Efficient Self-Supervised Learning for Earth Observation via Dynamic Dataset Curation

Thomas Kerdreux[1,*]   Alexandre Tuel[1]   Quentin Febvre[2]   Alexis Mouche[2]

Bertrand Chapron[2]

[1]Galeio, Paris, France   [2]Ifremer, UMR CNRS LOPS, Brest, France

[*]Corresponding author: tkerdreux@galeio.fr

Abstract

Self-supervised learning (SSL) has enabled the development of vision foundation models for Earth Observation (EO), demonstrating strong transferability across diverse remote sensing tasks. While much research has focused on network architectures and training strategies, the role of dataset curation – particularly in balancing and diversifying pre-training datasets – remains underexplored. In EO, this challenge is exacerbated by the strong redundancy and heavy-tailed distributions of satellite imagery, which can lead to biased representations and inefficient training.

In this work, we introduce a dynamic dataset pruning strategy designed to enhance SSL pre-training efficiency by maximizing dataset diversity and balancedness. Our method iteratively refines the training set without relying on a pre-existing feature extractor, making it well-suited for domains where curated datasets are unavailable. We illustrate our approach on the Sentinel-1 Wave Mode (WV) Synthetic Aperture Radar (SAR) archive, a challenging dataset primarily composed of ocean observations. We train models from scratch on the entire Sentinel-1 WV data archive over 10 years. Our results, validated across three downstream tasks, show that dynamic pruning improves both computational efficiency and feature quality, leading to better transferability in real-world applications. This work provides a scalable and adaptable solution for dataset curation in EO, paving the way for more efficient and generalizable foundation models in remote sensing.

We release the weights of Nereus-SAR-1, the first foundation model in our Nereus models family — a series of models dedicated to ocean observation and analysis using SAR imagery, at github.com/galeio-research/nereus-sar-models/.

1. Introduction

Self-supervised learning (SSL) has recently demonstrated much potential for Earth Observation (EO). Vision founda-

tion models, in particular, have shown a remarkable ability to learn task-agnostic image representations without requiring any human labels. A wide range of EO foundation models are now available, covering multiple modalities, whether optical, radar, multispectral or hyperspectral data [e.g., 6, 21, 25, 30, 32, 33]. Much effort has been put into exploring new network architectures and training strategies to improve the performance of these foundation models, and in particular their ability to handle multiple remote sensing modalities. However, one aspect of model pre-training has so far attracted relatively little attention, namely the selection of images to build pre-training datasets.

There is now a large body of evidence that SSL model performance on downstream prediction tasks is strongly dependent on the scale, on the diversity, and on the balancedness of the pre-training datasets [3, 19, 37, 55]. By contrast, training vision models on random collections of images leads to a drop in performance [51], likely because of the heavy-tailed distribution of concepts in uncurated datasets [31]. Large Language Models (LLMs) are trained on carefully curated, high-quality datasets [35, 53]. The challenge of dataset balancedness is particularly acute in EO because remote sensing imagery exhibits a highly non-uniform distribution, with some concepts, like oceans, deserts, or tropical forests, being much more frequent than others. Such a strong imbalance may lead to biases toward a few dominant scenes in the learned representations, as well as a waste of computational resources during training.

To overcome this challenge, most existing EO foundation models rely on a handful of pre-compiled, sometimes carefully curated datasets, like BigEarthNet [47], SSL4EO [59], SAtlasPretrain [4] or FMoW [10]. Data curation in these datasets is typically achieved by selecting images according to a criterion, often based on land cover, meant to ensure balancedness in the resulting dataset [47, 48, 59]. A similar approach was taken in the Prithvi model [49], whereby individual Harmonized Landsat/Sentinel-2 images were selected based on average climatology and land cover

to build the model's pre-training dataset.

However, such approaches have important drawbacks. First, they require an a priori selection criterion to build the pre-training dataset. While land cover may certainly make sense for land-based imagery, it is not the only choice one could think of. Besides, for some modalities, like SAR imagery over the ocean, devising such a criterion may require substantial efforts that are not replicable for other modalities (for instance, diversity in ocean imagery is largely driven by changing weather conditions). Furthermore, there is no guarantee that the resulting datasets are truly balanced, since the selection criteria, like land cover, are always reductive (diversity in land imagery is not only driven by differences in land cover, but also by seasonality, weather, extreme events, etc.)

We address this challenge by introducing a simple, yet effective strategy to automatically prune the pre-training dataset in SSL to maximize its diversity and balancedness. To do so, we build on previous work [55] to define a dynamic coreset selection strategy over time for highly redundant SSL datasets, particularly in scenarios where a reliable pre-existing feature extractor is unavailable for pruning. We apply our method to radar imagery over the oceans, a choice example of a *strongly redundant* EO dataset. Using three different downstream tasks, we show how dynamic pruning increases model performance while improving computational efficiency. Our results demonstrate that selecting a diverse and balanced subset of pre-training data leads to improved feature representations, ultimately enhancing the transferability of self-supervised models to real-world applications. By validating our approach across multiple tasks, we highlight its robustness and generalizability, offering a practical solution for curating large-scale, redundant EO datasets in the absence of predefined feature extractors. We also show that a carefully-trained, single-modality foundation model can outperform large, multi-modal models.

2. Related Work

Dataset pruning/coreset selection Scaling laws highlight the balance between dataset size, diversity, and model size as key factors in achieving optimal performance in self-supervised learning. These empirical laws suggest that performance improvements require exponentially increasing dataset sizes. However, it is known that naive data pruning can be a promising approach to mitigate the constraints imposed by these scaling laws [44].

In practice, self-supervised training across different domains, including images [5], language [2, 17], and vision-language models [7, 36, 56], commonly involves an initial curation phase of the large datasets. This curation process often goes beyond simple deduplication [52] and aims to remove semantic redundancies within the dataset.

Numerous dataset pruning, coreset selection, and dataset distillation strategies have been developed within machine learning and deep learning to influence model training by effectively modifying the training dataset. These strategies follow various heuristics and objectives, such as leveraging model optimization status to encourage learning from harder examples—such as focal loss or self-paced reweighting [13, 26, 54]; progressively increasing data sample difficulty using augmentation techniques [11, 20, 66]; utilizing label information for pruning [34, 43, 63]; or pruning to ensure better data coverage [65, 67].

The curation of large datasets for self-supervised training presents distinct challenges, particularly in cases where no effective feature extractor exists or where datasets exhibit high redundancy. For EO datasets, most existing approaches rely on metadata to guide dataset sampling, such as balancing datasets based on land cover types or urbanization patterns. However, in cases where metadata are unavailable or insufficiently informative, alternative/complementary strategies must be explored to ensure dataset balance and effective training outcomes.

Recent work has sought to perform unsupervised dataset curation via hierarchical clustering in the latent space [55]. However, the proposed approach still requires an existing feature extractor and cannot therefore be considered fully unsupervised. We build on these efforts to develop a dynamic sampling strategy that functions even without a pre-existing feature extractor. As a case study, we focus on a niche modality – Sentinel-1 Synthetic Aperture Radar (SAR) imagery in wave mode – to demonstrate the effectiveness of this approach. In this example, the training dataset can be considered overly redundant, as wave mode data mainly consists of images of the ocean, which are exceedingly similar.

Sentinel-1 SAR foundation models Multiple SAR foundation models have already been published in the literature.
- CROMA (Contrastive Radar-Optical Masked Autoencoders) [15]: CROMA is based on a hybrid training strategy mixing MAE-type reconstruction with a contrastive (InfoNCE) loss, using optical and radar images as positive pairs. CROMA was pre-trained on the SSL4EO-S12 [59] dataset – a geographically and seasonally diverse dataset of 1 million paired Sentinel-2 L2A and Sentinel-1 IW GRD samples.
- DeCUR (Decoupling Common and Unique Representations) [60]: the idea behind the approach is to take into account the fact that, in multi-modal self-supervised learning, some modalities usually contain information that is not present in other modalities (for example, radar sensors will see through clouds, but not optical ones [16]). DeCUR therefore relies on cross-correlation matrices between embeddings from different modalities to separate between cross-modal common ones and modality-unique

ones. Intra-modal representation are also enhanced using a regularization term on modality-specific embeddings to avoid collapse. The model was also pre-trained on the SSL4EO-S12 [59] dataset.

- MoCo (Momentum Contrast) [59]: like DINO [19], the method we use in our experiments, MoCo is a contrastive self-supervised learning strategy that involves building dynamic dictionaries of embeddings. In the InfoNCE loss, instead of comparing an image to the ones from the same batch, MoCo allows to query from a larger queue, in which embeddings are computed on-the-fly using a momentum-updated encoder. The model was pre-trained on the SSL4EO-S12 [59] dataset.

- DOFA (Dynamic One-For-All) [64]: DOFA avoids using separate encoders when training on multiple modalities by using a dynamic weight generator network that adjusts the weights of a shared vision backbone according to the wavelength(s) of the spectral bands of the input image. The model was pre-trained on a variety of datasets, including Sentinel-1 IW GRD images from the SAtlasPretrain [4] dataset.

- SoftCon [61]: unlike other strategies, the SoftCon (Soft Contrastive learning) strategy relies on automatically generated image labels to compute a modified InfoNCE loss, which weighs the similarity between images in the same batch according to their labels. Labels are obtained by matching data from Google's Dynamic World land cover maps with images from the SSL4EO-S12 [59] dataset, which is used for pre-training. The training also involves Siamese masking of image patches for enhanced robustness.

- WV-net [18]: a ResNet50 model trained on Sentinel-1 WV images using a SimCLR strategy, along with a set of complex data augmentation techniques designed specifically for SAR imagery. WV-net was pre-trained on the entire Sentinel-1 WV archive, at the time consisting of about 9.9 million images.

All these models (except for WV-net) were pre-trained on Sentinel-1 IW GRD only, which consists of intensity-only, dual polarization (usually VV+VH) SAR images. WV images, by contrast, are only acquired with the VV polarization.

3. Methodology and Results

3.1. Sentinel-1 WV data

Our work is based on Synthetic Aperture Radar (SAR) data acquired by the Sentinel-1 constellation, specifically *WaVe mode* (WV) data acquired over the oceans. The ESA Sentinel-1 mission is a constellation of two polar-orbiting, sun-synchronous satellites (S-1 A and S-1 B) launched in April of 2014 and 2016, respectively. They are equipped with a C-band SAR instrument with frequency 5.405 GHz

(5.5 cm wavelength). They have a 12-day repeat cycle at the equator, and are phased at $180°$ to provide an effective 6-day repeat cycle. S-1 B ceased to function in late 2021, while S1-C was launched in December 2024 and is still in commissionning phase. The launch of S1-D is planned for late 2025. These satellites operate in four different acquisition modes: WV, Interferometric Wide (IW), Extra Wide (EW) swath and Stripmap (SM) modes. These modes represent different acquisition strategies (chiefly in terms of swath width, image footprint, resolution and polarisations), and are deployed over different geographies. WV data is acquired almost exclusively over oceans (a few images cover Africa, Australia and the Great Lakes region of North America). It consists in $\approx 20 \times 20$ km images with a spatial resolution of 5×5 m, every 100 km along the orbit, acquired alternately on two different incidence angles ($\approx 23°$ and $36°$). WV data are only available in single-look complex (SLC) format, while the more generally used IW data are also available in ground range detected (GRD) format.

The WV data archive we used includes about 12 million different images, covering the full 2015-2024 period (each satellite produces approximately 60,000 WV images per month). Their spatial distribution is very uneven. Some regions, such as the South Pacific Ocean, have a much higher concentration of images, while others, notably the North Atlantic, are more sparsely covered (European waters are observed in IW and EW modes). To build an initial training database, we therefore sample from these 12 million images uniformly in space and time, retaining about 2 million different images. We make sure during this step to exclude all images used in our benchmarks (see section 3.5). To reduce the computational load, we downsample images to a 50-meter resolution, which preserves sufficient information richness for model training and relevant benchmarking.

Sentinel-1 WV images are characterized by a high level of redundancy. Many images exclusively display ocean waves, and even when other phenomena are present (sea-ice or rain cells, for instance), the background still chiefly consists of ocean waves. As a consequence, many phenomena that can be seen in WV images are rare, relatively speaking, compared to ocean waves. This results in significant unbalancedness in the dataset.

3.2. Training Framework

For our experiments, we employ the DINO self-supervised learning strategy [8]. DINO is a contrastive learning strategy that relies on self-distillation between a teacher and a student model. It consists in learning a global representation of an image using a mixture of global and local views of that image. Global views pass through the teacher model, which is tasked with predicting high-level features, while the student model is tasked with predicting

these same features from the local views. Regarding network architectures, we use Vision Transformers (ViTs), specifically ViT-Small and ViT-Base, with respectively 27 and 82 million parameters. For a fair comparison with some of the existing SAR foundation models, we also train a ResNet50 in the same conditions. This choice aligns with the unimodal nature and high redundancy of the dataset, where excessively large models would probably be unnecessary. Note, however, that our proposed approach is compatible with any training strategy and network architecture.

3.3. Dynamic coreset selection for highly redundant dataset

We present our simple dynamic sampling strategy in Algorithm 1. We start by training a model with the selected SSL loss (in this case, the DINO loss) for a pre-specified number of warmup epochs, until a sufficiently effective feature extractor is learned. Then, at scheduled intervals during training, the original dataset is clustered, and a subset of it is sampled. The SSL model then continues training for the remaining epochs until the next resampling step on this dynamically selected data.

Algorithm 1 SSL on an Overly Redundant Dataset (SSL-ORD)

Require: Dataset $\mathcal{D}$, Total training epochs T, Warm-up epochs T_w, Sampling epoch frequency T_s, SSL training algorithm $\mathcal{A}(\cdot, \cdot, \cdot)$, Pruning strategy $\mathcal{S}(\cdot, \cdot, \cdot, \cdot)$, Reduction ratio ρ, Sampling sequence (η_i).
Ensure: Trained model θ
 1: Initialize model parameters θ_0
 2: $\theta_{T_\mathrm{w}-1} \leftarrow \mathcal{A}(\theta_0, \mathcal{D}, T_\mathrm{w})$ ▷ Train without dataset pruning
 3: **for** epoch $e = T_\mathrm{w}$ to T **do**
 4:     **if** $e \mod T_\mathrm{s} = 0$ **then**   ▷ Apply dataset pruning at scheduled intervals
 5:        $\mathcal{D}_{e+1} \leftarrow \mathcal{S}(\theta_e, \mathcal{D}, \rho, \eta_e)$
 6:     **else**
 7:        $\mathcal{D}_{e+1} \leftarrow \mathcal{D}_e$
 8:     **end if**
 9:     $\theta_{e+1} \leftarrow \mathcal{A}(\theta_e, \mathcal{D}_e, 1)$
10: **end for**

Algorithm 1 relies on access to two key procedures. The self-supervised learning (SSL) training algorithm $\mathcal{A}(\cdot, \cdot, \cdot)$ takes a model checkpoint θ and performs SSL on the dataset $\mathcal{D}$ for T epochs. This approach is therefore highly versatile with respect to SSL methods. While we conducted experiments using DINO for simplicity, any SSL strategy could be employed within this framework. It provides a flexible foundation for conducting extensive numerical experiments

across diverse training structures by relying solely on the features extracted from the SSL-trained model. Interestingly, certain SSL methods compute additional latent concepts beyond extracted features, potentially providing valuable insights for improving unsupervised dataset clustering. Exploring a deeper integration between the pruning strategy and SSL mechanisms would be a fruitful direction for future research.

Because this approach does not require a pre-existing feature extractor, it is especially useful in scenarios where none exists – such as for *de novo* modalities where reliable feature extractors have yet to be developed. This occurs when a balanced and diverse dataset like ImageNet is unavailable or when no self-supervised models have been trained (or publicly released) on the specific modality.

Algorithm 1 relies on a pruning strategy $\mathcal{S}(\cdot, \cdot, \cdot, \cdot)$, which, given a (self-supervised) trained feature extractor θ, prunes the dataset $\mathcal{D}$ to retain only a fraction ρ of the original datapoints. The pruning strategy typically consists of two steps: a clustering phase followed by a sampling phase. The sampling parameter η controls the evolution of the sampling strategy during training, allowing for dynamic adjustments that may enhance learning. In practice, we experimented with a modified version of [55, Algorithm 1], as described in Algorithm 2. This clustering-based pruning approach was designed to rebalance large training datasets in self-supervised learning, ensuring that the retained subset remains diverse and representative throughout training.

Algorithm 2 *Modified* Hierarchical k-Means with Resampling Algorithm [55]

Require: Dataset $\mathcal{D} \in \mathbb{R}^{n \times d}$, Size of dataset subset n_c, Number of levels L, Clusters per level $(k_i)_{1 \leq i \leq L}$, Sampling diversity $\eta \in [0, 1]$, Reduction ratio ρ, Hierarchical clustering algorithm H-KMEANS [55, Algorithm 1].
Ensure: Balanced dataset $\mathcal{D}_\rho$ where $\left\lceil \frac{|\mathcal{D}_\rho|}{|\mathcal{D}|} \right\rceil = \rho$
 1: $I_{n_\mathrm{c}} \leftarrow \mathrm{RandomSubset}(\{1, \ldots, n\}, n_\mathrm{c})$ ▷ Randomly sample a subset of indices
 2: $\mathcal{D}_{\mathrm{small}} \leftarrow \mathcal{D}[I_{n_\mathrm{c}}]$ ▷ Extract subset from dataset
 3: $(C_t)_{1 \leq t \leq L} \leftarrow \mathrm{H\text{-}KMEANS}(\mathcal{D}_{\mathrm{small}}, (k_i)_{1 \leq i \leq L})$ ▷ Perform hierarchical clustering
 4: $L_T \leftarrow \mathrm{ASSIGN}(\mathcal{D}, C_L)$ ▷ Assign clusters based on final centroids C_L
 5: $\mathcal{D}_\rho \leftarrow \mathrm{RESAMPLE}(L_T, \rho, \eta)$ ▷ Resample data to achieve desired reduction

---

Algorithm 2 differs from [55, Algorithm 1] in that it does not perform clustering on the entire dataset. Instead, our primary motivation for pruning is to significantly reduce dataset redundancy rather than to capture its underlying concepts. This approach ensures that dynamic dataset

sampling incurs no additional computational burden on the GPUs, as the clustering process is lightweight enough to be offloaded to the CPU during training. If pruning is scheduled for epoch e, we store the embeddings of the first N data batches, run step 3 of Algorithm 2 on the CPU, and progressively run steps 4 and 5 as the remaining data batches within epoch e are processed by the GPU.

A second difference is the introduction of the sampling diversity parameter $\eta \in [0, 1]$, which controls how datapoints are selected within each cluster. Specifically, η determines whether sampling favors points near the centroid (*i.e.*, more representative) or those on the cluster boundary (*i.e.*, less representative). As previously observed [55], empirical results suggest that sampling closer to the centroids performs best when using a single pruned dataset, as this promotes balance and well-discriminated concepts.

3.4. Experiments

We run experiments using a ResNet50, a ViT-S/16, a ViT-S/8 and a ViT-B/8[14]. In each case, we pre-train the model on our WV dataset without any dataset pruning, with pruning that discards 50% of the dataset, and with pruning that discards 75% of the dataset. We therefore conduct a total of 12 runs. For each run, we train the model for 200 epochs. For the runs that include pruning, we adjust the number of epochs so that the models are trained on the same number of batches as the ones without pruning (with a 20-epoch warmup before the first pruning, this means a total of 380 epochs for the 50% pruning and 740 epochs for the 75% pruning). All runs are conducted with a batch size of 256, an AdamW optimizer, a cosine scheduler with annealing, and a 0.0001 initial learning rate.

3.5. Downstream Tasks

To quantify the performance of foundation models on Sentinel-1 WV imagery, we use three different labelled datasets, covering two kinds of tasks: classification and regression. All the datasets are randomly split between training (80%) and validation (20%).

We report the performance of available, SOTA models trained on Sentinel-1 dual-polarization imagery, namely CROMA[15], MoCo[59], DeCUR[60], DOFA[64], and SoftCon[61]. Note that the WV-Net[18] model is not publicly available, and therefore we did not include it in this study. For MoCo and DeCUR, we use the model versions available from the torchgeo package [46].

3.5.1. TenGeoP dataset

The TenGeoP dataset [57] consists of 37,553 WV images that were manually annotated by human SAR experts into ten categories. These correspond to ten different geophysical phenomena, of both oceanic and meteorologic nature (pure ocean waves, wind streaks, micro convective cells,

rain cells, biological slicks, sea ice, icebergs, low wind areas, atmospheric fronts and oceanic fronts). For each image, only one category was selected. The dataset covers the entire ocean and is composed of Sentinel-1A WV images acquired in 2016. Sample images are shown in Figure 1 The dataset is available at https://doi.org/10.17882/56796.

3.5.2. Significant Wave Height dataset

We randomly selected 10,000 WV images and labeled them with their Significant Wave Height (H_s), defined as the mean height of the highest third of the waves passing through a point over a given period. The H_s values were obtained from co-located altimeter measurements available in the CMEMS database [12]. WV and altimeter data were considered co-located if acquired within 3 hours of each other and if the altimeter measurement was within $\pm 2°$ of the WV image center. When multiple altimeter measurements met these criteria, we retained the one closest in space. The data will be made publicly available.

3.5.3. Wind speed dataset

SAR images capture information on the sea-surface roughness, which is related to surface wind speed [1, 22]. We used a random subset of 50,000 images among those selected by O'Driscoll et al. [38] and labeled with wind speed using ERA5 reanalysis data [23] as ground truth. The data are available at https://doi.org/10.5281/zenodo.7784019.

3.6. Results

3.6.1. Impact of dataset pruning

Table 1 summarizes the results of our experiments and highlights the impact of data pruning on model performance on downstream tasks. We find that pruning the dataset leads to an overall increase in performance on downstream tasks, that can be very significant (for instance from 2.5 to 3% for the TenGeoP classification task), at no additional computation cost. The increase is smaller, however, for the two regression tasks (on the order of 0.01-0.02 RMSE). A possible reason for this is that DINO may not be an ideal training strategy for such tasks, as it favors a holistic understanding of images and may overlook finer-scale texture details that are important for SWH and wind speed prediction.

Furthermore, validation loss and k-NN probing curves during training suggest that experiments with pruning would benefit from even more training, since model performance does not plateau when reaching our limit of 200 epochs, which is not the case for the no-pruning experiments (Figures 2 and 3). Since images from downstream datasets were excluded from the pre-training dataset, this indicates better and stronger generalizability by the with-pruning models, since they achieve a higher accuracy while being (effectively) trained on a smaller set of data.

Dataset pruning also substantially reduces the computational cost required to reach a given level of model per-
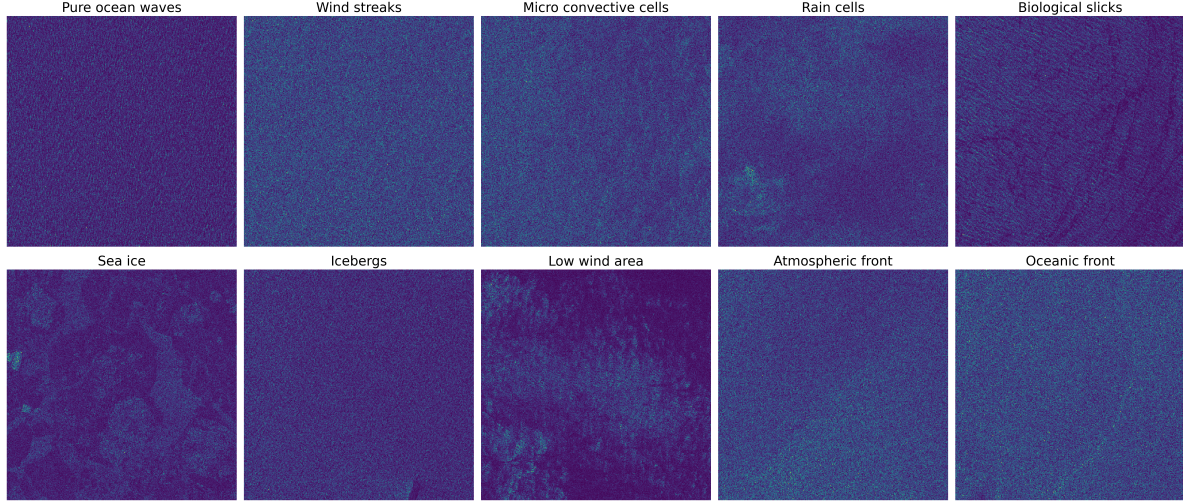
Figure 1. Illustration of the 10 geophysical classes in the TenGeoP dataset.

Model	Pruning	TenGeoP	SWH	Wind speed
ResNet50	0%	72.7	0.72	1.43
ResNet50	50%	73.8	0.71	1.42
ResNet50	75%	75.5	0.70	1.42
ViT-S/16	0%	76.0	0.70	1.40
ViT-S/16	50%	77.1	0.69	1.40
ViT-S/16	75%	78.6	0.69	1.39
ViT-S/8	0%	79.9	0.66	1.40
ViT-S/8	50%	81.6	0.64	1.39
ViT-S/8	75%	82.1	0.64	1.38
ViT-B/8	0%	80.5	0.65	1.38
ViT-B/8	50%	81.9	0.64	1.37
ViT-B/8	75%	83.6	0.63	1.37

Table 1. Summary of k-NN performance on the three downstream tasks across our experiments. For each experiment, we show the highest accuracy reached during training, regardless of epoch. For experiments that involve pruning, this is usually around 200 epochs, but for the no-pruning experiments, the highest accuracy is often reached earlier during training (Figure 3), after which the accuracy plateaus, or decreases slightly.

formance for TenGeoP classification. For instance, the best TenGeoP classification accuracy for a ViT-S/16 is 76% when no pruning is applied during training, and it is reached after about 145 epochs. By contrast, the same accuracy is reached as early as after 85 epochs under the 75% pruning scenario, a more than 40% gain in computational resources.

3.6.2. Comparison to other SAR foundation models

Table 2 compares the performance of existing SAR Sentinel-1 foundation models with ours. The performance tends to vary significantly across different downstream
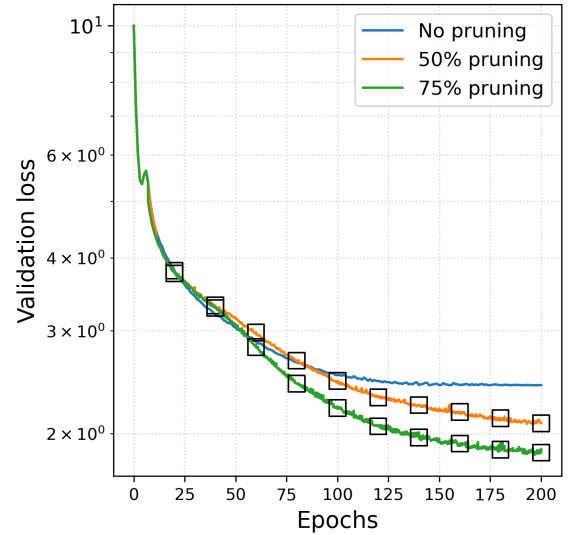

Figure 2. DINO validation loss as a function of training epoch for a ViT-S/16 model, with (blue) no data pruning, (orange) pruning that discards 50% of the training dataset and (green) pruning that discards 75% of the training dataset. For the latter two curves, black squares indicate epochs at which dataset pruning is performed.

tasks, as the relevant information is embedded at different levels within the image. For example, class differences in TenGeoP are strongly linked to global-level features, whereas wind field estimation depends on the finer texture details of the image.

Most self-supervised models developed for SAR imagery have been trained on Ground Range Detected (GRD) images in Interferometric Wide (IW) swath mode, which are available in dual polarization (VV+VH). As a result,
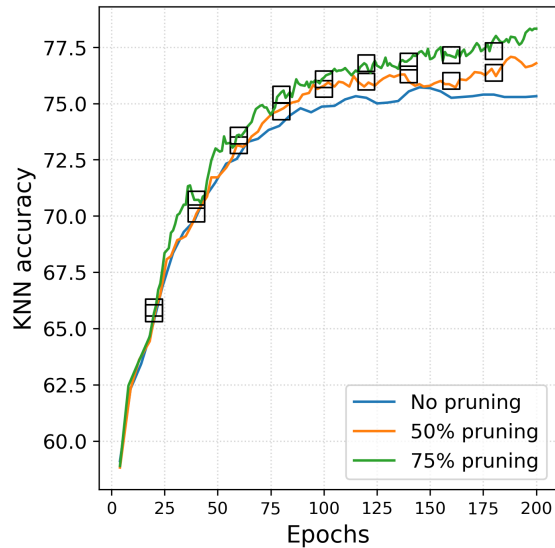
Figure 3. Same as Figure 2, but for the k-NN classification accuracy on TenGeoP.

Method	Backbone	TenGeoP	SWH	Wind
CROMA	ViT-B/8	65.4	0.78	1.95
MoCo	ResNet50	60.9	0.83	1.98
DeCUR	ResNet50	58.3	0.84	2.10
DOFA	ViT-B/16	58.4	0.96	2.40
DOFA	ViT-L/16	63.4	0.95	2.43
SoftCon	ViT-S/14	73.2	0.98	2.08
SoftCon	ViT-B/14	74.8	1.01	2.08
Nereus-SAR-1	ResNet50	75.5	0.75	1.62
Nereus-SAR-1	ViT-S/16	78.6	0.69	1.39
Nereus-SAR-1	ViT-S/8	82.1	0.64	1.38
Nereus-SAR-1	ViT-B/8	**83.6**	**0.63**	**1.37**

Table 2. Comparison of k-NN performance of existing models and ours across the three selected downstream tasks: (1) classification accuracy on TenGeoP, (2) regression RMSE for significant wave height prediction, and (3) regression RMSE for wind speed prediction.

these models expect two-channel images as input. However, WV images are only available in single-polarization (VV). In Table 2, the results for WV images are obtained by simply duplicating the single-channel VV input to match the expected dual-channel format.

For further evaluation, we also perform supervised fine-tuning for existing SAR SSL models. A wide range of different methods were designed to perform Parameter Efficient Fine-Tuning (PEFT) on pre-trained backbones, like visual fine-tuning [28] or adapters [9, 45]. Most methods specifically aim at improving the trade-off between the

computational cost and performance degradation in fine-tuning. Note, however, that most of the proposed fine-tuning methods are supervised: in particular, they do not lead to a fine-tuned module that is transferable across downstream tasks. This goes against the spirit of foundation models, and would severely limit the applicability of such fine-tuning approaches for downstream tasks with little annotated data.

Some recent studies developed PEFT methods specific to geospatial foundation models [27, 42], and started to explore the challenge of self-supervised fine-tuning [29, 41]. Others, by contrast, developed supervised fine-tuning methods to address the specific case of cross-domain adaptation by incorporating domain inductive bias for specific cases, e.g. for multi-spectral [50], thermal [62], or RGB-Depth [24] imagery.

As it is not the scope of this paper to design a radar-specific adaptation module, we proceed as follows. For ResNet backbones like MoCo and DeCUR, we perform full fine-tuning. For SoftCon, the best performing ViT-based model, we perform a classical LoRA fine-tuning on the patch embedding and attention layers. Additionally, in each case, we use a linear layer as the classification head to probe model performance. This head is trained jointly with the backbone adapters, using different learning rates. We then compare these results to our self-supervised model under the same evaluation protocol. It is important to note that this setting places our model at a disadvantage, as only the head parameters are trainable—resulting in significantly fewer parameters available to learn the task-specific objective.

The fine-tuning results are shown in Table 3. Accuracies on TenGeoP are improved by 10-25% for existing foundation models and by 5-10% for our models. Similarly, the RMSE for wave height and wind speed decreases by 0.05-0.2 for existing models and by 0.1-0.3 for our models.

Our specialized foundation models consistently outperform existing foundation models on the three SAR WV tasks, and the gap is especially large for the wave height and wind speed regression (RMSE difference of 0.2-0.6). This can be attributed, in part, to the fact that existing models were trained on different data, which leads to both sensor and domain shifts. Notably, our training dataset consists exclusively of oceanic SAR SLC (single-look complex) images, whereas existing models have been predominantly trained on SAR GRD (ground range detected) imagery over land. Future work could investigate the relative impact of sensor versus domain shift in driving this performance gap.

The smaller performance gaps between models on the TenGeoP benchmark dataset suggest that TenGeoP may be relatively simplistic as a benchmark and may lack suffi-

cient discriminative power to effectively evaluate the performance of self-supervised learning models for SAR imagery. Notably, even with a simple linear probing head—i.e., solving a convex problem—we achieve results that equal the current state-of-the-art in supervised learning on this task [58].

For our models, we find that the best-performing one on TenGeoP is the ViT-S/16, closely followed by ViT-B/8. However, for the wave height and wind speed regression tasks, larger models with smaller patch sizes perform better. This likely results from TenGeoP classes being primarily characterized by global image features, and being therefore less sensitive to small-scale structures that are better captured by models with smaller patch sizes. By contrast, the other two tasks rely more heavily on local texture information to accurately estimate geophysical parameters.

Method	Backbone	TenGeoP	SWH	Wind
MoCo-FT	ResNet50	86.5	0.77	1.80
DeCUR-FT	ResNet50	81.9	0.82	1.93
Softcon-LoRA	ViT-S/14	84.2	0.78	1.98
Softcon-LoRA	ViT-B/14	85.1	0.79	1.95
Nereus-SAR-1	ResNet50	80.9	0.63	1.33
Nereus-SAR-1	ViT-S/16	**89.0**	0.57	1.34
Nereus-SAR-1	ViT-S/8	85.8	0.55	1.30
Nereus-SAR-1	ViT-B/8	88.3	**0.54**	**1.29**

Table 3. Linear probing accuracy of existing models (adapted to the single WV polarization or fine-tuned with LoRA) and ours on the three selected downstream tasks.

4. Discussion and Conclusion

In this study, we introduced a dynamic pruning methodology for self-supervised training of an EO foundation model on Sentinel-1 WV imagery. Given the high redundancy inherent in the dataset, our pruning strategy significantly accelerates convergence.

Beyond efficiency gains, we find that dynamic pruning also improves downstream performance. We attribute this to the regularizing effect of resampling from a redundant dataset, which helps avoid overfitting. From an optimization standpoint, this insight suggests promising connections to classical strategies—such as restart-based methods—that have seen limited exploration in this setting [39, 40].

We also address a gap in existing EO foundation models by focusing on a modality—Sentinel-1 WV—for which no pretrained model is well-suited. To support further research, we release a benchmark framework tailored to this modality, relevant for applications ranging from wind field

estimation to oil slick detection. We also release the weights for *Nereus-SAR-1*, the first in a family of models designed specifically for ocean observation using SAR imagery.

Finally, our findings emphasize the complexity of satellite data, where varying acquisition modes result in divergent data distributions. We show that models trained across multiple modes do not necessarily generalize to unseen ones. In fact, for specialized modalities, unimodal self-supervised models can outperform multimodal counterparts while requiring significantly fewer computational resources—further amplified by our pruning strategy.

5. Acknowledgements

6. Data and Code

The Nereus-SAR-1 model weights are available on HuggingFace at https://huggingface.co/galeio-research/nereus-sar-1. We also release the probing weights for the TenGeoP classification task and the wave height and wind speed regression tasks.

References

[1] T. Ahsbahs, G. Maclaurin, C. Draxl, C. R. Jackson, F. Monaldo, and M. Badger. Us east coast synthetic aperture radar wind atlas for offshore wind energy. *Wind Energy Science*, 5 (3):1191–1210, 2020. 5

[2] Alon Albalak, Yanai Elazar, Sang Michael Xie, Shayne Longpre, Nathan Lambert, Xinyi Wang, Niklas Muennighoff, Bairu Hou, Liangming Pan, Haewon Jeong, et al. A survey on data selection for language models. *arXiv preprint arXiv:2402.16827*, 2024. 2

[3] Randall Balestriero, Mark Ibrahim, Vlad Sobal, Ari Morcos, Shashank Shekhar, Tom Goldstein, Florian Bordes, Adrien Bardes, Gregoire Mialon, Yuandong Tian, Avi Schwarzschild, Andrew Gordon Wilson, Jonas Geiping, Quentin Garrido, Pierre Fernandez, Amir Bar, Hamed Pirsiavash, Yann LeCun, and Micah Goldblum. A Cookbook of Self-Supervised Learning. *arXiv preprint arXiv:2304.12210*, 2023. 1

[4] Favyen Bastani, Piper Wolters, Ritwik Gupta, Joe Ferdinando, and Aniruddha Kembhavi. SatlasPretrain: A Large-Scale Dataset for Remote Sensing Image Understanding. *arXiv preprint arXiv:2211.15660*, 2023. 1, 3

[5] Vighnesh Birodkar, Hossein Mobahi, and Samy Bengio. Semantic redundancies in image-classification datasets: The 10% you don't need. *arXiv preprint arXiv:1901.11409*, 2019. 2

[6] Nassim Ait Ali Braham, Conrad M Albrecht, Julien Mairal, Jocelyn Chanussot, Yi Wang, and Xiao Xiang Zhu. SpectralEarth: Training Hyperspectral Foundation Models at Scale. *arXiv preprint arXiv:2408.08447*, 2024. 1

[7] Liangliang Cao, Bowen Zhang, Chen Chen, Yinfei Yang, Xianzhi Du, Wencong Zhang, Zhiyun Lu, and Yantao Zheng. Less is more: Removing text-regions improves clip training efficiency and robustness, 2023. 2

[8] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9650–9660, 2021. 3

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

[10] Gordon Christie, Neil Fendley, James Wilson, and Ryan Mukherjee. Functional Map of the World, 2018. 1

[11] Guanyi Chu, Xiao Wang, Chuan Shi, and Xunqiang Jiang. CuCo: Graph Representation with Curriculum Contrastive Learning. In *IJCAI*, pages 2300–2306, 2021. 2

[12] CMEMS. Copernicus Marine Service (CMEMS) In Situ TAC. http://marineinsitu.eu/dashboard/, 2024. Accessed: 2023-11-06. 5

[13] Yikang Ding, Zhenyang Li, Dihe Huang, Zhiheng Li, and Kai Zhang. Enhancing multi-view stereo with contrastive matching and weighted focal loss. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 821–825. IEEE, 2022. 2

[14] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 5

[15] Anthony Fuller, Koreen Millard, and James Green. Croma: Remote sensing representations with contrastive radar-optical masked autoencoders. *Advances in Neural Information Processing Systems*, 36:5506–5538, 2023. 2, 5

[16] Jakob Gawlikowski and Nina Maria Gottschling. On the relevance of SAR and optical modalities in deep learning-based data fusion. In *ICLR 2024 Machine Learning for Remote Sensing (ML4RS) Workshop*, 2024. 2

[17] Jonas Geiping and Tom Goldstein. Cramming: Training a Language Model on a single GPU in one day. In *International Conference on Machine Learning*, pages 11117–11143. PMLR, 2023. 2

[18] Yannik Glaser, Justin E Stopa, Linnea M Wolniewicz, Ralph Foster, Doug Vandemark, Alexis Mouche, Bertrand Chapron, and Peter Sadowski. WV-Net: A foundation model for SAR WV-mode satellite imagery trained using contrastive self-supervised learning on 10 million images. *arXiv preprint arXiv:2406.18765*, 2024. 3, 5

[19] Priya Goyal, Mathilde Caron, Benjamin Lefaudeux, Min Xu, Pengchao Wang, Vivek Pai, Mannat Singh, Vitaliy Liptchinsky, Ishan Misra, Armand Joulin, et al. Self-supervised pretraining of visual features in the wild. *arXiv preprint arXiv:2103.01988*, 2021. 1, 3

[20] Han Guo, Sai Ashish Somayajula, Ramtin Hosseini, and Pengtao Xie. Improving image classification of gastrointestinal endoscopy using curriculum self-supervised learning. *Scientific Reports*, 14(1):6100, 2024. 2

[21] Xin Guo, Jiangwei Lao, Bo Dang, Yingying Zhang, Lei Yu, Lixiang Ru, Liheng Zhong, Ziyuan Huang, Kang Wu, Dingxiang Hu, Huimei He, Jian Wang, Jingdong Chen, Ming Yang, Yongjun Zhang, and Yansheng Li. SkySense: A Multi-Modal Remote Sensing Foundation Model Towards Universal Interpretation for Earth Observation Imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 27672–27683, 2024. 1

[22] Danièle Hauser, Saleh Abdalla, Fabrice Ardhuin, Jean-Raymond Bidlot, Mark Bourassa, David Cotton, Christine Gommenginger, Hayley Evers-King, Harald Johnsen, John Knaff, Samantha Lavender, Alexis Mouche, Nicolas Reul, Charles Sampson, Edward C.C Steele, and Ad Stoffelen. Satellite Remote Sensing of Surface Winds, Waves, and Currents: Where are we Now? *Surveys in Geophysics*, 44:1357–1446, 2023. 5

[23] Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, Adrian Simmons, Cornel Soci, Saleh Abdalla, Xavier Abellan, Gianpaolo Balsamo, Peter Bechtold, Gionata Biavati, Jean Bidlot, Massimo Bonavita, Giovanna De Chiara, Per Dahlgren, Dick Dee, Michail Diamantakis, Rossana Dragani, Johannes Flemming, Richard Forbes, Manuel Fuentes, Alan Geer, Leo Haimberger, Sean Healy, Robin J. Hogan, Elías Hólm, Marta

Janisková, Sarah Keeley, Patrick Laloyaux, Philippe Lopez, Cristina Lupu, Gabor Radnoti, Patricia de Rosnay, Iryna Rozum, Freja Vamborg, Sebastien Villaume, and Jean-Noël Thépaut. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020. 5

[24] Judy Hoffman, Saurabh Gupta, Jian Leong, Sergio Guadarrama, and Trevor Darrell. Cross-modal adaptation for rgb-d detection. In *2016 IEEE international conference on robotics and automation (ICRA)*, pages 5032–5039. IEEE, 2016. 7

[25] Danfeng Hong, Bing Zhang, Xuyang Li, Yuxuan Li, Chenyu Li, Jing Yao, Naoto Yokoya, Hao Li, Pedram Ghamisi, Xiuping Jia, Antonio Plaza, Paolo Gamba, Jon Atli Benediktsson, and Jocelyn Chanussot. SpectralGPT: Spectral Remote Sensing Foundation Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(8):5227–5244, 2024. 1

[26] Pengyue Hou and Xingyu Li. Improving Contrastive Learning of Sentence Embeddings with Focal-InfoNCE. *arXiv preprint arXiv:2310.06918*, 2023. 2

[27] Leiyi Hu, Wanxuan Lu, Hongfeng Yu, Dongshuo Yin, Xian Sun, and Kun Fu. Tea: A training-efficient adapting framework for tuning foundation models in remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 2024. 7

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

[29] Samar Khanna, Medhanie Irgau, David B Lobell, and Stefano Ermon. Explora: Parameter-efficient extended pretraining to adapt vision transformers under domain shifts. 7

[30] Weijie Li, Wei Yang, Yuenan Hou, Li Liu, Yongxiang Liu, and Xiang Li. SARATR-X: Toward Building a Foundation Model for SAR Target Recognition. *IEEE Transactions on Image Processing*, 34:869–884, 2025. 1

[31] Ziwei Liu, Zhongqi Miao, Xiaohang Zhan, Jiayun Wang, Boqing Gong, and Stella X. Yu. Large-Scale Long-Tailed Recognition in an Open World. *arXiv preprint arXiv:1904.05160*, 2019. 1

[32] Siqi Lu, Junlin Guo, James R Zimmer-Dauphinee, Jordan M Nieusma, Xiao Wang, Parker VanValkenburgh, Steven A Wernke, and Yuankai Huo. Vision Foundation Models in Remote Sensing: A Survey. *arXiv preprint arXiv:2408.03464*, 2025. 1

[33] 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: A Global and Inclusive Benchmark for Geospatial Foundation Models. *arXiv preprint arXiv:2412.04204*, 2024. 1

[34] Baharan Mirzasoleiman, Jeff Bilmes, and Jure Leskovec. Coresets for data-efficient training of machine learning models. In *International Conference on Machine Learning*, pages 6950–6960. PMLR, 2020. 2

[35] Humza Naveed, Asad Ullah Khan, Shi Qiu, Muhammad Saqib, Saeed Anwar, Muhammad Usman, Naveed

Akhtar, Nick Barnes, and Ajmal Mian. A Comprehensive Overview of Large Language Models. *arXiv preprint arXiv:2307.06435*, 2024. 1

[36] Thao Nguyen, Gabriel Ilharco, Mitchell Wortsman, Sewoong Oh, and Ludwig Schmidt. Quality not quantity: On the interaction between dataset design and robustness of clip. *Advances in Neural Information Processing Systems*, 35:21455–21469, 2022. 2

[37] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023. 1

[38] Owen O'Driscoll, Alexis Mouche, Bertrand Chapron, Marcel Kleinherenbrink, and Paco López-Dekker. Obukhov Length Estimation From Spaceborne Radars. *Geophysical Research Letters*, 50(15):e2023GL104228, 2023. 5

[39] Sebastian Pokutta. Restarting algorithms: Sometimes there is free lunch. In *International Conference on Integration of Constraint Programming, Artificial Intelligence, and Operations Research*, pages 22–38. Springer, 2020. 8

[40] James Renegar and Benjamin Grimmer. A simple nearly optimal restart scheme for speeding up first-order methods. *Foundations of computational mathematics*, 22(1):211–256, 2022. 8

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

[42] Karthick Panner Selvam, Raul Ramos-Pollan, and Freddie Kalaitzis. Rapid adaptation of earth observation foundation models for segmentation. *arXiv preprint arXiv:2409.09907*, 2024. 7

[43] Ozan Sener and Silvio Savarese. Active learning for convolutional neural networks: A core-set approach, 2018. 2

[44] Ben Sorscher, Robert Geirhos, Shashank Shekhar, Surya Ganguli, and Ari Morcos. Beyond neural scaling laws: beating power law scaling via data pruning. *Advances in Neural Information Processing Systems*, 35:19523–19536, 2022. 2

[45] Jan-Martin O Steitz and Stefan Roth. Adapters strike back. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23449–23459, 2024. 7

[46] Adam J. Stewart, Caleb Robinson, Isaac A. Corley, Anthony Ortiz, Juan M. Lavista Ferres, and Arindam Banerjee. TorchGeo: Deep learning with geospatial data. In *Proceedings of the 30th International Conference on Advances in Geographic Information Systems*, pages 1–12, Seattle, Washington, 2022. Association for Computing Machinery. 5

[47] Gencer Sumbul, Marcela Charfuelan, Begum Demir, and Volker Markl. BigEarthNet: A Large-Scale Benchmark Archive for Remote Sensing Image Understanding. In *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2019. 1

[48] Gencer Sumbul, Jian Kang, Tristan Kreuziger, Filipe Marcelino, Hugo Costa, Pedro Benevides, Mario Caetano,

and Begüm Demir. BigEarthNet Dataset with A New Class-Nomenclature for Remote Sensing Image Understanding. *arXiv preprint arXiv:2001.06372*, 2021. 1

[49] Daniela Szwarcman, Sujit Roy, Paolo Fraccaro, Þorsteinn Elí Gíslason, Benedikt Blumenstiel, Rinki Ghosal, Pedro Henrique de Oliveira, Joao Lucas de Sousa Almeida, Rocco Sedona, Yanghui Kang, Srija Chakraborty, Sizhe Wang, Carlos Gomes, Ankur Kumar, Myscon Truong, Denys Godwin, Hyunho Lee, Chia-Yu Hsu, Ata Akbari Asanjan, Besart Mujeci, Disha Shidham, Trevor Keenan, Paulo Arevalo, Wenwen Li, Hamed Alemohammad, Pontus Olofsson, Christopher Hain, Robert Kennedy, Bianca Zadrozny, David Bell, Gabriele Cavallaro, Campbell Watson, Manil Maskey, Rahul Ramachandran, and Juan Bernabe Moreno. Prithvi-EO-2.0: A Versatile Multi-Temporal Foundation Model for Earth Observation Applications. *arXiv preprint arXiv:2412.02732*, 2025. 1

[50] Romain Thoreau, Valerio Marsocci, and Dawa Derksen. Parameter-efficient adaptation of geospatial foundation models through embedding deflection, 2025. 7

[51] Yonglong Tian, Olivier J. Henaff, and Aaron van den Oord. Divide and Contrast: Self-supervised Learning from Uncurated Data. *arXiv preprint arXiv:2105.08054*, 2021. 1

[52] Kushal Tirumala, Daniel Simig, Armen Aghajanyan, and Ari Morcos. D4: Improving llm pretraining via document deduplication and diversification. *Advances in Neural Information Processing Systems*, 36, 2024. 2

[53] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. LLaMA: Open and Efficient Foundation Language Models. *arXiv preprint arXiv:2302.13971*, 2023. 1

[54] Valentino Vito and Lim Yohanes Stefanus. An asymmetric contrastive loss for handling imbalanced datasets. *Entropy*, 24(9):1303, 2022. 2

[55] Huy V. Vo, Vasil Khalidov, Timothée Darcet, Théo Moutakanni, Nikita Smetanin, Marc Szafraniec, Hugo Touvron, Camille Couprie, Maxime Oquab, Armand Joulin, Hervé Jégou, Patrick Labatut, and Piotr Bojanowski. Automatic data curation for self-supervised learning: A clustering-based approach, 2024. 1, 2, 4, 5

[56] Alex Jinpeng Wang, Kevin Qinghong Lin, David Junhao Zhang, Stan Weixian Lei, and Mike Zheng Shou. Too large; data reduction for vision-language pre-training. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3147–3157, 2023. 2

[57] Chen Wang, Alexis Mouche, Pierre Tandeo, Justin E Stopa, Nicolas Longépé, Guillaume Erhard, Ralph C Foster, Douglas Vandemark, and Bertrand Chapron. A labelled ocean sar imagery dataset of ten geophysical phenomena from sentinel-1 wave mode. *Geoscience Data Journal*, 6(2):105–115, 2019. 5

[58] Chen Wang, Pierre Tandeo, Alexis Mouche, Justin E. Stopa, Victor Gressani, Nicolas Longepe, Douglas Vandemark, Ralph C. Foster, and Bertrand Chapron. Classification of the global Sentinel-1 SAR vignettes for ocean surface pro-

cess studies. *Remote Sensing of Environment*, 234:111457, 2019. 8

[59] Yi Wang, Nassim Ait Ali Braham, Zhitong Xiong, Chenying Liu, Conrad M Albrecht, and Xiao Xiang Zhu. Ssl4eos12: A large-scale multimodal, multitemporal dataset for self-supervised learning in earth observation [software and data sets]. *IEEE Geoscience and Remote Sensing Magazine*, 11(3):98–106, 2023. 1, 2, 3, 5

[60] Yi Wang, Conrad M Albrecht, Nassim Ait Ali Braham, Chenying Liu, Zhitong Xiong, and Xiao Xiang Zhu. Decoupling common and unique representations for multimodal self-supervised learning. In *European Conference on Computer Vision*, pages 286–303. Springer, 2024. 2, 5

[61] Yi Wang, Conrad M Albrecht, and Xiao Xiang Zhu. Multilabel guided soft contrastive learning for efficient earth observation pretraining. *IEEE Transactions on Geoscience and Remote Sensing*, 2024. 3, 5

[62] Zeyu Wang, Shuaiting Li, and Kejie Huang. Cross-modal adaptation for object detection in thermal infrared remote sensing imagery. *IEEE Geoscience and Remote Sensing Letters*, 2025. 7

[63] Xiaobo Xia, Jiale Liu, Jun Yu, Xu Shen, Bo Han, and Tongliang Liu. Moderate Coreset: A Universal Method of Data Selection for Real-world Data-efficient Deep Learning. In *The Eleventh International Conference on Learning Representations*, 2023. 2

[64] 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. 3, 5

[65] Shuo Yang, Zhe Cao, Sheng Guo, Ruiheng Zhang, Ping Luo, Shengping Zhang, and Liqiang Nie. Mind the Boundary: Coreset Selection via Reconstructing the Decision Boundary. In *Forty-first International Conference on Machine Learning*, 2024. 2

[66] Seonghyeon Ye, Jiseon Kim, and Alice Oh. Efficient contrastive learning via novel data augmentation and curriculum learning. *arXiv preprint arXiv:2109.05941*, 2021. 2

[67] Haizhong Zheng, Rui Liu, Fan Lai, and Atul Prakash. Coverage-centric coreset selection for high pruning rates, 2023. 2