transport the embeddings in the latent space through attention blocks: the input of the l^{th} attention block is denoted as $\boldsymbol{z}^{(l)}$, *e.g.* $\boldsymbol{z}^{(1)} = \boldsymbol{x}$. The self-attention module, comprising query, key and value matrices denoted as $W_l^Q, W_l^K, W_l^V \in \mathbb{R}^{d \times d}$, respectively, computes a first displacement:

$$\Delta_1 \boldsymbol{z}_i^{(l)} = \sum_{j=1}^{n} \frac{\exp(\alpha_{ij})}{\sum_{j'=1}^{n} \exp(\alpha_{ij'})} (\boldsymbol{z}_j^{(l)} W_l^V), \text{ where } \alpha_{ij} = \frac{1}{\sqrt{d}} (\boldsymbol{z}_i^{(l)} W_l^Q)(\boldsymbol{z}_j^{(l)} W_l^K)^T. \tag{1}$$

Second, a standard MLP computes a second displacement:

$$\Delta_2 \boldsymbol{z}_i^{(l)} = \text{MLP}^{(l)}(\boldsymbol{z}_i^{(l)} + \Delta_1 \boldsymbol{z}_i^{(l)}) \tag{2}$$

Embedding decomposition

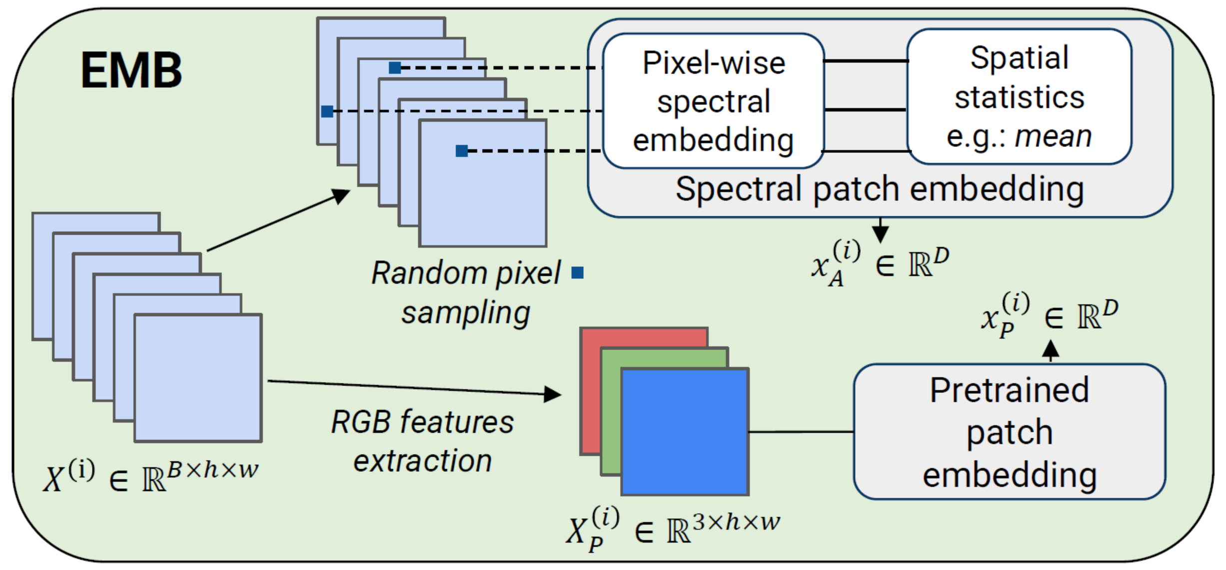
We assume that the patch embeddings $\boldsymbol{x}$, computed by an arbitrary neural network, can be decomposed as follows:

$$\boldsymbol{x} = \boldsymbol{x}_P + \boldsymbol{x}_A \tag{3}$$

where $\boldsymbol{x}_P$ encodes the *radiometric* and *geometric* information of the **RGB** channels, while $\boldsymbol{x}_A$ encodes the *spectral* information beyond RGB.

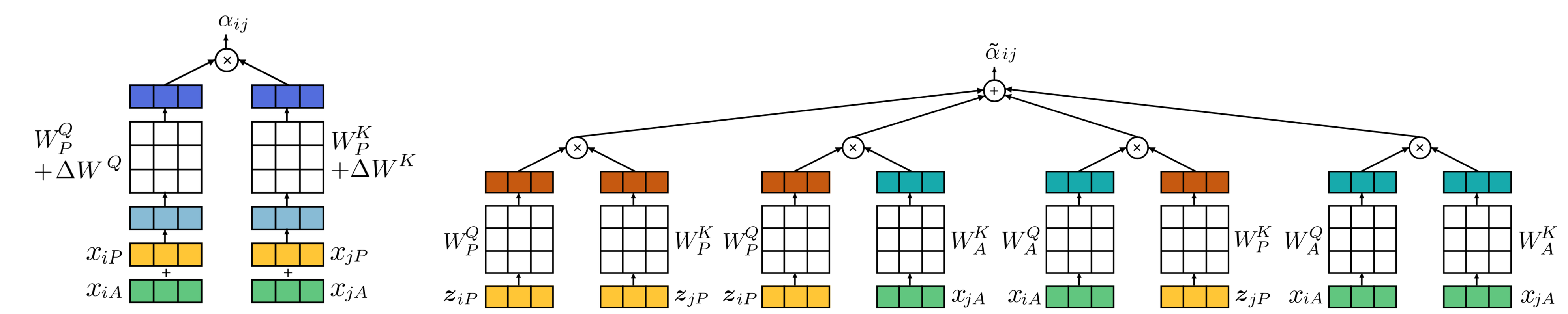
Untangled Patch Embedding (EMB)

In **EMB**, we compute the embeddings $\boldsymbol{x}_P$ and $\boldsymbol{x}_A$, as shown below. We aim to extract the radiometric information only of spectral channels beyond RGB, without overloading the model with new parameters.



Untangled Attention (uAtt)

uAtt mitigates the effects of the linear arithmetic of self-attention. It introduces new parameters to process the auxiliary spectral information. This yields extra **RGB-to-spectral**, **spectral-to-RGB** and **spectral-to-spectral** attention products. It can be combined with low-rank updates.



(a) Low-rank adaptation with standard attention (1st layer)

(b) DEFLECT adaptation with untangled attention

Figure 1. Illustration of the attention product during finetuning.

Embedding Deflection

To preserve the structure of the pretrained GFM latent space, we *constrain the norm of the displacement computed* by adapted uAtt blocks to match the one pretrained standard attention blocks would have computed.

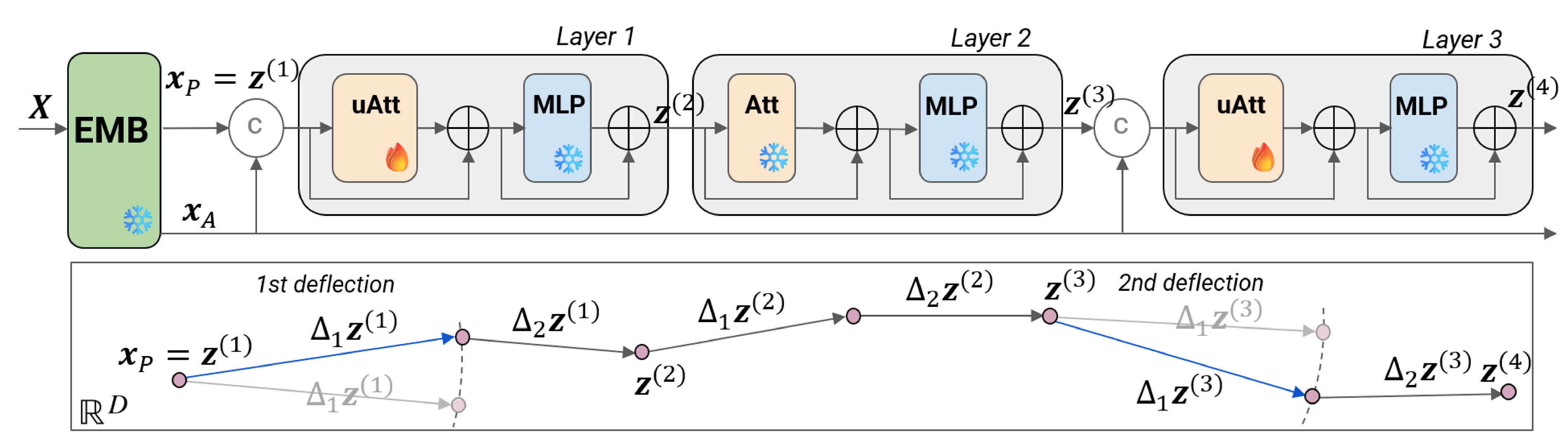


Figure 2. Illustration of DEFLECT, our parameter-efficient finetuning method.

Conclusions

We showed several benefits of DEFLECT compared to competing methods:

- **more stable performance** across tasks, datasets, models, and weights initialization,
- **higher accuracy** with *5-10× less tuned parameters* that low-rank based techniques.

DEFLECT is also consistent across low-rank dimensions.

Limitations of DEFLECT include i) higher FLOPs compared to low-rank techniques, and ii) a sensitivity to the choice of adapted layers.

In future work, we will experiment DEFLECT on *hyperspectral* and *multi-temporal* satellite images.

Numerical experiments

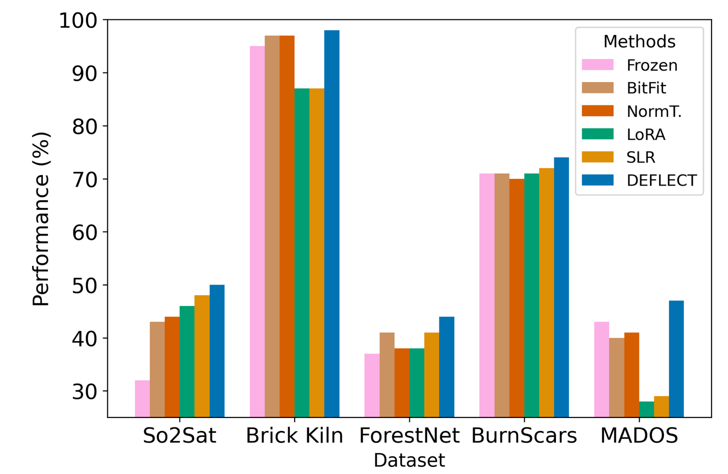


Figure 3. Performance averaged across models (Cross-scale MAE, Scale-MAE and DINO-MC).

Table 1. Comparison of F1-score (classification) and IoU (segmentation) results across downstream tasks for DEFLECT and competing PEFT methods using Scale-MAE [1], with standard deviation across 3 runs.

Method	Encoder Tuned Params	So2Sat (mF1)	Brick Kiln (mF1)	ForestNet (mF1)	Burn Scars (IoU)	MADOS (mIoU)	Avg. Class.	Avg. Segm.	Avg. Perf.
Finetuning (oracle)	*100%*	*52.7* ±0.1	*96.6* ±2.1	*44.9* ±0.2	*78.6* ±1.2	*46.2* ±2.5	*64.8*	*62.4*	*63.8*
Frozen	0.0%	32.1 ±0.6	94.3 ±0.2	41.3 ±0.5	75.3 ±2.2	36.3 ±0.7	55.9	55.8	55.9
Norm Tuning [2]	0.03%	45.5 ±6.6	97.0 ±0.5	41.8 ±0.8	70.7 ±3.8	30.6 ±21.1	61.4	50.6	57.1
BitFit [3]	0.09%	41.0 ±5.4	97.4 ±0.5	41.9 ±0.9	75.9 ±0.1	28.3 ±15.4	60.1	52.1	56.9
LoRA [4]	2.1%	**54.7** ±0.2	97.4 ±0.1	39.4 ±3.2	78.8 ±2.2	45.2 ±0.5	63.8	62.0	63.1
SLR [5]	2.2%	52.0 ±0.5	96.4 ±1.9	43.0 ±1.0	**80.3** ±0.7	45.5 ±2.2	63.8	62.9	63.4
DEFLECT (ours)	0.2%	52.7 ±0.5	**97.6** ±0.2	**43.2** ±0.9	77.0 ±1.2	**50.5** ±0.2	**64.5**	**63.7**	**64.2**

References

[1] C. J. Reed, R. Gupta, S. Li, S. Brockman, C. Funk, B. Clipp, K. Keutzer, S. Candido, M. Uyttendaele, and T. Darrell, "Scale-mae: A scale-aware masked autoencoder for multiscale geospatial representation learning," 2023.

[2] B. Zhao, H. Tu, C. Wei, J. Mei, and C. Xie, "Tuning layernorm in attention: Towards efficient multi-modal llm finetuning," 2023.

[3] E. B. Zaken, S. Ravfogel, and Y. Goldberg, "Bitfit: Simple parameter-efficient fine-tuning for transformer-based masked language-models," *arXiv preprint arXiv:2106.10199*, 2021.

[4] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "Lora: Low-rank adaptation of large language models," 2021.

[5] L. Scheibenreif, M. Mommert, and D. Borth, "Parameter efficient self-supervised geospatial domain adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 27841–27851, June 2024.

[6] N. Houlsby, A. Giurgiu, S. Jastrzebski, B. Morrone, Q. De Laroussilhe, A. Gesmundo, M. Attariyan, and S. Gelly, "Parameter-efficient transfer learning for nlp," in *International conference on machine learning*, pp. 2790–2799, PMLR, 2019.

romain.thoreau@agroparistech.fr, valerio.marsocci@esa.int