REUSE OR RELEARN? A SPECTRAL VIEW OF EARTH OBSERVATION FOUNDATION MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Foundation models are rarely used as generic, frozen feature extractors; instead, they are fine-tuned for the target downstream application. This practice is particularly prevalent in Earth observation (EO), and it raises a question that downstream accuracy alone cannot answer: does fine-tuning reuse the pretrained representation, or does it relearn a new one? We study this with spectral diagnostics that compare a model before and after adaptation, quantifying how well its dominant singular subspaces are preserved, how broadly the weight update is distributed, and how large it is. Using natural image models such as CLIP and DINO as a reference, we find that, under the evaluated fine-tuning settings, EO models undergo far larger, higher-rank updates and retain much less of their pretrained structure, so their downstream performance is often obtained with substantial changes to the pretrained weight structure. The diagnostics further provide insight into how cheaply a model can be adapted: where the pretrained subspaces are preserved, adapting a small fraction of the parameters can match full fine-tuning, and where they are not, it can fall behind. More broadly, foundation models, and EO foundation models in particular, should be assessed not only by benchmark accuracy, but also by how reusable their pretrained representation is.

1 INTRODUCTION

Foundation models aim to learn representations that can be reused across downstream tasks (Bommasani et al., 2021). This promise has motivated large-scale pretraining in Earth observation (EO), with models intended to transfer across sensors, regions, and applications (Cong et al., 2022; Fuller et al., 2023; Szwarcman et al., 2025). However, strong performance after fine-tuning does not reveal how much of the pretrained structure is retained or how it contributes to solving the downstream task. Head-only evaluation also provides an incomplete picture. A linear head measures how readily the task can be solved from frozen features, which may favor representations already aligned with that task without capturing their usefulness after adaptation. More expressive decoders, such as UPerNet (Xiao et al., 2018), introduce substantial task-specific learning capacity, making it harder to attribute strong performance specifically to the pretrained features. Understanding reuse therefore requires examining both the benefits of pretraining and the changes through which the pretrained model adapts to a new task.

Work in AI interpretability provides a basis for this analysis. Learned features have been associated with linear directions and subspaces, including interpretable singular directions in transformer weights (Elhage et al., 2022; Millidge & Black, 2022), while spectral analyses indicate that information is distributed across singular modes (Staats et al., 2025). Singular-value fine-tuning builds on this perspective by adjusting the strength of pretrained modes while preserving their directions (Sun et al., 2022; Turkoglu et al., 2026). These findings motivate studying whether pretrained singular subspaces remain useful during adaptation.

We propose a *spectral diagnostic toolkit* that characterizes adaptation along three complementary dimensions. The *Spectral Stability Index* (SSI) measures preservation of dominant pretrained subspaces, the *Effective Rank of Adaptation* (ERA) measures how many directions contribute to the update, and relative weight change measures its magnitude. Comparisons across pretrained model generations and downstream task requirements ground their interpretation. Together with downstream performance, these diagnostics provide a structural view of representation reuse.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

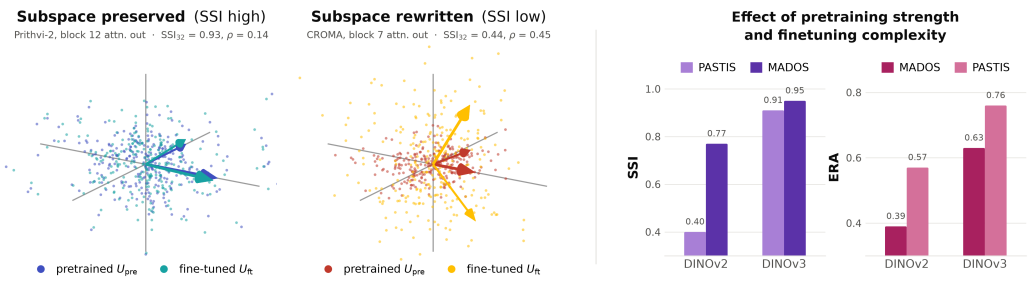


Figure 1: **Left:** One attention output projection $W \in \mathbb{R}^{n\times n}$ per panel, pretrained and after full fine-tuning on MADOS (Prithvi-2 block 12, $n = 768$ and CROMA block 7, $n = 1024$). Points are 260 sampled columns of W, arrows are the two leading left singular vectors, scaled by their singular values ($U_2\Sigma_2$). Both are projected onto the 3-D subspace that best fits the four arrows. SSI_k measures how well the pretrained and fine-tuned top-k subspaces align, independently of singular-value magnitude. **Right:** DINOv3 exhibits higher SSI than DINOv2 on both datasets, suggesting that stronger pretraining is associated with greater preservation of pretrained subspaces during fine-tuning. Both models exhibit higher ERA on PASTIS than on MADOS, suggesting that adaptation to the more complex task involves a broader range of update directions.

We apply this framework to full fine-tuning across five EO benchmarks and compare downstream performance with parameter-efficient alternatives: singular-value fine-tuning (SVF) and LoRA (Hu et al., 2022). Experiments with three computer vision (CV) backbones provide a reference for adaptation and reuse. To assess which benchmarks reveal the value of pretraining, we compare pretrained models with architecture-matched, randomly initialized counterparts. We also provide additional analyses, including experiments on cross-modal transfer. Our contributions are twofold:

1. **A spectral perspective on the usefulness and reusability of pretraining.** We introduce and empirically ground a diagnostic toolkit for examining how pretrained weight structure supports downstream adaptation, extending evaluation beyond task-level performance.
2. **Distinct adaptation behaviors in EO and CV foundation models.** We show that the evaluated EO models generally undergo larger updates with higher effective ranks and weaker subspace preservation, suggesting less mature representation reuse than in the CV reference. Our analyses also identify cases where strong performance requires comparatively limited changes, highlighting opportunities for more reusable pretraining.

2 Spectral Diagnostics of Fine-Tuning

We study how fine-tuning modifies a pretrained encoder in weight space. For a pretrained weight matrix $W_{\text{pre}} \in \mathbb{R}^{m\times n}$ and its fine-tuned counterpart W_{ft}, we define the update

$$\Delta W = W_{\text{ft}} - W_{\text{pre}}. \tag{1}$$

We characterize this update along three complementary dimensions: *how much* the weights change, *whether* their dominant pretrained subspaces are preserved, and *how many directions* participate in the update. These are measured by relative weight change, the Spectral Stability Index (SSI), and the Effective Rank of Adaptation (ERA), respectively.

2.1 Relative Weight Change

We first measure the magnitude of adaptation relative to the scale of the pretrained weights:

$$\rho = \frac{\|\Delta W\|_F}{\|W_{\text{pre}}\|_F}. \tag{2}$$

The relative weight change ρ quantifies how strongly a matrix is modified during fine-tuning. Small values indicate that the fine-tuned weights remain close to their pretrained initialization, whereas larger values indicate more substantial changes.

2.2 Spectral Stability Index (SSI)

Relative weight change measures the size of an update, but not whether fine-tuning preserves the dominant directions of the pretrained weights. To capture this, we compare the leading singular subspaces before and after fine-tuning. We write the singular value decompositions as

$$W_{\text{pre}} = U_{\text{pre}}\Sigma_{\text{pre}}V_{\text{pre}}^{\top}, \qquad W_{\text{ft}} = U_{\text{ft}}\Sigma_{\text{ft}}V_{\text{ft}}^{\top}, \tag{3}$$

with singular vectors ordered by decreasing singular value. The first k columns of U and V define the dominant left and right singular subspaces. Their principal-angle cosines are given by

$$\cos\theta_i^{(L)} = \sigma_i\big(U_{\text{pre},k}^{\top}U_{\text{ft},k}\big), \qquad \cos\theta_i^{(R)} = \sigma_i\big(V_{\text{pre},k}^{\top}V_{\text{ft},k}\big), \tag{4}$$

where $\sigma_i(\cdot)$ denotes the i-th singular value (A.2). We define the Spectral Stability Index (SSI) as

$$\text{SSI}_k = \frac{1}{2}\left[\frac{1}{k}\sum_{i=1}^{k}\cos^2\theta_i^{(L)} + \frac{1}{k}\sum_{i=1}^{k}\cos^2\theta_i^{(R)}\right]. \tag{5}$$

$\text{SSI}_k \in [0,1]$ measures the overlap between the dominant pretrained and fine-tuned weight subspaces: values close to one indicate strong preservation, while lower values indicate greater rotation away from the pretrained subspace. We use $k = 32$ to focus on the dominant singular directions, which capture the strongest linear transformations of each weight matrix. Increasing k includes progressively weaker modes and dilutes this focus; at $k = \min(m,n)$, at least one subspace overlap becomes trivially equal to one because the corresponding subspaces span the full ambient space. We report sensitivity to $k \in 64, 128$ in the Appendix B.2.

2.3 Effective Rank of Adaptation (ERA)

SSI measures how much of the pretrained weight subspace is preserved, whereas ERA measures how many directions are involved in the weight update. We denote the singular values of ΔW by $\lambda_1 \geq \lambda_2 \geq \cdots$. The singular values are normalized as $p_i = \lambda_i / \sum_j \lambda_j$, with entropy $H = -\sum_i p_i \log p_i$. Following the entropy-based definition of effective rank, we define

$$\text{ERA} = \frac{\exp(H)}{\min(m,n)}. \tag{6}$$

The numerator $\exp(H)$ is the entropy-based effective rank of Roy & Vetterli (2007), providing a soft estimate of the number of active singular directions in ΔW. We normalize it by $\min(m,n)$, the maximum possible rank, to obtain a score comparable across matrix sizes. Low ERA therefore indicates that the update is concentrated in a small number of directions, while high ERA indicates that the update is distributed across a larger fraction of the available weight space.

2.4 Joint Interpretation

The three diagnostics capture complementary aspects of adaptation: ρ measures update magnitude, SSI measures preservation of dominant pretrained subspaces, and ERA measures the dimensional breadth of the update. We expect a reusable foundation model to provide a feature basis that can be readily reweighted and combined to solve new tasks, supporting efficient adaptation with limited changes to the pretrained structure. High SSI, together with strong downstream performance, is consistent with this expectation. Modest ρ and low ERA further indicate small, concentrated adjustments, whereas low SSI and high ERA indicate broader reorganization. We interpret these diagnostics jointly with performance, since high SSI alone can also reflect insufficient adaptation. All diagnostics are computed independently for each weight matrix and aggregated over attention and MLP layers. Details of matrix extraction, aggregation, and the choice of k are provided in Appendix A.

3 Experiment

We first examine the spectral diagnostics, then assess benchmark suitability and compare adaptation across EO and CV models.

		Resolution	
		0.25°	0.1°
SSI	Attn	0.992	0.902
	MLP	0.997	0.953
ERA	Attn	0.150	0.528
	MLP	0.141	0.762
ρ	Attn	0.055	0.218
	MLP	0.060	0.235

Table 1: Aurora adaptation from the same $0.25°$ pretrained checkpoint. Higher-resolution forecasting exhibits lower SSI and higher ERA and relative weight change.

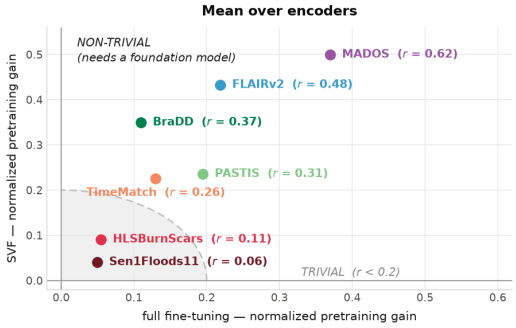


Figure 2: Normalized pretraining gains over architecture-matched random initialization under full fine-tuning and SVF, averaged across encoders.

Building on PANGAEA (Marsocci et al., 2026), our EO evaluation covers optical segmentation at different spatial resolutions with MADOS (Kikaki et al., 2024) and FLAIR-2 (Garioud et al., 2023), temporal crop mapping with PASTIS (Sainte Fare Garnot & Landrieu, 2021), pixel-level time-series classification with TimeMatch (Nyborg et al., 2022), and SAR segmentation with BraDD (Karaman et al., 2023). We select CROMA (Fuller et al., 2023), DOFA (Xiong et al., 2024), Prithvi-2 (Szwarcman et al., 2025), TerraMind (Jakubik et al., 2025), and SSL4EO-MAE (Wang et al., 2023) to cover a broad range of EO pretraining approaches. The selection rationale and additional encoders are detailed in Appendix E. Additional experiments examine SSI rank and learning rate sensitivity, decoder capacity, convolutional architectures, and cross-modal transfer in Appendix B.

For the CV reference, we use three widely established vision backbones: CLIP (Radford et al., 2021), DINOv2 (Oquab et al., 2024), and DINOv3 (Siméoni et al., 2025). We evaluate them on standard classification benchmarks (Flowers-102, CIFAR-10, and Oxford-IIIT Pets) Nilsback & Zisserman (2008); Krizhevsky (2009); Parkhi et al. (2012) and on PASCAL VOC Everingham et al. (2010) for semantic segmentation.

We compare full fine-tuning (FT) with two parameter-efficient adaptation methods chosen for complementary reasons. SVF is particularly relevant to our analysis because it preserves the pretrained singular vectors and adapts only their associated singular values, directly testing whether a downstream task can be solved by reweighting existing pretrained directions. LoRA provides a widely adopted low-rank adaptation baseline that permits the model to introduce new update directions. Together, they contrast reuse of the pretrained spectral basis with more flexible low-rank adaptation. Spectral diagnostics characterize full fine-tuning, with SSI reported at $k = 32$. We use standard learning rates and decay schedules and monitor validation performance to assess convergence. Full optimization details, as well as dataset splits, model configurations, and adaptation settings, are provided in Appendix F.

3.1 Interpreting the Spectral Diagnostics

We first examine whether SSI and ERA reflect meaningful differences in pretraining and downstream adaptation. We consider two complementary comparisons: changing the pretrained model while holding the downstream task fixed, and changing the downstream task while holding the pretrained model fixed.

SSI captures preservation of dominant pretrained subspaces. Moving from DINOv2 to DINOv3 increases SSI from 0.771 to 0.951 on MADOS, while test mIoU improves from 41.9 to 64.9, and from 0.400 to 0.908 on PASTIS, with mIoU increasing from 31.5 to 33.6 (Fig. 1). These observations are consistent with stronger pretraining providing directions that require less modification during adaptation, although the comparison also includes differences in architecture and pretraining recipe. Aurora (Bodnar et al., 2025), a large-scale weather foundation model, provides a complementary example: adapting its $0.25°$ pretrained model to $0.1°$ forecasting yields lower SSI than

fine-tuning at $0.25°$, in both attention ($0.992 \rightarrow 0.902$) and MLP layers ($0.997 \rightarrow 0.953$). Refer to Appendix B.1 for additional Aurora experiments on other tasks.

ERA captures the dimensional breadth of adaptation. For DINOv2 and DINOv3, ERA increases from MADOS to PASTIS, from 0.386 to 0.572 and from 0.630 to 0.763, respectively (Fig. 1). The temporal crop-mapping task therefore involves updates distributed across more singular directions than the single-frame segmentation task. Similarly, Aurora's higher-resolution adaptation increases ERA from 0.150 to 0.528 in attention and from 0.141 to 0.762 in MLP layers. Together, these examples illustrate the complementary roles of the diagnostics: SSI measures whether dominant pretrained subspaces persist, whereas ERA measures how broadly the update is distributed. High subspace preservation can coexist with a high-dimensional update, as observed for DINOv3 on PASTIS.

3.2 Which Benchmarks Reveal the Value of Pretraining?

Not every downstream benchmark is equally informative about the value of a foundation model. Geospatial foundation models are evaluated on a highly fragmented and inconsistently reported set of benchmarks, with little agreement on which tasks should be used for comparison (Corley et al., 2026). Moreover, relatively simple tasks may already be solved well by models trained from scratch, so high scores alone do not demonstrate that large-scale pretraining has produced useful transferable features. We therefore first assess whether each benchmark reveals a benefit from pretraining, and restrict our main analysis to those that do.

We compare pretrained and architecture-matched, randomly initialized models under full fine-tuning and SVF. We compute the normalized pretraining gain as $\text{gain} = (s_{\text{pre}} - s_{\text{rand}})/(s_{\text{pre}} - s_{\text{chance}})$, where s_{pre} and s_{rand} are downstream scores (accuracy or mIoU, in percent) obtained with pretrained and random initialization, and s_{chance} is the corresponding chance-level score. This expresses the improvement over training from scratch as a fraction of the pretrained model's performance above chance.

Averaged across encoders, Fig. 2 shows small gains on HLSBurnScars (Phillips et al., 2023) and Sen1Floods11 (Bonafilia et al., 2020), but clearer benefits on MADOS, FLAIR-2, BraDD, PASTIS, and TimeMatch under the evaluated training budgets. We therefore focus our main analysis on these five benchmarks. MADOS exhibits the largest gains under both adaptation methods, exceeding those on the more complex temporal task PASTIS. Its limited labeled training data may make pretrained features particularly valuable, suggesting that task complexity alone does not determine the benefit of pretraining.

3.3 Main Results

EO models substantially reorganize pretrained subspaces. Across all EO benchmarks, full fine-tuning departs far from the pretrained representation (Tab. 2–4). Averaged over datasets, SSI is only 0.543/0.548 for attention/MLP layers, whereas the vision models retain values close to one (Tab. 5), and ERA and relative weight change ρ are several times higher. Although the benchmarks are not matched across domains, the gap is consistent: EO adaptation typically reorganizes the dominant pretrained subspaces instead of adjusting features within a stable basis.

Reuse diagnostics provide insight into parameter-efficient transfer. The diagnostics also indicate when a pretrained model can be adapted cheaply. When full fine-tuning preserves the dominant subspaces (high SSI) with small, low-dimensional updates (low ρ and ERA), the task-relevant structure may already be present in the pretrained weights, and parameter-efficient methods can recover full fine-tuning performance. MADOS is the only EO benchmark where all three diagnostics point towards reuse, combining the highest SSI (0.765/0.815) with the most compact updates (ERA 0.464/0.596) and the smallest weight change (ρ = 0.284/0.277). Here, both LoRA (58.5) and SVF (57.7) exceed full fine-tuning (56.9), even though SVF updates fewer than 0.1% of the parameters (see Tab.10 for number of trainable parameters). TimeMatch shows the opposite profile: the lowest SSI (0.388/0.366), a higher-rank update than MADOS (ERA 0.612/0.687), and by far the largest weight change (ρ = 0.882/0.901), indicating that the pretrained basis is substantially rewritten. Accordingly, LoRA and SVF fall 4.2 and 18.0 points below full fine-tuning on average, and

Dataset	Backbone	Full-FT spectral diagnostics						mIoU (%) ↑		
		SSI_{32}		ERA		ρ		Full	LoRA	SVF
		Attn	MLP	Attn	MLP	Attn	MLP			
MADOS	CROMA	.545	.695	.466	.542	.452	.462	63.6	63.9	**64.4**
	DOFA	.673	.762	.513	.674	.345	.332	58.5	55.9	**58.9**
	Prithvi-2	.786	.787	.558	.718	.292	.261	50.9	**62.1**	61.0
	TerraMind	.890	.909	.459	.611	.195	.187	**65.2**	63.9	62.3
	SSL4EO-MAE	.930	.923	.324	.437	.136	.141	46.1	**46.9**	41.9
	Mean	.765	.815	.464	.596	.284	.277	56.9	**58.5**	57.7
PASTIS	CROMA	.239	.276	.665	.726	.905	.936	**37.2**	31.6	36.8
	DOFA	.684	.764	.584	.723	.372	.370	**26.2**	8.9	13.5
	Prithvi-2	.460	.378	.745	.874	.602	.604	33.7	**37.0**	33.9
	TerraMind	.919	.895	.612	.759	.202	.240	**43.2**	41.6	38.9
	SSL4EO-MAE	.419	.360	.764	.859	.653	.785	**34.2**	33.0	30.2
	Mean	.544	.535	.674	.788	.547	.587	**34.9**	30.4	30.7
FLAIR-2	CROMA	.348	.446	.525	.616	.606	.694	33.2	**34.0**	31.8
	DOFA	.450	.516	.657	.775	.596	.641	39.7	**45.5**	44.6
	Prithvi-2	.551	.420	.684	.803	.500	.522	37.4	**37.6**	37.1
	TerraMind	.566	.513	.693	.807	.491	.533	38.9	**41.7**	36.9
	SSL4EO-MAE	.522	.448	.699	.783	.518	.646	**32.8**	28.7	21.6
	Mean	.487	.469	.652	.757	.542	.607	36.4	**37.5**	34.4
Overall mean		.599	.606	.597	.714	.458	.490	**42.7**	42.2	40.9

Table 2: Full-FT spectral diagnostics and test mIoU on three EO benchmarks. Attn and MLP denote layer groups, and ρ denotes relative weight change. Dataset means average the five backbones; the overall mean weights all 15 model–dataset pairs equally. Bold performance values identify the best adaptation method within each backbone, and dataset mean.

Dataset	Backbone	Full-FT spectral diagnostics						mIoU (%) ↑		
		SSI_{32}		ERA		ρ		Full	LoRA	SVF
		Attn	MLP	Attn	MLP	Attn	MLP			
BraDD	CROMA	.287	.357	.672	.743	.914	1.090	54.1	55.8	**57.4**
	DOFA	.637	.715	.609	.737	.422	.427	**48.1**	47.3	47.7
	TerraMind	.449	.457	.721	.838	.638	.645	59.1	58.7	**60.3**
	SSL4EO-MAE	.745	.694	.685	.796	.349	.400	52.3	51.8	**53.7**
	Mean	.530	.556	.672	.779	.581	.641	53.4	53.4	**54.8**

Table 3: Full-FT spectral diagnostics and test mIoU on BraDD. Attn and MLP denote layer groups; ρ is relative weight change. The mean row weights all four backbones equally. Bold marks the best performance per row.

SVF underperforms full fine-tuning on all seven backbones. The remaining benchmarks lie between these extremes, with SVF slightly outperforming full fine-tuning on BraDD (54.8 vs. 53.4, with LoRA on par) and showing moderate drops on PASTIS and FLAIR-2.

Vision models illustrate the reuse regime. With strong pretraining, full fine-tuning largely preserves vision foundation models' dominant subspaces. On VOC, DINOv3 has SSI of 0.998/0.999 (attention/MLP) and $\rho \leq 0.021$, indicating near-complete preservation. Its ERA values of 0.048/0.065 are approximately an order of magnitude below the corresponding EO means. LoRA and SVF reach 78.3 and 81.4 mIoU, respectively, versus 77.5 for full fine-tuning. High SSI and small relative weight changes persist for DINOv3 even with $10\times$ more optimization steps and a larger initial learning rate (Appendix B.4).

Dataset	Backbone	Full-FT spectral diagnostics						Accuracy (%) ↑		
		SSI_{32}		ERA		ρ		Full	LoRA	SVF
		Attn	MLP	Attn	MLP	Attn	MLP			
TimeMatch	CROMA	.116	.074	.589	.628	1.330	1.445	**85.6**	85.4	73.7
	DOFA	.188	.182	.587	.701	.851	.883	**86.2**	81.5	70.0
	Prithvi-2	.327	.175	.597	.707	.654	.726	**86.6**	84.7	72.5
	SSL4EO-MAE	.437	.310	.696	.712	.639	.834	84.3	**85.5**	73.0
	AnySat	.217	.224	.639	.700	.845	.952	**86.9**	69.6	62.7
	Presto	.715	.759	.416	.519	1.429	1.058	**70.0**	66.0	36.7
	Tessera	.716	.836	.760	.840	.424	.406	**87.9**	84.9	72.6
	Mean	.388	.366	.612	.687	.882	.901	**83.9**	79.7	65.9

Table 4: Full-FT spectral diagnostics and test accuracy on the canonical single-target TimeMatch experiment, with FR/31TCJ/2017 held out. Attn and MLP denote layer groups; ρ is relative weight change. The mean row weights all seven backbones equally. Bold marks best performance per row.

Dataset	Backbone	Full-FT spectral diagnostics						Performance (%) ↑		
		SSI_{32}		ERA		ρ		Full	LoRA	SVF
		Attn	MLP	Attn	MLP	Attn	MLP			
Flowers-102	CLIP	.982	.992	.160	.246	.051	.051	91.2	90.7	**93.7**
	DINOv2	.994	.990	.160	.286	.045	.046	97.0	**98.7**	98.1
	DINOv3	.998	.999	.069	.102	.021	.020	**98.1**	97.4	94.7
CIFAR-10	CLIP	.976	.984	.237	.337	.065	.061	**98.1**	97.2	97.6
	DINOv2	.986	.987	.248	.371	.057	.057	98.8	98.5	**99.0**
	DINOv3	.998	.999	.048	.055	.019	.017	98.7	98.7	**99.0**
Pets	CLIP	.944	.969	.310	.493	.105	.099	91.3	**93.9**	87.9
	DINOv2	.966	.972	.337	.584	.101	.100	93.1	**95.2**	**95.2**
	DINOv3	.996	.997	.136	.218	.034	.033	94.8	94.4	**95.4**
VOC	CLIP	.995	.998	.102	.144	.027	.030	71.7	70.3	**72.9**
	DINOv2	.998	.998	.090	.143	.024	.026	**79.6**	77.8	75.9
	DINOv3	.998	.999	.048	.065	.021	.020	77.5	78.3	**81.4**
Overall mean		.986	.990	.162	.254	.048	.047	–	–	–

Table 5: Full-FT spectral diagnostics and downstream accuracy (mIoU for VOC). Attn and MLP denote layer groups; ρ is relative weight change. Spectral means weight datasets and backbones equally. Bold marks the best performance per row.

Stronger pretraining is associated with more concentrated updates. Comparing DINOv2 with its stronger successor DINOv3 illustrates the effect of improved pretraining. DINOv3 consistently exhibits a clearer reuse signature, with higher SSI and lower ρ and ERA. On CIFAR-10, for instance, SSI rises from 0.986/0.987 to 0.998/0.999, ρ drops from 0.057/0.057 to 0.019/0.017, and ERA drops from 0.248/0.371 to 0.048/0.055, so the update is concentrated in several times fewer directions. This suggests that stronger pretraining yields a richer and more reusable basis, within which a downstream task can be solved by adjusting a few existing directions rather than introducing new ones. This agrees with Turkoglu et al. (2026), who show that SVF benefits increasingly from stronger pretraining, often exceeding full fine-tuning, also in the language domain.

4 Discussion and Future Perspective

Reuse and relearning in EO and CV. The evaluated EO and CV models generally occupy different adaptation regimes. CV models largely preserve pretrained subspaces, whereas EO models undergo larger updates with higher effective ranks and greater subspace reorganization. EO pretrain-

ing thus often appears to provide a useful initialization for task-specific learning, with less readily reusable structure. The smaller pretraining gains on MADOS and PASTIS compared with VOC support this interpretation (Fig. 3c), although benchmarks and training budgets differ.

A trade-off between performance and reuse. For most EO backbones, higher performance on MADOS and PASTIS is associated with larger relative weight changes (Fig. 3b). Together with their lower SSI, this suggests that these models often solve downstream tasks by substantially departing from their dominant pretrained subspaces. Such behavior may fall short of the foundation model premise of learning representations that can be readily reused across tasks. However, models achieving strong performance with limited changes show that this trade-off is not inevitable, highlighting the potential for more reusable EO pretraining.

Subspace preservation can coexist with broad adaptation. High SSI does not necessarily imply low-dimensional adaptation. On TimeMatch, Tessera (Feng et al., 2026) achieves the highest accuracy while combining relatively high SSI with high ERA (Tab. 4). This illustrates that successful adaptation can preserve much of the dominant pretrained subspaces while distributing updates across many directions. Subspace preservation and adaptation breadth therefore capture complementary aspects of transfer and should be interpreted jointly.

Cross-modal transfer and adaptation depth. Subspace changes generally concentrate in early and middle layers, with larger preservation toward the output (Fig. 3a). CROMA's cross-modal experiments illustrate why performance and reuse must be distinguished (Appendix C). On BraDD, the optical encoder matches the SAR encoder at 54.1 mIoU while substantially reorganizing its early and middle layers. Conversely, the SAR encoder preserves more pretrained structure on MADOS but reaches only 40.7 mIoU, compared with 63.6 for the optical encoder. Strong performance can therefore coexist with extensive rewriting, while high subspace preservation can accompany poor transfer. Neither accuracy nor SSI alone establishes successful reuse.

Limitation. Our main analyses use fixed fine-tuning configurations, complemented by targeted learning-rate sweeps for one CV and two EO encoders (Appendix B.3). These sweeps support our main conclusions but also show that a shared learning rate can under-train individual encoders. Extending them to all configurations was beyond our computational budget, as each configuration would require $10\times$ the fine-tuning compute. Future work should broaden these sweeps and explore aggregate measures, such as normalized area-under-the-curve scores over a shared log-scaled learning-rate range, to assess adaptation more broadly than a single configuration.

Future work. Our diagnostics could guide the design of pretraining objectives that produce more broadly reusable representations, supporting strong downstream performance with less structural reorganization. Controlled studies varying pretraining data size and diversity, architecture, model capacity, and compute budget could further disentangle their contributions to reuse and investigate whether representation reuse follows predictable scaling relationships.

5 Related Work

Linear subspaces as feature bases. Recent work suggests that knowledge in neural networks is organized along meaningful linear subspaces. In transformer weight matrices, singular vectors have been associated with semantically meaningful directions in representation space (Elhage et al., 2022; Millidge & Black, 2022), while analyses of their spectra suggest that information is distributed across singular modes (Staats et al., 2025). Related work on model editing further shows that targeted low-rank changes to weight matrices can modify specific learned associations (Meng et al., 2022). Fine-tuning itself has also been shown to operate in a low-dimensional intrinsic subspace (Aghajanyan et al., 2021). Methods such as Singular Value Fine-tuning build on this perspective by preserving pretrained singular directions while adapting their associated singular values (Sun et al., 2022; Turkoglu et al., 2026). Related analyses of transfer learning have examined how pretrained models change during adaptation in feature and parameter space, including the spectral structure of individual modules (Neyshabur et al., 2020). Together, these findings motivate treating the dominant singular subspace of a pretrained model as a meaningful feature basis whose preservation can be studied during downstream adaptation.

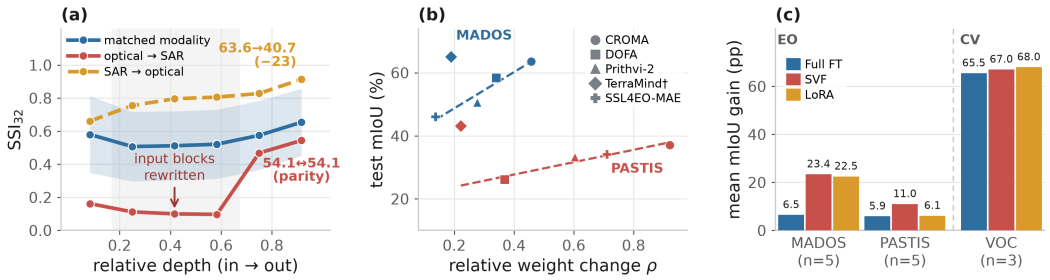


Figure 3: (a) Depth-resolved SSI_{32} under full-FT for matched-modality transfer (mean ± 1 std across encoders and benchmarks) and CROMA cross-modal transfer on BraDD. Annotations compare cross-modal and modality-matched mIoU. (b) Full-FT mIoU versus relative weight change ρ on MADOS and PASTIS. Dashed lines are least-squares fits excluding TerraMind. (c) Test mIoU gains over architecture-matched random initialization, in pp, for full-FT, SVF, and LoRA: five EO backbones on MADOS/PASTIS and three CV backbones on VOC, under the evaluated training budgets.

Spectral structure of model adaptation. Several parameter-efficient methods exploit low-rank or spectral structure during adaptation. LoRA constrains weight updates to be low rank (Hu et al., 2022), while Spectral Adapter and related methods parameterize adaptation directly in the singular basis (Zhang & Pilanci, 2024; Zhao et al., 2025). Recent work has also used spectral structure to analyze fine-tuning itself. Si et al. (2025) study changes in singular values and singular-vector alignment between pretrained and fine-tuned weights. Similarly, Shuttleworth et al. (2025) show that LoRA and full fine-tuning can reach similar downstream performance while inducing substantially different singular-vector structure, including LoRA-specific "intruder" directions. Relatedly, Diffract studies the spectral evolution of LLMs during continued pretraining and finds that substantial adaptation can occur through changes in singular directions despite comparatively stable singular-value spectra (Borodin et al., 2026). Our objective is complementary: rather than constraining adaptation or tracking individual singular directions, we use spectral structure to *audit representation reuse*. SSI measures preservation of the dominant pretrained left and right singular subspaces through principal angles, while ERA characterizes the effective dimensionality of the update ΔW.

Foundation models for EO and Earth system. Foundation models aim to learn representations that can be reused across downstream tasks with limited adaptation (Bommasani et al., 2021). This paradigm, established in language and general vision (Brown et al., 2020; Oquab et al., 2024), has increasingly been adopted in Earth observation (EO). SatMAE targets temporal and multispectral imagery (Cong et al., 2022), CROMA jointly learns from optical and SAR observations (Fuller et al., 2023), and Prithvi-EO-2.0 targets transfer across multitemporal EO applications (Szwarcman et al., 2025). At the same time, benchmark studies show that geospatial foundation models do not consistently outperform conventional or task-specific alternatives (Marsocci et al., 2026; Corley et al., 2026). Related developments in weather and Earth-system modelling range from data-driven forecasters such as Pangu-Weather and GraphCast (Bi et al., 2023; Lam et al., 2023) to broadly adaptable models such as Aurora (Bodnar et al., 2025). These developments motivate evaluating not only downstream accuracy, but also whether adaptation actually reuses the pretrained representation.

6 Conclusion

We introduced spectral diagnostics to examine representation reuse through subspace preservation, update magnitude, and adaptation breadth. Across the evaluated benchmarks, EO models generally undergo larger, higher-rank updates and greater subspace reorganization than natural-image models, showing that strong downstream performance can coexist with extensive rewriting of pretrained structure. These patterns also help explain when parameter-efficient methods can match full fine-tuning. Our findings motivate evaluating foundation models through downstream performance, pretraining benefits, and structural preservation together, since neither accuracy nor preservation alone establishes successful pretraining representation reuse. This perspective can guide EO pretraining toward representations that support strong transfer with less relearning.

AI USE STATEMENT

We used generative AI tools as writing aids to improve the language, clarity, grammar, and readability of the manuscript. Generative AI tools were *not* used for substantive research tasks, including developing the methodology, formulating hypotheses or mathematical claims, conducting experiments, analyzing data, or interpreting results. They were also used to assist in preparing the public code release. All AI-assisted text and code were reviewed by the authors to ensure correctness and consistency with the intended content. The authors take full responsibility for the final manuscript and released code.

ETHICS STATEMENT

This work studies existing Earth-observation foundation models and does not involve human subjects or real-world deployment. We note that EO models can have dual-use applications and may reflect geographic or dataset biases.

REPRODUCIBILITY STATEMENT

We provide the definitions and implementation details of all spectral diagnostics in Section 2 and Appendix A. The experimental setup, including dataset splits, preprocessing, model configurations, adaptation methods, trainable parameter counts, optimization settings, and random seeds, is documented in Section 3 and Appendix F. We also provide the code required to reproduce experiments.

REFERENCES

Armen Aghajanyan, Sonal Gupta, and Luke Zettlemoyer. Intrinsic dimensionality explains the effectiveness of language model fine-tuning. In *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)*, pp. 7319–7328. Association for Computational Linguistics, 2021. doi: 10.18653/v1/2021.acl-long.568.

Guillaume Astruc, Nicolas Gonthier, Clément Mallet, and Loic Landrieu. AnySat: One Earth observation model for many resolutions, scales, and modalities. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19530–19540, 2025.

Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, and Qi Tian. Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, 619:533–538, 2023. doi: 10.1038/s41586-023-06185-3.

Åke Björck and Gene H. Golub. Numerical methods for computing angles between linear subspaces. *Mathematics of Computation*, 27(123):579–594, 1973.

Cristian Bodnar, Wessel P. Bruinsma, Ana Lucic, et al. A foundation model for the Earth system. *Nature*, 641:1180–1187, 2025. doi: 10.1038/s41586-025-09005-y.

Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, et al. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*, 2021.

Derrick Bonafilia, Beth Tellman, Tyler Anderson, and Erica Issenberg. Sen1Floods11: A georeferenced dataset to train and test deep learning flood algorithms for Sentinel-1. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 835–845, 2020. doi: 10.1109/CVPRW50498.2020.00113.

Nikita Borodin, Maria Krylova, Artem Zabolotnyi, Dmitry Aspisov, Egor Shikov, Nikita Tyuplyaev, Oleg Travkin, Roman Alferov, and Dmitry Vinichenko. Diffract: Spectral view of LLM domain adaptation. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=XBUHoiAGDE.

Tom B. Brown, Benjamin Mann, Nick Ryder, et al. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Yezhen Cong, Samar Khanna, Chenlin Meng, Patrick Liu, Erik Rozi, Yutong He, Marshall Burke, David B. Lobell, and Stefano Ermon. SatMAE: Pre-training transformers for temporal and multi-spectral satellite imagery. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Isaac Corley, Nils Lehmann, Caleb Robinson, Gabriel Tseng, Anthony Fuller, Hamed Alemohammad, Evan Shelhamer, Jennifer Marcus, and Hannah Kerner. No one knows the state of the art in geospatial foundation models. *arXiv preprint arXiv:2605.12678*, 2026.

Nelson Elhage, Tristan Hume, Catherine Olsson, Nicholas Schiefer, Tom Henighan, Shauna Kravec, Zac Hatfield-Dodds, Robert Lasenby, Dawn Drain, Carol Chen, Roger Grosse, Sam McCandlish, Jared Kaplan, Dario Amodei, Martin Wattenberg, and Christopher Olah. Toy models of superposition. *Transformer Circuits Thread*, 2022. URL https://transformer-circuits.pub/2022/toy_model/.

Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John Winn, and Andrew Zisserman. The Pascal visual object classes (VOC) challenge. *International Journal of Computer Vision*, 88 (2):303–338, 2010. doi: 10.1007/s11263-009-0275-4.

Zhengpeng Feng, Clement Atzberger, Sadiq Jaffer, Jovana Knezevic, Silja Sormunen, Robin Young, Madeline C. Lisaius, Markus Immitzer, Toby Jackson, James Ball, David A. Coomes, Anil Madhavapeddy, Andrew Blake, and Srinivasan Keshav. TESSERA: Temporal embeddings of surface spectra for Earth representation and analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 34818–34831, 2026.

Anthony Fuller, Koreen Millard, and James R. Green. CROMA: Remote sensing representations with contrastive radar-optical masked autoencoders. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Anatol Garioud, Apolline De Wit, Marc Poupée, Marion Valette, Sébastien Giordano, and Boris Wattrelos. FLAIR #2: Textural and temporal information for semantic segmentation from multi-source optical imagery. *arXiv preprint arXiv:2305.14467*, 2023.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022.

Johannes Jakubik, Felix Yang, Benedikt Blumenstiel, Erik Scheurer, Rocco Sedona, Stefano Maurogiovanni, Jente Bosmans, Nikolaos Dionelis, Valerio Marsocci, Niklas Kopp, Rahul Ramachandran, Paolo Fraccaro, Thomas Brunschwiler, Gabriele Cavallaro, Juan Bernabe-Moreno, and Nicolas Longépé. TerraMind: Large-scale generative multimodality for Earth observation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 7383–7394, 2025. doi: 10.1109/ICCV51701.2025.00693.

Kaan Karaman, Vivien Sainte Fare Garnot, and Jan Dirk Wegner. Deforestation detection in the Amazon with Sentinel-1 SAR image time series. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, X-1/W1-2023:835–842, 2023. doi: 10.5194/isprs-annals-X-1-W1-2023-835-2023.

Katerina Kikaki, Ioannis Kakogeorgiou, Ibrahim Hoteit, and Konstantinos Karantzalos. Detecting marine pollutants and sea surface features with deep learning in Sentinel-2 imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 210:39–54, 2024. doi: 10.1016/j.isprsjprs.2024.02.017.

Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009. URL https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf.

Remi Lam, Alvaro Sanchez-Gonzalez, Matthew Willson, Peter Wirnsberger, Meire Fortunato, Ferran Alet, Suman Ravuri, Timo Ewalds, Zach Eaton-Rosen, Weihua Hu, Alexander Merose, Stephan Hoyer, George Holland, Oriol Vinyals, Jacklynn Stott, Alexander Pritzel, Shakir Mohamed, and Peter Battaglia. Learning skillful medium-range global weather forecasting. *Science*, 382(6677):1416–1421, 2023. doi: 10.1126/science.adi2336.

Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A ConvNet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11976–11986, 2022. doi: 10.1109/CVPR52688.2022.01167.

Valerio Marsocci, Yuru Jia, Georges Le Bellier, David Kerekes, Liang Zeng, Sebastian Hafner, Sebastian Gerard, Eric Brune, Ritu Yadav, Ali Shibli, Heng Fang, Yifang Ban, Maarten Vergauwen, Nicolas Audebert, and Andrea Nascetti. PANGAEA: Assessing geospatial foundation models capabilities through a global and inclusive benchmark. *IEEE Geoscience and Remote Sensing Magazine*, 14(1):245–285, 2026. doi: 10.1109/MGRS.2025.3628194.

Kevin Meng, David Bau, Alex Andonian, and Yonatan Belinkov. Locating and editing factual associations in GPT. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Beren Millidge and Sid Black. The singular value decompositions of transformer weight matrices are highly interpretable. AI Alignment Forum, 2022. URL https://www.alignmentforum.org/posts/mkbGjzxD8d8XqKHzA/the-singular-value-decompositions-of-transformer-weight.

Vishal Nedungadi, Ankit Kariryaa, Stefan Oehmcke, Serge Belongie, Christian Igel, and Nico Lang. MMEarth: Exploring multi-modal pretext tasks for geospatial representation learning. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 164–182, 2024. doi: 10.1007/978-3-031-73039-9_10.

Behnam Neyshabur, Hanie Sedghi, and Chiyuan Zhang. What is being transferred in transfer learning? In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Maria-Elena Nilsback and Andrew Zisserman. Automated flower classification over a large number of classes. In *2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing*, pp. 722–729. IEEE, 2008. doi: 10.1109/ICVGIP.2008.47.

Joachim Nyborg, Charlotte Pelletier, Sébastien Lefèvre, and Ira Assent. TimeMatch: Unsupervised cross-region adaptation by temporal shift estimation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 188:301–313, 2022. doi: 10.1016/j.isprsjprs.2022.04.018.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, et al. DINOv2: Learning robust visual features without supervision. *Transactions on Machine Learning Research*, 2024.

Omkar M. Parkhi, Andrea Vedaldi, Andrew Zisserman, and C. V. Jawahar. Cats and dogs. In *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3498–3505. IEEE, 2012. doi: 10.1109/CVPR.2012.6248092.

Christopher Phillips, Sujit Roy, Kumar Ankur, and Rahul Ramachandran. HLS foundation burnscars dataset. Hugging Face dataset, 2023. URL https://huggingface.co/datasets/ibm-nasa-geospatial/hls_burn_scars.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8748–8763. PMLR, 2021. URL https://proceedings.mlr.press/v139/radford21a.html.

Olivier Roy and Martin Vetterli. The effective rank: A measure of effective dimensionality. In *2007 15th European Signal Processing Conference*, pp. 606–610, 2007.

Vivien Sainte Fare Garnot and Loic Landrieu. Lightweight temporal self-attention for classifying satellite images time series. In *Advanced Analytics and Learning on Temporal Data (AALTD 2020)*, volume 12588 of *Lecture Notes in Computer Science*, pp. 171–181. Springer, 2020. doi: 10.1007/978-3-030-65742-0_12.

Vivien Sainte Fare Garnot and Loic Landrieu. Panoptic segmentation of satellite image time series with convolutional temporal attention networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4852–4861, 2021. doi: 10.1109/ICCV48922.2021.00483.

Vivien Sainte Fare Garnot, Loic Landrieu, Sebastien Giordano, and Nesrine Chehata. Satellite image time series classification with pixel-set encoders and temporal self-attention. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12322–12331, 2020.

Reece Shuttleworth, Jacob Andreas, Antonio Torralba, and Pratyusha Sharma. LoRA vs full fine-tuning: An illusion of equivalence. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Chongjie Si, Xuankun Yang, Muqing Liu, Yadao Wang, Xiaokang Yang, Wenbo Su, Bo Zheng, and Wei Shen. Weight spectra induced efficient model adaptation. *arXiv preprint arXiv:2505.23099*, 2025.

Oriane Siméoni, Huy V. Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seungeun Yi, Michaël Ramamonjisoa, Francisco Massa, Daniel Haziza, Luca Wehrstedt, Jianyuan Wang, Timothée Darcet, Théo Moutakanni, Leonel Sentana, Claire Roberts, Andrea Vedaldi, Jamie Tolan, John Brandt, Camille Couprie, Julien Mairal, Hervé Jégou, Patrick Labatut, and Piotr Bojanowski. DINOv3. *arXiv preprint arXiv:2508.10104*, 2025.

Max Staats, Matthias Thamm, and Bernd Rosenow. Small singular values matter: A random matrix analysis of transformer models. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Yanpeng Sun, Qiang Chen, Xiangyu He, Jian Wang, Haocheng Feng, Junyu Han, Errui Ding, Jian Cheng, Zechao Li, and Jingdong Wang. Singular value fine-tuning: Few-shot segmentation requires few-parameters fine-tuning. In *Advances in Neural Information Processing Systems*, volume 35, 2022. doi: 10.52202/068431-2717.

Daniela Szwarcman, Sujit Roy, Paolo Fraccaro, et al. Prithvi-EO-2.0: A versatile multitemporal foundation model for Earth observation applications. *IEEE Transactions on Geoscience and Remote Sensing*, 64:1–20, 2025. doi: 10.1109/TGRS.2025.3642610.

Gabriel Tseng, Ruben Cartuyvels, Ivan Zvonkov, Mirali Purohit, David Rolnick, and Hannah Kerner. Lightweight, pre-trained transformers for remote sensing timeseries. *arXiv preprint arXiv:2304.14065*, 2023.

Mehmet Ozgur Turkoglu, Dominik J. Mühlematter, Alexander Becker, Konrad Schindler, and Helge Aasen. Quantifying the uncertainty of foundation models with singular value ensembles. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026.

Yi Wang, Nassim Ait Ali Braham, Zhitong Xiong, Chenying Liu, Conrad M. Albrecht, and Xiao Xiang Zhu. SSL4EO-S12: A large-scale multimodal, multitemporal dataset for self-supervised learning in Earth observation. *IEEE Geoscience and Remote Sensing Magazine*, 11(3):98–106, 2023. doi: 10.1109/MGRS.2023.3281651.

Tete Xiao, Yingcheng Liu, Bolei Zhou, Yuning Jiang, and Jian Sun. Unified perceptual parsing for scene understanding. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 432–448, 2018.

Zhitong Xiong, Yi Wang, Fahong Zhang, Adam J. Stewart, Joëlle Hanna, Damian Borth, Ioannis Papoutsis, Bertrand Le Saux, Gustau Camps-Valls, and Xiao Xiang Zhu. Neural plasticity-inspired multimodal foundation model for Earth observation. *arXiv preprint arXiv:2403.15356*, 2024.

Fangzhao Zhang and Mert Pilanci. Spectral adapter: Fine-tuning in spectral space. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-4158.

Jialin Zhao, Yingtao Zhang, Xinghang Li, Huaping Liu, and Carlo Vittorio Cannistraci. Sparse spectral training and inference on Euclidean and hyperbolic neural networks. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 77976–78002. PMLR, 2025.

A Additional Details on Spectral Diagnostics

A.1 Matrix Extraction and Aggregation

We compute the spectral diagnostics independently for each analyzed weight matrix in the encoder. Attention query, key, value, and output projections are grouped as ATTN, while feed-forward layers are grouped as MLP. Task-specific heads and normalization parameters are excluded because they do not have a directly comparable pretrained counterpart. For convolutional encoders, convolutional kernels of shape $[C_{\text{out}}, C_{\text{in}}, k_h, k_w]$ are reshaped to $[C_{\text{out}}, C_{\text{in}} k_h k_w]$ before computing the diagnostics. Diagnostics are computed per matrix and subsequently aggregated within the corresponding layer group.

A.2 Computing SSI Principal Angles

Let $U_{\text{pre},k}, U_{\text{ft},k} \in \mathbb{R}^{m\times k}$ contain the top-k left singular vectors of W_{pre} and W_{ft}, and let $V_{\text{pre},k}, V_{\text{ft},k} \in \mathbb{R}^{n\times k}$ contain the corresponding right singular vectors. Their columns are orthonormal and therefore span k-dimensional subspaces. The principal angles $0 \le \theta_1 \le \cdots \le \theta_k \le \pi/2$ between two such subspaces are obtained from the singular values of the cross-Gram matrix between their orthonormal bases (Björck & Golub, 1973):

$$\cos\theta_i^{(L)} = \sigma_i\big(U_{\text{pre},k}^\top U_{\text{ft},k}\big), \qquad \cos\theta_i^{(R)} = \sigma_i\big(V_{\text{pre},k}^\top V_{\text{ft},k}\big). \tag{7}$$

Each $\cos\theta_i \in [0,1]$ measures the overlap of a principal direction between the two subspaces. The principal angles depend only on the subspaces and are invariant to the particular orthonormal bases used to represent them.

A.3 Choice of SSI Subspace Rank k

Why truncate the singular subspaces? SSI measures whether fine-tuning preserves the dominant pretrained subspaces. Truncation is essential because a complete basis spans the same ambient space regardless of how its vectors rotate. At $k = \min(m,n)$, at least one subspace overlap is therefore identically one; for square matrices, $\text{SSI}_k = 1$ regardless of the update. Restricting the comparison to the leading k singular directions avoids this degeneracy and focuses the diagnostic on the strongest linear transformations represented by each matrix. Related spectral analyses also use truncation: Diffract (Borodin et al., 2026), for example, assesses adaptation redundancy by retaining only the leading singular components of the weight update ΔW and evaluating the resulting model.

Choice of k. We use $k = 32$ across architectures to compare the same number of dominant directions while remaining below full rank for every analyzed matrix. Larger k progressively includes weaker singular modes and broadens the comparison beyond the dominant subspaces. We report $k \in \{64, 128\}$ as sensitivity checks.

B Additional Results and Ablations

B.1 Aurora Results

Aurora provides a useful within-model comparison because the same pretrained backbone is adapted to several operational forecasting settings (Bodnar et al., 2025). These include weather forecasting at the pretraining resolution of $0.25°$, higher-resolution weather forecasting at $0.1°$, ocean-wave prediction at $0.25°$, and air-pollution forecasting at $0.4°$. The resulting spectral changes are consistent with the increasing departure from the original forecasting setting. Adaptation at $0.25°$ leaves the pretrained subspaces almost unchanged, with SSI above 0.99 and small ERA and relative weight change. Moving to $0.1°$ weather forecasting produces substantially lower SSI and higher ERA and ρ, indicating that higher spatial resolution requires a broader reorganization of the pretrained weights. Wave and air-pollution forecasting similarly induce larger, higher-rank updates, with the strongest changes observed for air pollution. These results provide an additional within-architecture sanity check that SSI and ERA respond meaningfully to changes in downstream task and resolution.

				SSI		ERA		Rel. Change	
Downstream Model	**Base Model**	**Res.**	**Task**	**Attn**	**MLP**	**Attn**	**MLP**	**Attn**	**MLP**
Aurora 0.25° Fine-Tuned	Aurora 0.25° Pretrained	0.25°	HRES T0 Forecast	0.992	0.997	0.150	0.141	0.055	0.060
Aurora 0.1° Fine-Tuned	Aurora 0.25° Pretrained	0.1°	High-Res Forecast	0.902	0.953	0.528	0.762	0.218	0.235
Aurora 0.25° Wave	Aurora 0.25° Pretrained	0.25°	Wave Prediction	0.918	0.945	0.563	0.779	0.220	0.277
Aurora 0.4° Air Pollution	Aurora 0.25° 12h Pretrained	0.4°	Air Pollution	0.865	0.917	0.683	0.864	0.319	0.356

Table 6: Spectral diagnostics for Aurora (1.3B) across downstream Earth-system tasks.

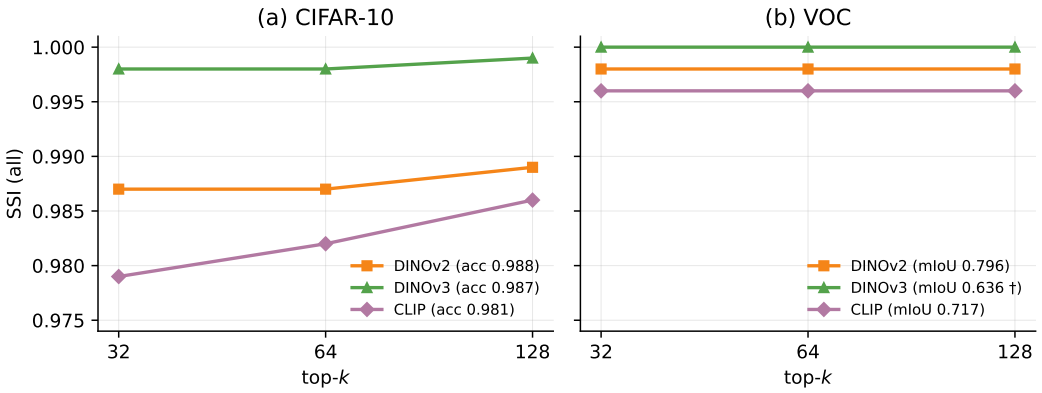


Figure 4: Sensitivity of SSI to the subspace rank k on (a) CIFAR-10 and (b) VOC. SSI remains largely stable across the evaluated ranks, showing that high subspace preservation is robust to the choice of k for the runs shown. Legends report accuracy and mIoU, respectively.

B.2 Sensitivity to the SSI Subspace Rank

We vary $k \in \{32, 64, 128\}$ to assess the stability of SSI and its interpretation. For CV backbones (Fig. 4), SSI remains close to one across ranks, with little change in absolute values and an unchanged model ordering. Among EO encoders (Fig. 5), SSI is comparatively stable on PASTIS, particularly for models with high preservation. On BraDD, scores increase more noticeably, but all models follow similar upward trends. Although individual EO rankings can vary, the broad preservation patterns remain consistent. The qualitative conclusions are therefore robust across the evaluated choices of k.

Increasing k includes progressively weaker singular modes and expands the compared subspaces, which can raise their overlap and compress differences between models. We therefore choose $k = 32$ as a practical trade-off: it summarizes dominant weight structure across multiple singular directions while retaining sensitivity to differences in subspace preservation. The larger-rank evaluations show that the overall interpretation extends beyond this particular choice.

B.3 Learning Rate Sensitivity and Under-Training

A high SSI can reflect either useful preservation or insufficient adaptation. We examine this distinction by sweeping the full fine-tuning learning rate for DINOv3 on Flowers-102 and VOC (Fig. 6) and for AnySat (Astruc et al., 2025) and Presto (Tseng et al., 2023) on TimeMatch (Fig. 7). For DINOv3, the spectral diagnostics remain largely stable across the sweep: over almost two orders of magnitude in learning rate, SSI stays above 0.93 and ρ at most 0.15, while ERA remains low up to the performance optimum. Very low learning rates yield high SSI and small weight changes despite suboptimal performance. Increasing the learning rate improves performance until it reaches a plateau, over which SSI remains high and largely stable. Higher learning rates broaden the update, with higher ERA and poorer performance.

The EO encoders behave differently. Their spectral diagnostics shift substantially across the sweep, with SSI falling from about 0.9 at the lowest learning rate to below 0.25 within one to two orders of magnitude. Presto improves monotonically with the learning rate while SSI decreases and ρ increases, so its best accuracy coincides with the lowest SSI in the sweep and the default learning rate under-trains it. AnySat reaches its best accuracy at the default with low SSI. Lower learning

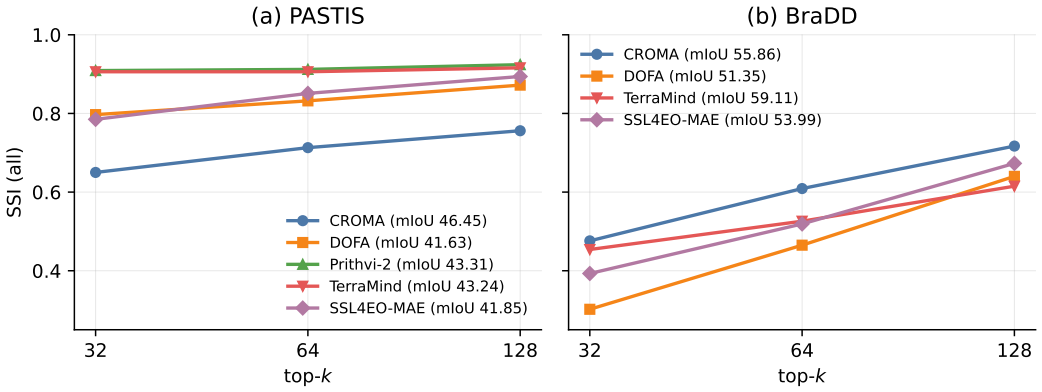
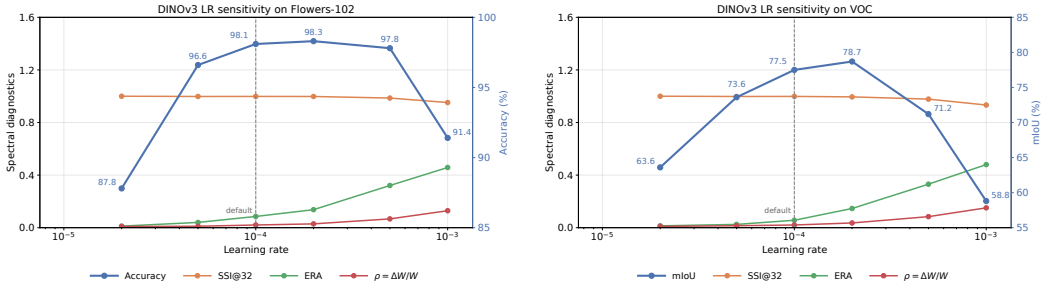


Figure 5: Sensitivity of SSI to rank k on (a) PASTIS and (b) BraDD. Legends report mIoU. Each curve varies k with trained weights fixed.



(a) Flowers-102 **(b)** VOC

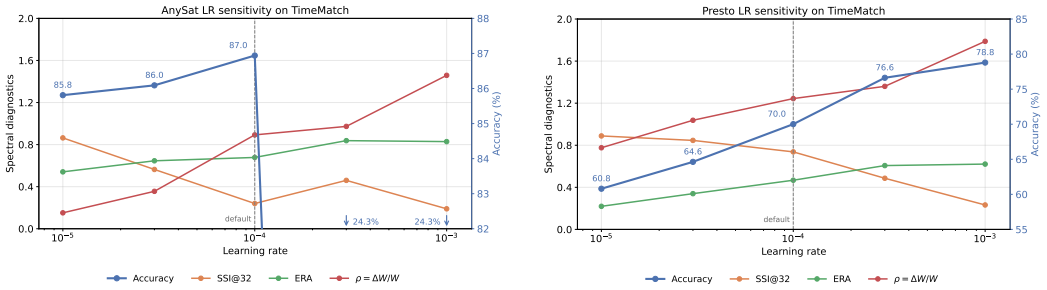
Figure 6: Learning-rate sensitivity of DINOv3 under full fine-tuning on (a) Flowers-102 and (b) VOC. Spectral diagnostics use the left axis and test performance the right axis; the dashed line marks the default learning rate. At low learning rates, the encoder changes very little and SSI remains close to one despite suboptimal performance. Performance peaks at 10^{-4} to 2×10^{-4}, where SSI_{32} remains above 0.99. Larger learning rates degrade performance and broaden the update (higher ERA), but relative weight change stays small ($\rho \leq 0.15$) and SSI_{32} stays above 0.93. Overall, subspace preservation is stable across the sweep, unlike for the EO encoders (Fig. 7).

rates raise SSI at a small cost in accuracy, whereas higher learning rates cause training to diverge. Neither encoder thus combines its best performance with preservation of its dominant subspaces, and the gap between CV and EO adaptation in Sec. 3.3 persists at each encoder's best learning rate.

These results motivate interpreting SSI, ERA, and ρ jointly with downstream performance. In particular, stable SSI across the performance plateau shows that substantial subspace preservation can accompany successful adaptation. High SSI with poor performance, however, may reflect insufficient learning, as observed for Presto at low learning rates. Diagnostics at a single learning rate therefore do not capture the full adaptation behavior of a model. Nevertheless, the sweeps support the conclusions drawn from the fixed configurations in our main analyses: the CV reference preserves its dominant subspaces across a wide range of learning rates, whereas the EO encoders reach their best performance only with substantial rewriting. We discuss this limitation in Section 4.

B.4 Sensitivity to the Fine-Tuning Budget

To assess whether subspace preservation persists under more extensive adaptation, we fine-tune DINOv3 on Flowers-102 and PASCAL VOC for 100 epochs, using $10\times$ more optimization steps than the default 10-epoch budget. We also double the initial learning rate from 10^{-4} to 2×10^{-4} to further test this behavior under larger optimization steps.



(a) AnySat **(b)** Presto

Figure 7: Learning rate sensitivity of (a) AnySat and (b) Presto under full fine-tuning on TimeMatch. Spectral diagnostics use the left axis and test accuracy the right axis; the dashed line marks the default learning rate used in Tab. 4. Unlike DINOv3 (Fig. 6), neither encoder reaches its best accuracy while preserving its dominant subspaces. AnySat peaks at the default with SSI_{32} of 0.24; lower learning rates raise SSI at a cost in accuracy, and learning rates of 3×10^{-4} and above diverge (24.3%, below the plotted range). Presto improves monotonically with the learning rate while SSI decreases and ρ increases, so its best accuracy coincides with the lowest SSI in the sweep. Overall, the EO diagnostics are far less stable across learning rates than their CV counterparts.

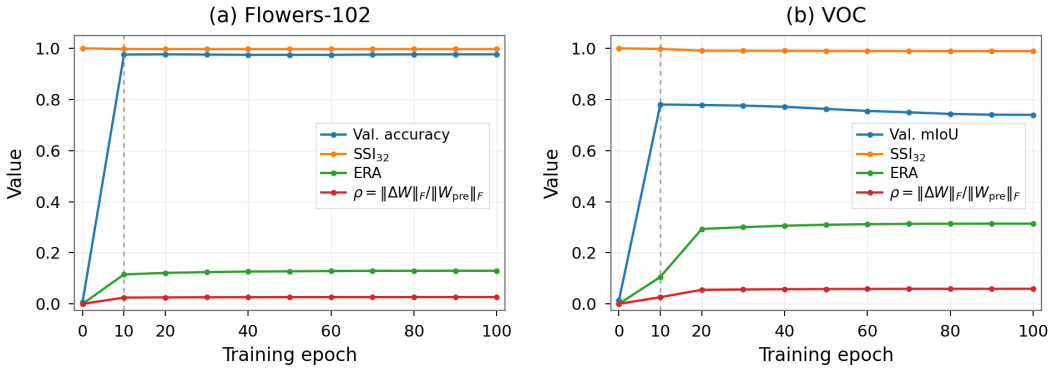


Figure 8: Extended full fine-tuning of DINOv3 on Flowers-102 and PASCAL VOC, using $10\times$ more optimization steps and twice the default initial learning rate. Curves show validation accuracy or mIoU as fractions, alongside SSI, ERA, and relative weight change ρ, aggregated over the analyzed encoder matrices. The dashed line marks epoch 10 within each extended run. SSI remains high and relative weight change remains small, while ERA increases more noticeably on VOC but still stays relatively low.

Figure 8 shows the evolution of validation performance and spectral diagnostics throughout these extended runs. Between epochs 10 and 100, Flowers-102 accuracy remains similar (97.55% to 97.65%), whereas VOC mIoU declines from 78.00% to 73.97%. SSI remains high on both datasets, ending at 0.9969 and 0.9893, respectively. Relative weight change increases only modestly in absolute terms and remains small ($\rho = 0.0266$ and 0.0589). ERA increases more noticeably on VOC, from 0.1054 to 0.3139, but remains below the EO benchmark averages reported in our main analysis. These results suggest that high subspace preservation and small relative updates persist in these two settings despite substantially longer training and a larger initial learning rate.

B.5 Effect of Decoder Capacity

We replace the linear decoder with UPerNet while keeping the encoder architecture and preprocessing unchanged (Fig. 9). UPerNet provides substantially greater task-specific capacity (Xiao et al., 2018): its more expressive decoder can accommodate more of the downstream learning, potentially

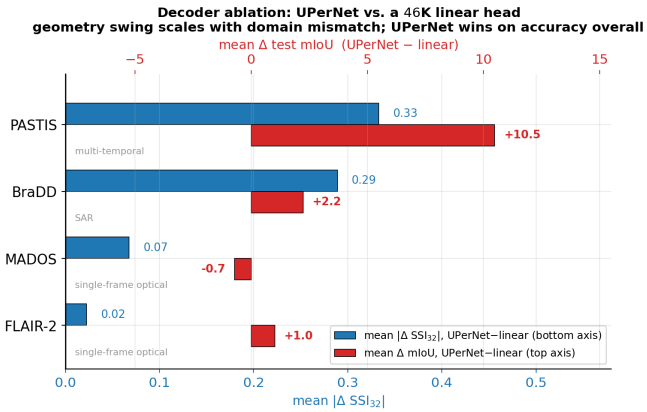


Figure 9: Decoder comparison across PASTIS, BraDD, MADOS, and FLAIR-2. Blue bars show the mean absolute change in SSI_{32}; red bars show the mean mIoU difference between UPerNet and the linear decoder.

reducing the need to reorganize the pretrained encoder. Greater preservation of encoder subspaces is therefore a plausible consequence of increasing decoder capacity.

The effect of decoder choice is small on single-frame optical tasks (mean $|\Delta\mathrm{SSI}_{32}| = 0.02$–$0.07$), but substantially larger on PASTIS and BraDD (0.29–0.33). UPerNet improves mean performance on PASTIS, BraDD, and FLAIR-2, with the largest gain on PASTIS, while slightly reducing mean performance on MADOS. These results show that decoder capacity influences both downstream performance and the extent of encoder adaptation.

B.6 Results on a Convolutional Encoder

To assess whether our observations extend beyond transformer encoders, we additionally evaluate MMEarth (Nedungadi et al., 2024), whose encoder is based on the ConvNeXt family of convolutional architectures (Liu et al., 2022) (Fig. 10). MMEarth exhibits the largest mean SVF deficit among the evaluated models, with SVF trailing full fine-tuning by 12.7 mIoU points. One possible contributing factor is the limited adaptation coverage of our SVF implementation, which modifies only 24 pointwise projection matrices in ConvNeXt.

Despite this performance gap, MMEarth retains high SSI_{32} (0.944–0.972) while also exhibiting high ERA (0.821–0.871). This combination shows that strong preservation of the dominant pretrained subspaces can coexist with updates distributed across many directions. Since this experiment includes only a single convolutional backbone, we cannot disentangle the effect of encoder architecture from that of MMEarth's narrower layer dimensions.

C Cross-Modal Transfer

We use CROMA's separate optical and SAR encoders to examine how cross-modal transfer relates to preservation of pretrained structure. Under full fine-tuning, we adapt the optical encoder to BraDD and the SAR encoder to MADOS, comparing each with the encoder pretrained on the target modality.

On BraDD, the optical encoder matches the SAR encoder at 54.1 mIoU, but this performance parity accompanies substantial subspace reorganization. Its SSI_{32} falls to approximately 0.10 in the early and middle blocks, recovering to 0.47–0.55 in the final blocks (Fig. 3). This accentuates the pattern observed in matched-modality experiments, where preservation generally increases toward the output. The concentration of changes in earlier layers is consistent with adaptation to unfamiliar input statistics.

Conversely, on MADOS, the SAR encoder reaches only 40.7 mIoU, compared with 63.6 for the optical encoder. Yet it preserves more of its pretrained structure, with SSI_{32} ranging from 0.66 to

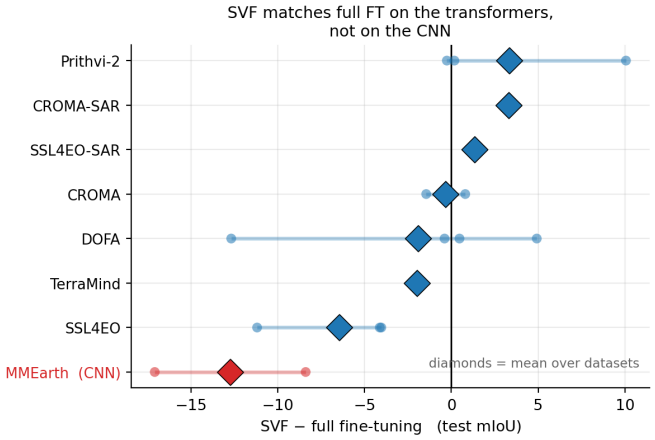


Figure 10: Difference in test mIoU between SVF and full fine-tuning across encoders. Diamonds indicate means over datasets. MMEarth exhibits the largest mean performance deficit.

0.92 across depth and a smaller relative weight change ($\rho = 0.28$ versus 0.46). Greater preservation therefore does not translate into better downstream performance in this case.

Together, these results reinforce the distinction between performance and reuse: strong performance can coexist with extensive rewriting, while high subspace preservation can accompany poor transfer. Neither performance nor SSI alone establishes successful reuse of the pretrained representation.

D Subspace Preservation Across Encoder Depth

We normalize layer positions to $[0, 1]$ and average SSI_{32} over the attention and MLP matrices within each block to compare encoders of different depths. Across all four benchmarks and most individual encoders, preservation is lowest in the early and middle layers and increases toward the output (Fig. 11). The pooled mean decreases from 0.58 in the first sixth of the network to approximately 0.51 through the early and middle blocks, before rising to 0.66 in the final sixth. This pattern suggests that fine-tuning reorganizes the earlier and intermediate transformations most strongly, while retaining more of the later pretrained structure.

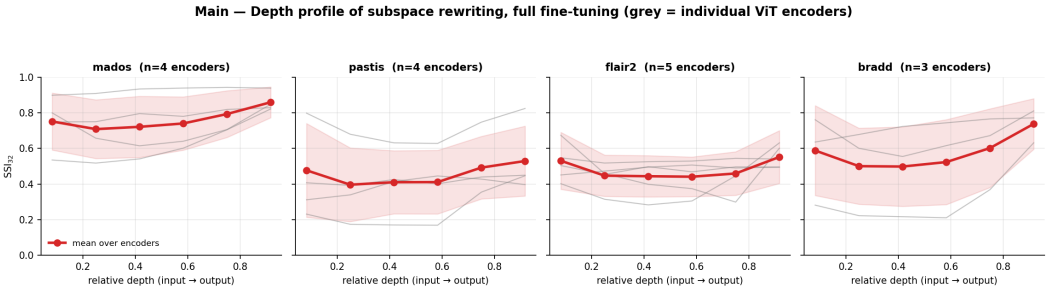


Figure 11: Depth profile of subspace preservation under full fine-tuning, per benchmark. Grey lines are individual encoders; red is the mean with ± 1 std.

E Model Selection

Main EO encoders. Tab. 7 lists the five encoders evaluated in the main set of experiments. They are chosen to vary along the axes on which current geospatial foundation models actually differ, so that a systematic difference in adaptation geometry cannot be attributed to a single shared design choice. The dominant pretraining recipes in the current literature are all represented: masked autoencoding in its canonical form (SSL4EO-MAE) and in a multitemporal variant (Prithvi-2), cross-modal

contrastive learning (CROMA), wavelength-conditioned supervision (DOFA), and multimodal pretraining with a diffusion decoder (TerraMind). Optical-only, SAR-capable and jointly optical-SAR encoders are all present, and CROMA additionally exposes a separate encoder per modality. Only Prithvi-2 was pretrained with a temporal path, so the multi-temporal benchmark contrasts a model built for it against four that were not. Widths span 384 to 1024. All five are publicly released with pretrained weights and span several model generations.

Additional encoders. Four further backbones appear in specific experiments rather than in the main sweep. *MMEarth* (ConvNeXt, median width 160) is the convolutional encoder used in the architecture ablation. *Tessera* and *Presto* (both 128-wide transformers with a native time-series path) and *AnySat* are evaluated on the crop-classification benchmark, where they provide the contrast between encoders pretrained on time series and image ViTs routed through a pixel-set path.

Encoder	Pretraining objective	Modality	$\min(m, n)$
CROMA	cross-modal contrastive + MAE	optical, SAR	1024
DOFA	wavelength-conditioned	multi-sensor	768
Prithvi-2	masked autoencoding, multitemporal	optical	768
TerraMind	multimodal generative	multi-sensor	768
SSL4EO-MAE	masked autoencoding	optical, SAR	384

Table 7: Main EO encoders, their pretraining objectives, and input modalities. Matrix dimensions are summarized by the median $\min(m, n)$ across analyzed attention and MLP weight matrices.

F Training Details

We describe the dataset, encoder, adaptation, and optimization settings below.

F.1 Datasets and Splits

We use the published split of each benchmark and do not resample it (Table 8).

Benchmark	Modality	Input	Bands	T	Classes	Train / Val / Test
MADOS	optical S2	240^2	11	1	15	1,433 / 642 / 728
PASTIS	optical S2	128^2	10	6	20	folds 1-3 / 4 / 5
FLAIR-2	aerial RGB	512^2	3	1	13	4,049 / 1,022 / 3,022
BraDD	SAR S1	48^2	2	6	2	official split
TimeMatch	optical pixel-set	16 px	10	30	9	cross-region

Table 8: Benchmarks. T is the number of time steps after subsampling. FLAIR-2 is at 20 cm ground sampling distance, the rest at 10 m.

- MADOS: 174 scenes, cropped. The 20 m and 60 m bands are resampled to 10 m by nearest neighbour. Label -1 (background) is ignored.
- PASTIS: five official geographic folds. Each series is reduced to $T = 6$ acquisitions at load time: series of at most 50 acquisitions are subsampled at evenly spaced indices in temporal order; longer series are first cut to a random, unordered subset of 50 (redrawn at every load), so their 6 frames are effectively a random draw without temporal order. This follows the PANGAEA loader. Class 19 (void) is ignored. The aerial and Sentinel-1 products are unused.
- BraDD: VV and VH polarizations. The only SAR benchmark in the panel.
- TimeMatch: 16 pixels per parcel, no spatial grid. We train on source tiles and test on a held-out tile.

F.2 Model Configurations

Each encoder runs at its released configuration and native resolution (Table 9). We tap selected blocks per encoder and pass them to the decoder as a feature pyramid.

Encoder	Backbone	Input	Bands (opt./SAR)	Tapped blocks
Main EO encoders				
CROMA	ViT-L	120	12 / 2	3, 5, 7, 11
DOFA	ViT-B	224	any[†]	3, 5, 7, 11
Prithvi-2	ViT-B	224	6 / –	3, 5, 7, 11
TerraMind	ViT-B	224	12 / 2	3, 5, 7, 11
SSL4EO-MAE	ViT-S/16	224	13 / 2	3, 5, 7, 11
DINO-on-EO				
DINOv2	ViT-L/14	224	3	5, 11, 17, 23
DINOv3	ViT-L/16	224	3	5, 11, 17, 23
Natural images				
CLIP	ViT-B/16	224	3	–
DINOv2	ViT-B/14	224	3	–
DINOv3	ViT-B/16	224	3	–

Table 9: Encoder configurations. Bands is the number of input channels expected (optical / SAR encoder). Natural-image backbones use their standard classification or segmentation head. [†]DOFA is wavelength-conditioned and accepts the dataset's bands.

- **Band matching.** By name. Bands the encoder expects but the dataset lacks are zero-filled, bands it does not expect are dropped. MADOS lacks B9 and B10, PASTIS additionally lacks B1, so CROMA, TerraMind and SSL4EO-MAE receive one to three zero-filled bands. DOFA is wavelength-conditioned and takes the dataset's bands directly. The natural-image encoders take RGB only.
- **Linear decoder** (main results). Concatenate the four tapped maps at native resolution, one 1×1 convolution to the class dimension, one bilinear upsample to label resolution. About 46 K parameters on a 768-wide encoder.
- **UPerNet decoder** (ablation, Appendix B.5). 512 channels, about 39 M parameters. On multi-temporal benchmarks it merges time with an L-TAE module (Sainte Fare Garnot & Landrieu, 2020), whereas the linear decoder stacks frames along the channel axis.
- **TimeMatch** follows a pixel-set formulation (Sainte Fare Garnot et al., 2020): each sampled pixel is processed independently by a shared encoder, and the resulting features are aggregated into a parcel representation for classification. We use a linear classification head throughout and exclude TimeMatch from the decoder ablation.
- **Preprocessing.** MADOS is randomly cropped to the encoder input size in training, biased towards labelled regions, and evaluated by sliding window over the full tile. The others are resized and evaluated whole. All inputs are standardized per band with training-split statistics.
- **DINO-on-EO** experiments use DINOv2-L/14 and DINOv3-L/16 with RGB inputs and the UPerNet decoder (with L-TAE on PASTIS), under the Earth-observation recipe of Appendix F.4: learning rate 10^{-4}, 80 epochs, batch size 8 on MADOS and 4 on PASTIS. The natural-image experiments use DINOv2-Base and DINOv3-Base.

F.3 Adaptation Settings

We compare three adaptation methods. Each is run from the pretrained weights and from an architecture-matched random initialization.

- **Full fine-tuning.** All encoder and decoder parameters are trained.
- **SVF.** Factorize each matrix once as $W = U\Sigma V^\top$, freeze U and V, train only Σ. All singular values are trained, none truncated. Adaptation can rescale the pretrained directions but not rotate them.

- **LoRA.** Freeze the matrix, add a trainable BA of rank $r = 8$ scaled by α/r with $\alpha = 16$. A is initialized from a zero-mean normal with standard deviation $1/\sqrt{r}$, B to zero, so the adapted matrix starts equal to the pretrained one. Rank and scaling are identical for Earth observation and natural images; only dropout on the low-rank branch differs (off for Earth observation, 0.05 for natural images).

Scope of SVF and LoRA. Both parameter-efficient methods wrap the attention projections and feed-forward layers of every encoder block. Normalization parameters, task heads and the patch-embedding projection are excluded, and the decoder is always trained in full. The one exception is DOFA: the two feed-forward layers of its wavelength-conditioned patch-embedding generator are wrapped as well, which is why DOFA has 50 adapted matrices in Table 10 rather than 48. Neither method reduces activation memory, only the number of updated parameters.

Random initialization. The random control uses the same architecture and skips loading the pretrained weights, so only the initial weights differ. In both cases we save the encoder weights just before adaptation, which gives an exact W_{pre} for the spectral diagnostics.

Encoder	Encoder params	Adapted matrices	SVF trainable	LoRA trainable
CROMA	303.0 M	96	98.3 K (0.03%)	3.15 M (1.04%)
DOFA	111.3 M	50	37.1 K (0.03%)	1.21 M (1.09%)
Prithvi-2	86.2 M	48	36.9 K (0.04%)	1.18 M (1.37%)
TerraMind	87.3 M	61	46.8 K (0.05%)	1.28 M (1.47%)
SSL4EO-MAE	22.6 M	48	18.4 K (0.08%)	0.59 M (2.60%)

Table 10: Trainable parameters, excluding the decoder. Percentages are relative to the encoder.

F.4 Optimization

Earth observation. AdamW, learning rate 10^{-4}, $\beta = (0.9, 0.999)$, weight decay 0.05, batch size 8 (4 for the two ViT-L DINO encoders on PASTIS, which process each frame separately), seed 234. Learning rate drops by $10\times$ at 60% and 90% of total iterations. No warmup, no gradient clipping. Validation every 5 epochs. Reported scores are from the checkpoint with the best validation mIoU, whereas the spectral diagnostics compare the encoder weights saved before adaptation with the weights at the end of training. The same learning rate is used for all three methods and both initializations, so Tables 2–4 carry no per-cell learning-rate exceptions. Segmentation runs 80 epochs, except CROMA on MADOS at 160 and FLAIR-2 at 20, the latter because its training split is about ten times larger. TimeMatch runs 30 epochs at batch size 32.

Natural images. 10 epochs, batch size 64, 224×224, weight decay 10^{-4}, seed 42. This arm needs a few learning-rate exceptions (Table 11).

Setting	Learning rate	Applies to
Default	10^{-4}	everything not listed below
Full fine-tuning	2×10^{-5}	CIFAR-10 (all backbones), VOC (CLIP, DINOv2)
LoRA	5×10^{-4}	VOC (DINOv3)
SVF	5×10^{-3}	VOC (DINOv3 and CLIP), Flowers-102 (CLIP)

Table 11: Learning rates for the natural-image experiments. Appendix B.3 examines this sensitivity.