Supplementary Material for Seasonal Contrast: Unsupervised Pre-Training from Uncurated Remote Sensing Data

Oscar Mañas[1,2] Alexandre Lacoste[1] Xavier Giró-i-Nieto[2] David Vazquez[1] Pau Rodriguez[1]
[1]Element AI [2]Universitat Politècnica de Catalunya
oscmansan@gmail.com, pau.rodriguez@servicenow.com

A. Appendix

A.1. Comparison to Similar Published Works

We compare our method to other published results [2, 4] that use the same benchmarks [3, 1]. Since the training procedures differ significantly (*e.g.*, in the amount of supervision or model capacity), in Tables 1 and 2 we clearly specify the differences among methods. *MS* means multi-spectral data, *RN18* denotes ResNet-18, and *RN50* denotes ResNet-50.

Method	Pretrain	Labels	Input	Arch.	mAP
Neumann et al. [2]	sup.	old	RGB	RN50	69.70
SeCo (ours)	unsup.	old	RGB	RN18	**85.67**
Vincenzi et al. [4]	unsup.	new	MS	RN18	86.00
SeCo (ours)	unsup.	new	RGB	RN18	**87.27**

Table 1. Comparison of fine-tuning performance on BigEarthNet.

Method	Pretrain	Arch.	Acc.
Neumann et al. [2]	sup.	RN50	**99.20**
SeCo (ours)	unsup.	RN18	96.60

Table 2. Comparison of fine-tuning performance on EuroSAT.

We were surprised to find that SeCo achieves 16% higher mAP than Neumann et al. [2] when using the old BigEarthNet label set despite pre-training in an unsupervised way and using a smaller backbone.

A.2. Additional Ablation on the Locations Sampling

In order to broaden our analysis of different location sampling strategies, we restrict the collection of images to only Europe by filtering our previous dataset of ~1M images. After filtering, we are left with ~83K images sampled around European cities. In Table 3, we provide results on the BigEarthNet benchmark when pre-training our method with this new dataset and our previous dataset of 100k images sampled worldwide.

Note that with the new setup, pre-training and fine-tuning images come from the same continent. This explains the slight increase in performance when pre-training with the European subset. However, SeCo aims to provide a good remote sensing representation regardless of the downstream task location.

Sampling	Size	Linear probing		Fine-tuning	
		10%	100%	10%	100%
Worldwide	100k	74.67	75.52	81.49	87.04
Europe	83K	**75.49**	**76.39**	**82.68**	**87.61**

Table 3. Comparison of mAP on BigEarthNet with different SeCo dataset sampling strategies. We use a ResNet-18 backbone.

A.3. Ablation on the Embedding Sub-spaces

We further analyze the individual contribution of each embedding sub-space $\mathcal{Z}_i$ on a pre-trained SeCo model. As a reminder, $\mathcal{V}$ is the common embedding space, $\mathcal{Z}_0$ is the embedding sub-space invariant to all augmentations, $\mathcal{Z}_1$ is invariant to seasonal augmentations but variant to artificial augmentations, and $\mathcal{Z}_2$ is invariant to artificial augmentations but variant to seasonal augmentations. In Table 4, we provide results when using the representations of each sub-space as initialization on the BigEarthNet benchmark.

Repr.	Linear probing		Fine-tuning	
	10%	100%	10%	100%
$\mathcal{Z}_0$	71.41	71.79	81.91	87.16
$\mathcal{Z}_1$	72.89	73.30	82.00	87.11
$\mathcal{Z}_2$	71.78	72.23	**82.11**	87.12
$\mathcal{V}$	**76.05**	**77.00**	81.86	**87.27**

Table 4. Comparison of mAP on BigEarthNet when projecting the "general" representation with each of the 3 sub-space heads. We use a ResNet-18 backbone.

Note that using the representation of a sub-space adds one more learnable linear layer during fine-tuning. We can observe that SeCo learns the best combination/weighting of sub-spaces and distills the information in the general embedding space, which allows the representation to be agnostic to downstream tasks.

References

[1] P. Helber, B. Bischke, A. Dengel, and D. Borth. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2019. 1

[2] M. Neumann, A. S. Pinto, X. Zhai, and N. Houlsby. In-domain representation learning for remote sensing. *arXiv preprint arXiv:1911.06721*, 2019. 1

[3] G. Sumbul, M. Charfuelan, B. Demir, and V. Markl. Bigearthnet: A large-scale benchmark archive for remote sensing image understanding. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 5901–5904. IEEE, 2019. 1

[4] S. Vincenzi, A. Porrello, P. Buzzega, M. Cipriano, P. Fronte, R. Cuccu, C. Ippoliti, A. Conte, and S. Calderara. The color out of space: learning self-supervised representations for earth observation imagery. *arXiv preprint arXiv:2006.12119*, 2020. 1