Parameter Efficient Self-Supervised Geospatial Domain Adaptation

Supplementary Material

7. Dataset Details

EuroSAT [15] is a land-cover classification dataset with 27k labeled 64×64 image patches from the Sentinel-2 satellite at 10m resolution. We utilize the train/test/validation split from [26] with the RGB bands of the dataset. **RESISC45** [6] contains 31.5k remote sensing RGB images with $0.3\text{-}30\text{m}$ resolution from 45 different classes. The image size is 256×256 pixels. We use the dataset split from [26]. **FireRisk** [33] is a dataset for classification of remote sensing imagery into 7 different fire-risk categories. The dataset contains 91k aerial RGB images with 1m resolution and 320×320 pixels. We utilize the train dataset from `torchgeo` and create validation and test datasets by randomly splitting the existing validation dataset into half. **TreeSatAI** [1] consists of aerial, Sentinel-2 and Sentinel-1 imagery of forest areas with 20 tree-type labels. We utilize the RGB+NIR aerial imagery at 0.2m resolution with 304×304 pixels (50k samples). We use the test dataset provided by the authors and randomly split a validation dataset (5k samples) from the training dataset. **EuroSAT-SAR** [43] contains 27k dual-pol Sentinel-1 images with VV and VH channels and land-cover labels. The images are 64×64 pixels and we arrange the channels as `[VV,VH,VV]` to utilize pre-trained patch-embedding layers. We use the same train/test/validation splits as for EuroSAT. **BENGE-8k** [25] is a subset of the BENGE dataset with 8k samples, based on the BigEarthNet [36, 37] dataset. In our experiments, we utilize the Sentinel-1 SAR image patches and arrange the VV and VH channels as `[VV,VH,VV/VH]` with land-cover segmentation masks as target. **UCMerced** [46] is a land-use classification dataset with 2k RGB samples in 21 classes and 30cm resolution. The images have 256×256 pixels and we utilize the dataset splits from [26]. See Fig. 7 for sample images from all datasets.

8. Training Setup

For self-supervised pre-training of the SLR adapters we use a default learning rate of $1 \cdot 10^{-4}$. The learning rate is reduced by a factor of 0.1 if the validation loss did not improve in 5 consecutive epochs. We train with a batch size of 32 for a maximum of 25 000 steps (15 000 for EuroSAT and BENGE-8k). For fine-tuning, the base learning rates for the backbone (or SLR adapters) and model head are $1 \cdot 10^{-4}$ and $1 \cdot 10^{-5}$, respectively. We use the same learning rate schedule as for pre-training. In experiments with SLR or LoRA adapters, the dimensionality of the hidden layer is set to $r = 8$ for EuroSAT and BENGE-8k and $r = 16$ for RESISC45, UCMerced, FireRisk, TreeSatAI and EuroSAT-SAR. We conduct all experiments with single NVIDIA V100 GPUs.

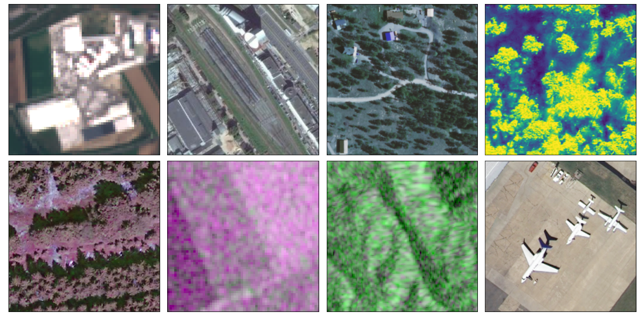

Figure 7. Example images from the remote sensing datasets with different data modalities used in this work. From top left to bottom right: EuroSAT-RGB, UCMerced, FireRisk, TreeSatAI-NIR, TreeSatAI-RGB, EuroSATSAR (`VV,VH,VV`), BENGE-8k-SAR (`VV,VH,VV/VH`) and RESISC45.

9. Continual Pre-training

We conduct continual pre-training on the EuroSAT-SAR and RESISC45 datasets to derive an upper bound on the possible performance gain with in-domain pre-training. In this experiment, the MAE model (pre-trained on ImageNet) is continually trained with masked autoencoding on the target dataset without supervision. Then, the resulting model is trained for the target task using supervised learning. This result in linear evaluation and fine-tuning accuracies of 88.68% and 88.68%, respectively, on EuroSAT-SAR (see Tab. 6). For RESISC45, full continual pre-training reaches 91.29% linear evaluation accuracy and 95.37% fine-tuning accuracy.

Method	EuroSAT-SAR	RESISC45
Lin. Eval	77.95	77.90
SLR Lin. Eval	84.22	87.08
CPT Lin. Eval	88.17	91.29
Fine-tuned	86.46	95.16
SLR FT	87.00	93.84
CPT FT	88.68	95.37

Table 6. Fine-tuning and linear evaluation accuracy on EuroSAT-SAR and RESISC45 datasets with the ImageNet MAE model. Comparison of standard linear evaluation/fine-tuning with **SLR** adapter training and self-supervised continual pre-training (**CPT**) of the full model on the target dataset.