Parameter Efficient Handling of Acquisition Artefacts

Paul Tresson[1[0000−0002−1275−4673]], Carmel Mounziegou-Mounziegou[1,2[0000−0002−8411−195X]], Diego Marcos[3[0000−0001−5607−4445]], Hadrien Tulet[1[0009−0004−8139−2380]], Gaelle Viennois[4[0000−0001−9772−343X]], and Nicolas Barbier[1[0000−0002−5323−3866]]

[1] AMAP, Univ. Montpellier, IRD, CNRS, CIRAD, INRAE, Montpellier, France
[2] Agence Gabonaise d'Études et d'Observations Spatiales, Libreville, Gabon
[3] Inria, EVERGREEN, UMR TETIS, Univ. Montpellier, Montpellier, France
[4] AMAP, Univ. Montpellier, CNRS, IRD, CIRAD, INRAE, Montpellier, France

Abstract. While deep learning has greatly improved Earth Observation, some areas such as tropical forests remain hard to monitor. Indeed, the high cloud cover, high solar angles, shadows within the canopy and low availability of labelled data make these areas susceptible to various acquisition artefacts that hinder robust monitoring. We propose a method based on Joint Embedding Predictive Architectures and Parameter Efficient Fine-Tuning to improve pre-trained model's robustness to acquisition artefacts at low computing cost. We feed a pre-trained model with random samples belonging to different acquisitions of the same places displaying varying conditions such as fog, clouds, or illumination. The model is then trained following LeJEPA's objective of projection into an isotropic normal space. We demonstrate the efficiency of our method on a dataset composed of Sentinel-2 tiles over the Congo Basin Forest and labelled test points for a forest typology mapping task. We tested this method on two foundation models, DOFA and DINOv3. With our method, we achieved up to 0.36 accuracy gain while training around 1 % of the model's parameters, using only a single GPU for a couple of hours. This method aims to be generalist and usable on any raster dataset composed of co-acquisitions and any pre-trained model.

Keywords: Unsupervised Learning · Sentinel-2 · Foundation Models · Bidirectional reflectance

1 Introduction

Space-borne Earth observation (EO) has become the foundation for studying tropical forests at large scales, with both synthetic aperture radar and optical imagery having proven useful for tasks such as monitoring deforestation [18], restoration [1] or estimating biomass [5]. However, although medium and high-resolution multispectral optical imagery, such as the one provided by the Landsat and Sentinel-2 missions, provides rich information about the forest canopy, its usefulness in tropical regions is limited by cloud cover, which can reach 90%

annually in some regions with tropical moist forests [22]. In addition to cloudiness, forests are also subject to other acquisition and atmospheric artefacts such as aerosols, bidirectional reflection, or the intricate shadows within the canopy (see [17] for a detailed analysis of spectral variability in tropical forests). Following in this paper, we regroup all these artefacts of different origin under the term "acquisition artefacts".

Advances based on deep learning, including a recent surge in EO foundation models trained via self-supervised learning (SSL) on massive, global datasets of unlabelled EO imagery, have substantially contributed to the extraction of features from EO imagery that are robust to such artefacts. However, these models have been shown degraded representations under small perturbations [14], which signals a potential lack of robustness to real-world artefacts in regions that are particularly unfavourable to optical EO, such as those with extremely high cloudiness. These regions are often underrepresented in EO datasets, which are often either biased towards the Global North, focus on densely inhabited, or are sampled in a regular grid that emphasizes large biomes [12].

One potential solution is to perform SSL on a custom dataset covering the area of interest or with a focus on the phenomenon of interest. This has proven useful for obtaining a better representation for tasks such as change detection [13] and cloud segmentation [9]. However, SSL training may require a substantial effort in terms of data collection and pre-training compute budget, giving rise to parameter-efficient finetuning (PEFT), which aims at leveraging existing pre-trained models to be adapted, in a cost-effective manner, using a small-scale dataset [10]. Although PEFT approaches, such as low-rank adaptation (LoRA [11]) are typically finetuned in supervised settings, they can also be leveraged for a second finetuning step on a specialized, unlabelled dataset.

In this paper we investigate the usefulness of a multi-temporal version of LeJEPA (Latent-Euclidean Joint-Embedding Predictive Architectures [3]), coupled with LoRA, in order to improve the representation of foundation models over a high-cloudiness tropical forest in the Congo basin, with Sentinel-2 imagery. The quality of the representation is probed with a fine-grained land-cover task. All experiments are run using a single GPU, highlighting the applicability of the approach in low-budget settings.

2 Methods

2.1 Overview

The proposed method is based on the LeJEPA [3] training objective. Instead of using artificially generated augmentations via masking and data augmentation, we obtain the different views by randomly sampling time steps of the same location, resulting in images naturally masked due to atmospheric artifacts. A pre-trained encoder is fed these samples in their different versions and must learn to project them in the same isotropic normal space. Then, samples depicting the same location at different times must be close in the projection space and samples

of different locations must be distinct. We chose LeJEPA due to its suitability for training with multiple positive pairs and its adaptability to a variety of different datasets and tasks. To ensure the lightness of the method, training is conducted using LoRA.

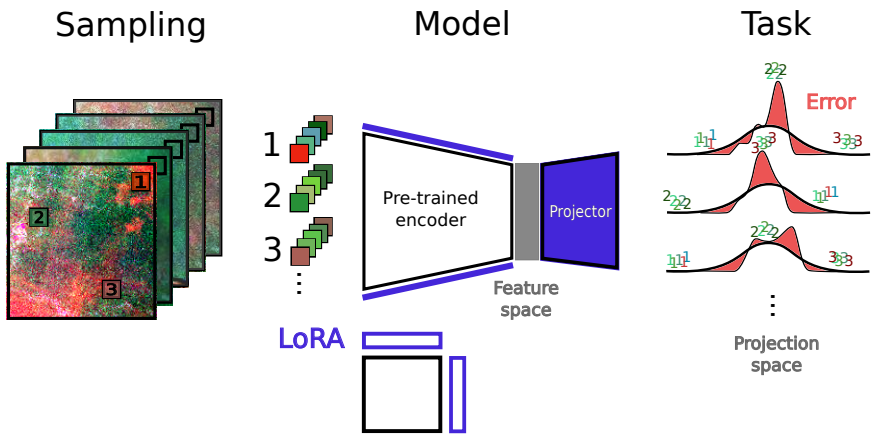

Fig. 1. Overview of the training method (samples not to scale). Sentinel 2 tiles at different dates are randomly sampled. These sampled are fed into a pre-trained encoder and projected into 128-dimension space with the constraint of (1) following an isotropic normal distribution, and (2) different versions of the samples being close in the projection space.

2.2 Datasets and sampling

The dataset is based on [20], and couples 10m resolution Sentinel-2 imagery with labels on forest typology in the Congo Basin, over Cameroon and Gabon. Following [20], only bands B03, B04, B05, B06, B07, B08, B8A, B11 and B12 were used. This choice was also done to ensure the versatility of the method for any number of band input. Training tiles correspond to the T33NTD Sentinel 2 L1C tile, validation tiles to the T33NUD tile, test tiles correspond to the T33NTC and T33NTE tiles. Different versions were chosen during the dry season of 2016 (except one train tile) to correspond to the labelled dataset of [20] (see Table 1). Additionally, acquisitions were selected to have less than 15 % cloud cover to ensure that land cover would still be visible on the majority of the samples. The objective being to make the model robust to factors varying between acquisitions, the model should in the end be used to map something that does not vary in short term. Here, we present a forest mapping task, with the hypothesis that land cover and forest typology experience little change within a single season. The labelled dataset features 2103 test points and 655 validation point (see Figure 2).

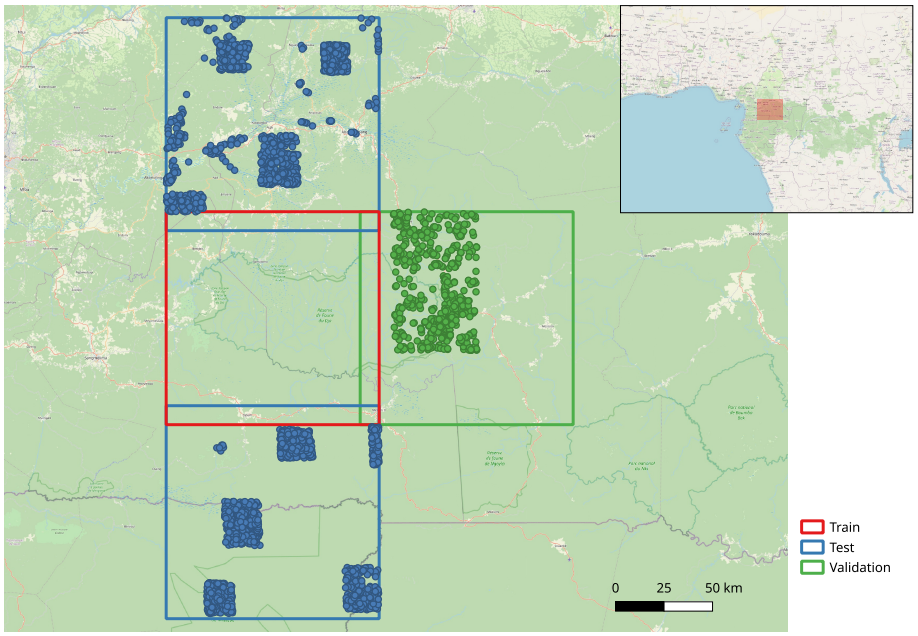

Fig. 2. Datasets extents and labelled points. Labelled validation and test points do not overlap on the train dataset extent.

Labels correspond to 10 land cover classes, mostly constituted of different forest typologies. Classes are unbalanced, with the most abundant class (Swamp forest) being represented 706 times and the least abundant (Wet Savannas) 20 times in the labelled test set (see Table 2). While some classes are relatively easy to distinguish, other like Terra Firme forest (*i.e.* solid ground forests), Gilbertio populated forests and Degraded forests require expert knowledge to be identified.

During training, tiles are sampled randomly across acquisitions with a 150×150 px bounding box (see Figures 1, 4 and 5). We chose this sampling size because this scale is relevant for a target task of forest mapping and lower sampling sizes display less spatial features and patterns. As an additional transformation, a random uniform centred crop between 30 and 150 px is applied to each sample separately. Samples are then resized to the ViT-base input size of 224×224 px and normalized using Sentinel 2 mean and standard deviation for each band (see Figure 4). We chose to include random crop in the data augmentation to provide different contexts to the model. Indeed, this kind of multi-crop based augmentations have widely been efficient in Joint Embedding Prediction Architecture settings (see [7, 2]).

Table 1. Tiles and acquisition dates of the rasters composing the different datasets

Dataset	Tile	Acquisition date
Train	T33NTD	06-12-2015
	T33NTD	16-12-2015
	T33NTD	26-12-2015
	T33NTD	25-01-2016
	T33NTD	18-02-2017
Validation	T33NUD	16-12-2015
	T33NUD	26-12-2015
Test	T33NTC	16-12-2016
	T33NTC	26-12-2016
	T33NTE	16-12-2015
	T33NTE	31-10-2016

Table 2. Class distribution within labelled tests datasets

Class	Test Zone 1	Test Zone 2
Swamp Forest	463	243
Terra Firme Forest	326	89
Gilbertio	275	11
Degraded Forest	72	254
Bare soil	2	135
Savana	0	114
Road	7	37
Swamp	0	29
Water	17	10
Wet Savanas	0	20

2.3 Validation scheme

During validation and test, the two different raster versions are sampled with a bounding box of 150 px centred on labelled points, resized to 224 px, normalized and fed to the encoder (see Figure 3 and 5). For each labelled point, a KNN (5 neighbours) is applied using the embeddings corresponding to the first version of the raster as coordinate for the target point but embeddings of the second version for its neighbours. This process is applied on all points using a stratified 5-fold scheme. Additionally to the KNN, a Linear Discriminant Analysis (LDA) is applied as a classifier on the test set to assess the generalizability of the method.

2.4 Chosen encoders

We chose DOFA [23] and DINOv3 [19] base encoders. Indeed, both represent current reliable and versatile foundation models in remote sensing and generalist

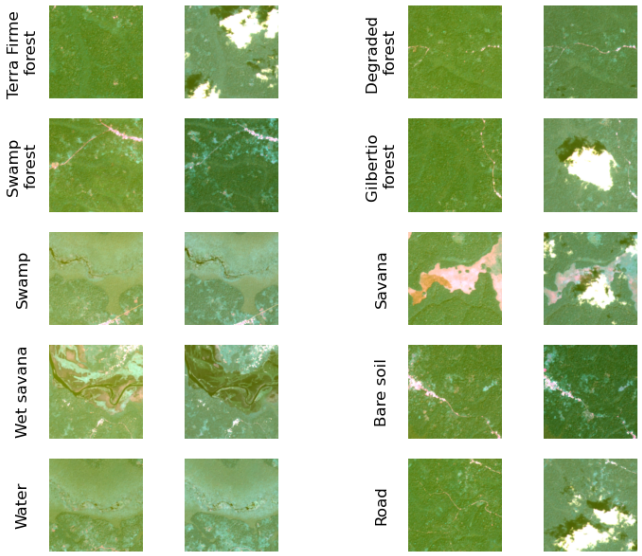

Fig. 3. RGB bands of test samples displaying two different versions of the test rasters for all represented classes. Samples are centred on the label point (see Roads).

computer vision respectively. DINOv3 being originally adapted to RGB images with three channels, the input layer is adapted following `timm`'s [21] process, *i.e.* copying RGB projection weights modulo 3.

2.5 Parameter efficient fine-tuning and training cost

Training is done using LoRA [11] with varying ranks on the encoder weights. Projector weights on the other hand, are fully trained. The projector head takes the embeddings of the encoder as an input and can have between 1 and 3 fully connected hidden layers of 2024 dimension. After these hidden layers, the projector head ends into a 128-dimension projection space were LeJEPA loss is computed.

All experiments were done on a single Nvidia H100 GPU, using only 40Gb of the 80Gb VRAM available, and with a 4-hour time limit.

3 Results

3.1 Accuracy improvement

Training leads to accuracy improvements on each LoRA setting, encoder and projector modality (see Table 3 and Figure 6), with up to 0.34 gain for DOFA and 0.36 gain for DINOv3.

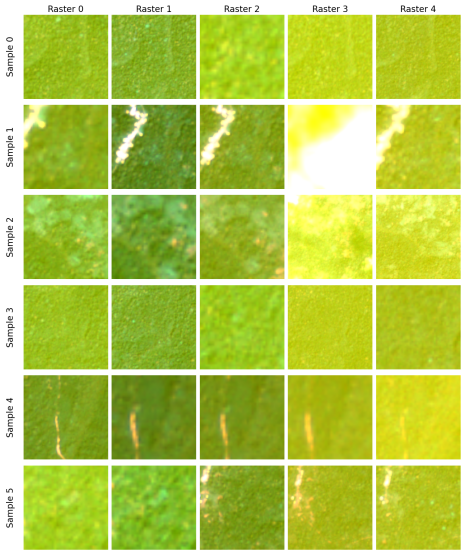

Fig. 4. RGB bands of train samples displaying five different versions (in colums) of the train rasters with additional transforms (random centred crop).

The confusion matrix of the best run is displayed on Figure 7. While the predictions improve overall with less confusion between forest types, confusion between Roads and Bare Soil label worsen, maybe due to the domain shift happening with training and the scale of the features compared to the fed image (see Figure 3). Overall, performances improve for all classes except wet Savanas over the frozen-backbone baseline, even for classes that are less represented in the dataset such as Swamps or Water (see Table 4).

3.2 Parameter efficient fine-tuning

The accuracy over the number of trained parameters is displayed Figure 8. With LoRA, most of the parameters updated during training belong to the projector head. While costing more, projector layers ensure a more stable training, with over 0.50 accuracy being ensured with more than two layers but not without. This is also visible with validation behaviour, where runs with deeper projector heads show higher validation accuracy (see Figure 6). In the best performing run, only 0.93M parameters were trained (including projector head), which represents around 1.1 % of the 85.6M parameters of the DINOv3 model in the the base version.

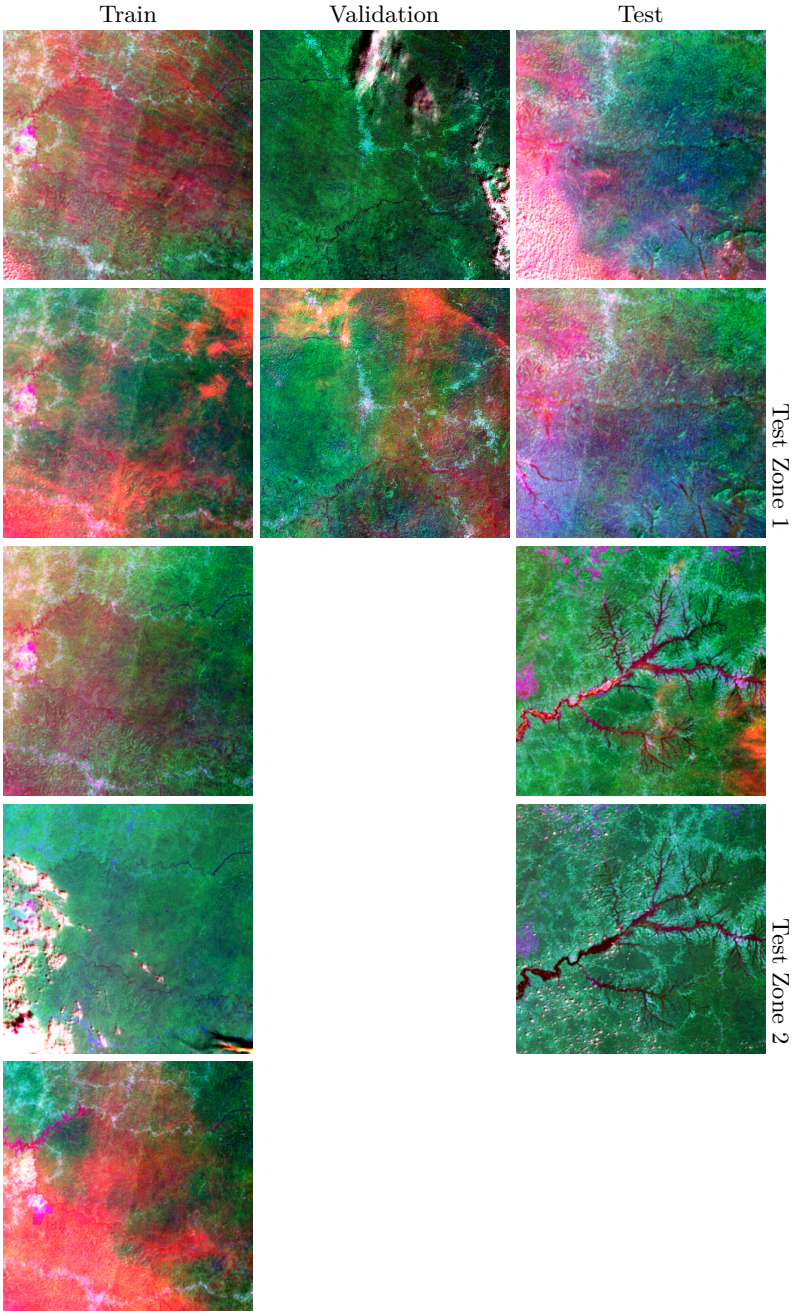

Fig. 5. RGB composition of the Sentinel 2 tiles in the different datasets (with bands 11, 8 and 2). This highlights the large scale variability in acquisition conditions between different versions within a dataset.

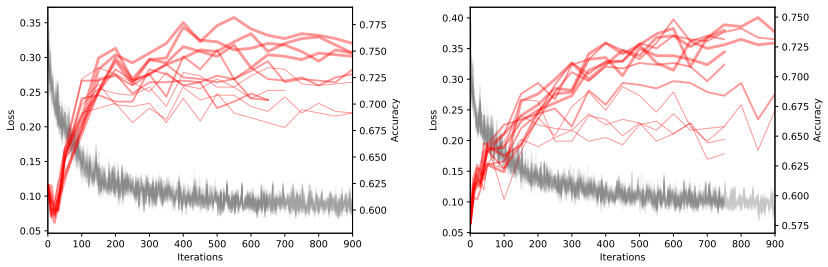

Fig. 6. Losses and validation accuracy for the different runs of DINOv3 (left) and DOFA (right). Accuracy line width corresponds to projector layers. Some runs end earlier due to the 4 hours time limit.

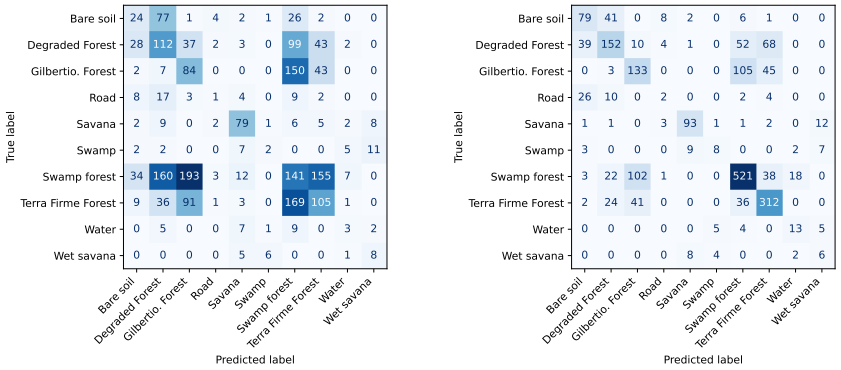

Fig. 7. Confusion matrixes before and after training for the run with the highest accuracy.

4 Discussion

4.1 Upscaling and qualitative results

As an upscaling experiment, we leveraged the raw L1C Sentinel 2 ortho-mosaic composed by [20] (see Figure 10). This mosaic overlaps partly train, validation and test sets but is composed by acquisitions never seen by the model and has a larger extent.

When projecting the model embeddings for different source tiles in a joint PCA or UMAP [15], one can see that the effet of source tile is still the defining factor in the global topology of the feature space (see Figure 9) rather than forest typology. However, points comming from different source tiles seem to mix more after training than before.

Several benefits of our method are demonstrated on the zooms displayed Figure 10. First, the main artefact due to Sentinel 2 swath is mostly erased. Second, buildings now appear as a distinct component (in cyan). Third, the

Table 3. Ablation over projector depth and LoRA rank for two backbones. KNN denotes top-1 KNN classification accuracy, with gain being the improvement over the frozen-backbone baseline. LDA accuracy is presented as well to assess generalizability to different classifiers. Best result per backbone is in **bold**.

Model	Proj. layers	LoRA rank	Trained params (M)	Iters.	KNN (gain)	LDA (gain)
DOFA-base	1	2	0.58	150	0.32 (0.07)	0.35 (0.07)
	1	4	0.70	150	0.34 (0.09)	0.38 (0.10)
	1	8	0.94	450	0.49 (0.24)	0.54 (0.26)
	1	16	1.43	350	0.50 (0.24)	0.55 (0.27)
	2	2	1.83	600	0.52 (0.27)	0.57 (0.29)
	2	4	1.95	500	0.53 (0.27)	0.56 (0.28)
	2	8	2.20	800	0.58 (0.32)	0.55 (0.27)
	2	16	2.68	600	0.57 (0.31)	0.58 (0.30)
	3	2	6.03	950	0.55 (0.29)	0.58 (0.30)
	3	4	6.15	850	**0.60 (0.34)**	**0.61 (0.33)**
	3	8	6.40	500	0.53 (0.27)	0.54 (0.26)
	3	16	6.88	750	0.58 (0.32)	0.56 (0.28)
DINOv3-base (lvd1689m)	1	2	0.58	150	0.52 (0.26)	0.51 (0.24)
	1	4	0.69	400	0.59 (0.33)	0.57 (0.29)
	1	8	0.93	400	**0.62 (0.36)**	**0.61 (0.33)**
	1	16	1.39	150	0.49 (0.23)	0.49 (0.21)
	2	2	1.83	300	0.59 (0.33)	0.58 (0.30)
	2	4	1.95	600	0.57 (0.32)	0.56 (0.28)
	2	8	2.18	350	0.57 (0.31)	0.58 (0.31)
	2	16	2.65	500	0.60 (0.35)	0.60 (0.32)
	3	2	6.03	400	0.60 (0.34)	0.58 (0.31)
	3	4	6.15	1200	0.61 (0.35)	0.58 (0.31)
	3	8	6.38	550	0.59 (0.33)	0.58 (0.31)
	3	16	6.85	600	0.59 (0.33)	0.58 (0.31)

impact area of clouds is reduced to the regions of highest cloud density. This suggests that, rather than focussing on a cloud when it is present, the model seems to now focus on the forest under it. Naturally, on areas with large and dense clouds, the representation is no longer aligned with the underlying land cover.

4.2 Transformation and sampling choices

As an ablation experiment, we trained the models with the modalities presented Table 3 but removing the random centred crop from the transformations. Runs without crops show no more than 3 % gain. This would tend to confirm that multiple crops providing different context are very efficient in JEPA settings (see discussion on the subject in [2]).

Table 4. Classification metrics for all land cover classes for the best performing model (with gain compared to frozen-backbone baseline).

Class	Precision (gain)	Recall (gain)	F1-score (gain)	Support
Bare soil	0.58 (0.40)	0.52 (0.30)	0.54 (0.35)	153
Degraded Forest	0.47 (0.12)	0.60 (0.34)	0.53 (0.23)	253
Gilbertio. Forest	0.47 (0.17)	0.47 (0.26)	0.47 (0.22)	286
Road	0.05 (0.02)	0.11 (0.03)	0.06 (0.03)	18
Savana	0.82 (0.12)	0.82 (0.18)	0.82 (0.15)	113
Swamp	0.28 (0.21)	0.44 (0.26)	0.34 (0.24)	18
Swamp forest	0.74 (0.54)	0.72 (0.49)	0.73 (0.51)	727
Terra Firme Forest	0.75 (0.50)	0.66 (0.37)	0.71 (0.43)	470
Water	0.48 (0.37)	0.37 (0.23)	0.42 (0.29)	35
Wet savana	0.30 (-0.10)	0.20 (-0.08)	0.24 (-0.09)	30
Accuracy	0.63 (0.36)	0.63 (0.36)	0.63 (0.36)	0.63
Macro avg.	0.49 (0.24)	0.49 (0.24)	0.49 (0.24)	2103
Weighted avg.	0.64 (0.36)	0.63 (0.36)	0.63 (0.36)	2103

With this method, one can choose the kind of artefacts the model should be robust to and chose co-acquisitions accordingly. In our case, forest typology can be hypothesised to experience slow changes and that most of the variation between different versions of a sample can be considered artefactual. However, the resulting model may not be relevant as such to study variables that vary between acquisition, such as tree phenology or short term stress responses, for instance.

4.3 Versatility of the method

In this study, we have focused on L1C Sentinel products over tropical forests because this modality incorporates atmospheric artefacts that remain challenging to correct [8] but also strong directional effects due to the dense canopy. However, inheriting from the versatile design philosophy of LeJEPA, this method should remain valid given any raster inputs (with different versions belonging to the same modality) and any model. For instance, it could be used on repeated UAV flights over a zone, such as those conducted over Central Africa [4] and allow to takle different artefact sources. Furthermore, similar methods based on JEPA frameworks have been successfully leveraged to mitigate acquisition artefacts over tropical forests with airborne hyperspectral imaging [16].

5 Conclusion

Our results show the advantage of combining performing self-supervised fine-tuning, combining LeJEPA with parameter-efficient LoRA, in order to obtain substantial gains on a downstream task (almost doubling the accuracy) at a low cost in terms of data and compute.

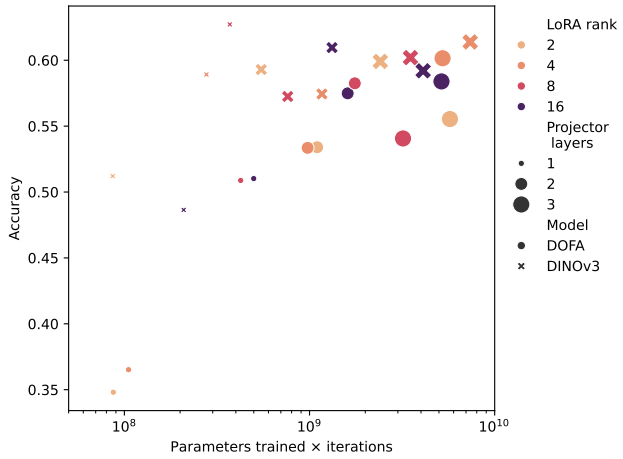

Fig. 8. Accuracy vs. training cost of different models and configurations.

Availability Source code is available at https://github.com/umr-amap/pehaa and https://forge.ird.fr/amap/pehaa.

References

1. de Almeida, D.R., Vedovato, L.B., Fuza, M., Molin, P., Cassol, H., Resende, A.F., Krainovic, P.M., de Almeida, C.T., Amaral, C., Haneda, L., et al.: Remote sensing approaches to monitor tropical forest restoration: Current methods and future possibilities. Journal of Applied Ecology **62**(2), 188–206 (2025)
2. Balestriero, R., Ibrahim, M., Sobal, V., Morcos, A., Shekhar, S., Goldstein, T., Bordes, F., Bardes, A., Mialon, G., Tian, Y., et al.: A cookbook of self-supervised learning. arXiv preprint arXiv:2304.12210 (2023)
3. Balestriero, R., LeCun, Y.: Lejepa: Provable and scalable self-supervised learning without the heuristics. arXiv preprint arXiv:2511.08544 (2025)
4. Barbier, N., Ploton, P., Heuret, P., Engel, J., Burban, B., Ball, J., Bastin, J.F., Ilondea, B.A., Mbock, G., Sonké, B., et al.: Bridging scales in tropical forest monitoring with uav observatories integrating ecosystem function, biodiversity, and satellite data. Tech. rep., Copernicus Meetings (2026)
5. Bossy, T., Ciais, P., Renaudineau, S., Wan, L., Ygorra, B., Adam, E., Barbier, N., Bauters, M., Delbart, N., Frappart, F., et al.: State of the art in remote sensing monitoring of carbon dynamics in african tropical forests. Frontiers in Remote Sensing **6**, 1532280 (2025)
6. Brown, C.F., Kazmierski, M.R., Pasquarella, V.J., Rucklidge, W.J., Samsikova, M., Zhang, C., Shelhamer, E., Lahera, E., Wiles, O., Ilyushchenko, S., et al.: Alphaearth foundations: An embedding field model for accurate and efficient global mapping from sparse label data. arXiv preprint arXiv:2507.22291 (2025)
7. Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., Joulin, A.: Emerging properties in self-supervised vision transformers. In: Proceedings of the International Conference on Computer Vision (ICCV). pp. 9650–9660 (2021)

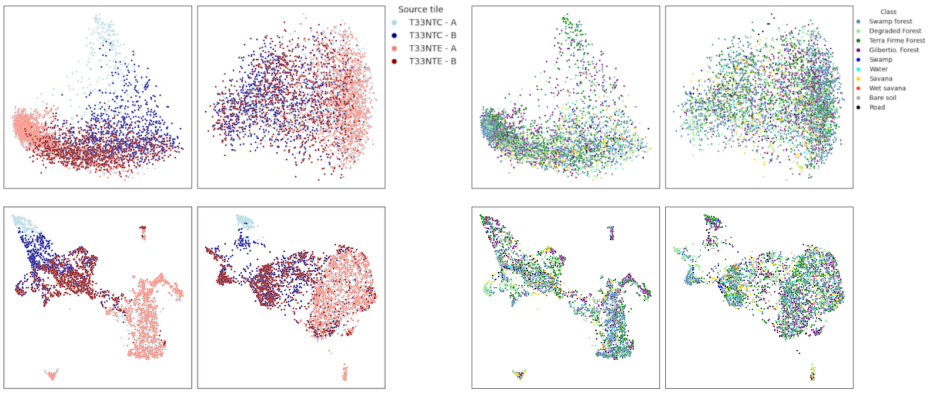

Fig. 9. PCA (top) and UMAP (bottom) projection of the best model's feature space before (left) and after (right) training. Points are colored by source images or by class. For source images, pale blue and dark blue points correspond to the same label points but feeding version A or B of the test tile into the model (same for pink and red points). Conversly, all points colored by class are featured two times, with projections coming from versions A or B of the test tiles.

8. Doxani, G., Vermote, E.F., Roger, J.C., Skakun, S., Gascon, F., Collison, A., De Keukelaere, L., Desjardins, C., Frantz, D., Hagolle, O., et al.: Atmospheric correction inter-comparison exercise, acix-ii land: An assessment of atmospheric correction processors for landsat 8 and sentinel-2 over land. Remote Sensing of Environment **285**, 113412 (2023)
9. Gbodjo, Y.J.E., Hughes, L.H., Molinier, M., Tuia, D., Li, J.: Self-supervised representation learning for cloud detection using sentinel-2 images. Remote Sensing of Environment **334**, 115205 (2026)
10. Han, Z., Gao, C., Liu, J., Zhang, J., Zhang, S.Q.: Parameter-efficient fine-tuning for large models: A comprehensive survey. Transactions on Machine Learning Research (2024), https://openreview.net/forum?id=lIsCS8b6zj
11. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al.: Lora: Low-rank adaptation of large language models. Iclr **1**(2), 3 (2022)
12. Kaur, A., Purohit, M., Muhawenayo, G., Rolf, E., Kerner, H.: Pretrain where? investigating how pretraining data diversity impacts geospatial foundation model performance. arXiv preprint arXiv:2604.21104 (2026)
13. Leenstra, M., Marcos, D., Bovolo, F., Tuia, D.: Self-supervised pre-training enhances change detection in sentinel-2 imagery. In: International conference on pattern recognition. pp. 578–590. Springer (2021)
14. Li, X., Tao, Y., Zhang, S., Liu, S., Xiong, Z., Luo, C., Liu, L., Pechenizkiy, M., Zhu, X.X., Huang, T.: Reobench: Benchmarking robustness of earth observation foundation models. In: The Thirty-ninth Annual Conference on Neural Information Processing Systems Datasets and Benchmarks Track (2025)
15. McInnes, L., Healy, J., Melville, J.: Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426 (2018)
16. Prieur, C., Braham, N.A.A., Tresson, P., Vincent, G., Chanussot, J.: Prospects for mitigating spectral variability in tropical species classification using self-supervised

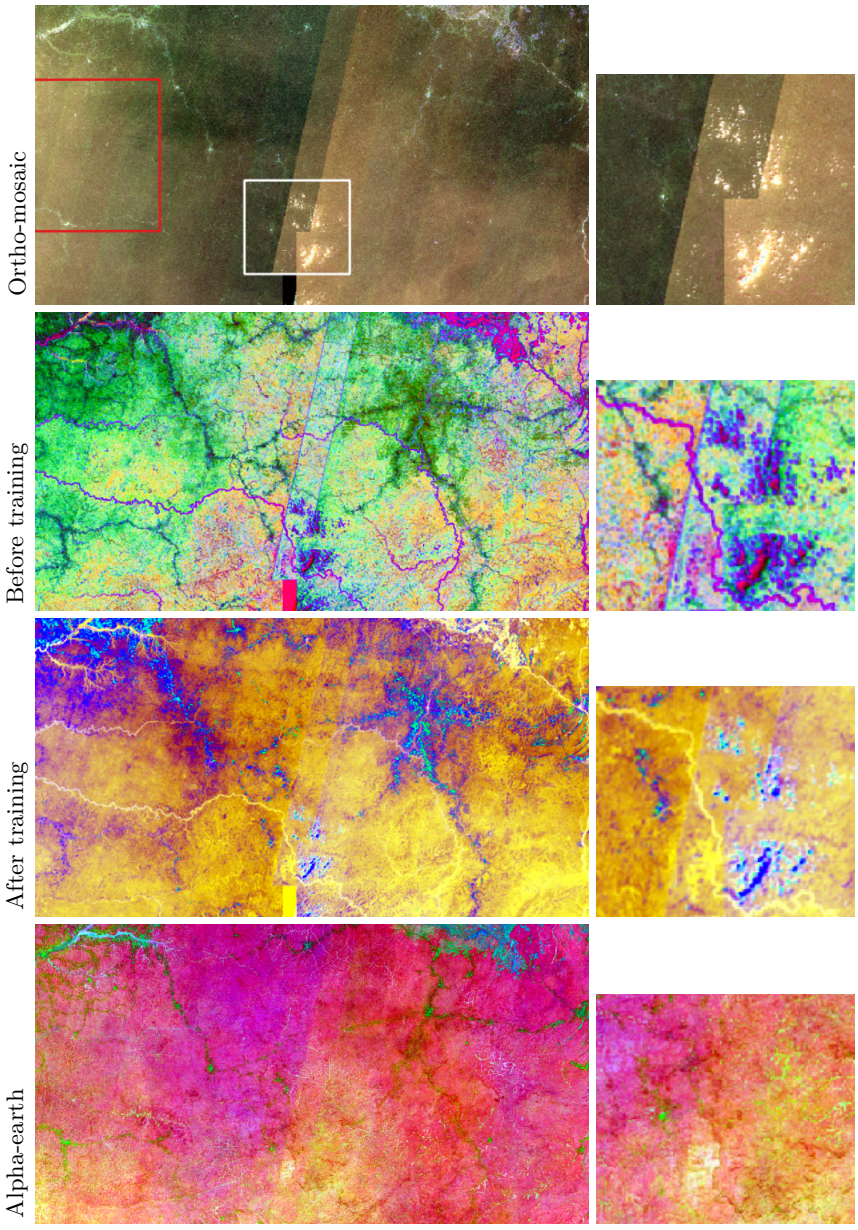

Fig. 10. Ortho-mosaic (with extent of the train datasets), PCA of the best model's embeddings before and after training, and a PCA of alpha-earth's [6] embeddings as a reference. This shows an improved robustness to acquisition artefacts such as Sentinel 2 swaths visible in the centre of the mosaic. A detail of a selected zone is displayed on the right. Alpha-earth embeddings summarize a year worth of acquisitions and multiple input modalities, as such, it does not suffer from clouds. However, Alpha-earth seem to inherit from some Sentinel-2 artefacts.

learning. In: 2024 14th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS). pp. 1–5. IEEE (2024)
17. Prieur, C., Laybros, A., Frati, G., Schläpfer, D., Chanussot, J., Vincent, G.: Investigating abiotic sources of spectral variability from multitemporal hyperspectral airborne acquisitions over the french guyana canopy. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **17**, 18751–18768 (2024)
18. Reiche, J., Lucas, R., Mitchell, A.L., Verbesselt, J., Hoekman, D.H., Haarpaintner, J., Kellndorfer, J.M., Rosenqvist, A., Lehmann, E.A., Woodcock, C.E., et al.: Combining satellite data for better tropical forest monitoring. Nature Climate Change **6**(2), 120–122 (2016)
19. Siméoni, O., Vo, H.V., Seitzer, M., Baldassarre, F., Oquab, M., Jose, C., Khalidov, V., Szafraniec, M., Yi, S., Ramamonjisoa, M., et al.: Dinov3. arXiv preprint arXiv:2508.10104 (2025)
20. Viennois, G., Tulet, H., Tresson, P., Ploton, P., Couteron, P., Barbier, N.: Sentinel-2 forest typology mapping in central africa: assessing deep learning and image preprocessing effects. Frontiers in Remote Sensing **6**, 1682132 (2025)
21. Wightman, R.: Pytorch image models. https://github.com/rwightman/pytorch-image-models (2019). https://doi.org/10.5281/zenodo.4414861
22. Wilson, A.M., Jetz, W.: Remotely sensed high-resolution global cloud dynamics for predicting ecosystem and biodiversity distributions. PLoS biology **14**(3), e1002415 (2016)
23. Xiong, Z., Wang, Y., Zhang, F., Stewart, A.J., Hanna, J., Borth, D., Papoutsis, I., Saux, B.L., Camps-Valls, G., Zhu, X.X.: Neural plasticity-inspired foundation model for observing the Earth crossing modalities. arXiv preprint arXiv:2403.15356 (2024)