From Natural Images to Spaceborne Imagery: An Empirical Study of Self-supervised Learning for Earth Observation

Yi Wang[1,2], Nassim Ait Ali Braham[1,2], Conrad M Albrecht[1,2], Lichao Mou[1,2], and Xiao Xiang Zhu[1,2]

{Yi.Wang, Nassim.AitAliBraham, Conrad.Albrecht, Lichao.Mou, Xiaoxiang.Zhu}@dlr.de

1: *Remote Sensing Technology Institute, German Aerospace Center, Germany*
2: *Data Science in Earth Observation, Technical University of Munich, Germany*

Introduction & Motivation

- **SELF-SUPERVISED LEARNING (SSL)** has raised wide interest in the remote sensing community with the advantage of learning generic representations from large-scale unlabeled data.

- Despite success in natural images, most of the potential of SSL in earth observation remains unlocked.

- Towards pushing SSL forwards in earth observation data science, we provide an empirical study of 4 modern SSL methods on 4 popular remote sensing datasets.
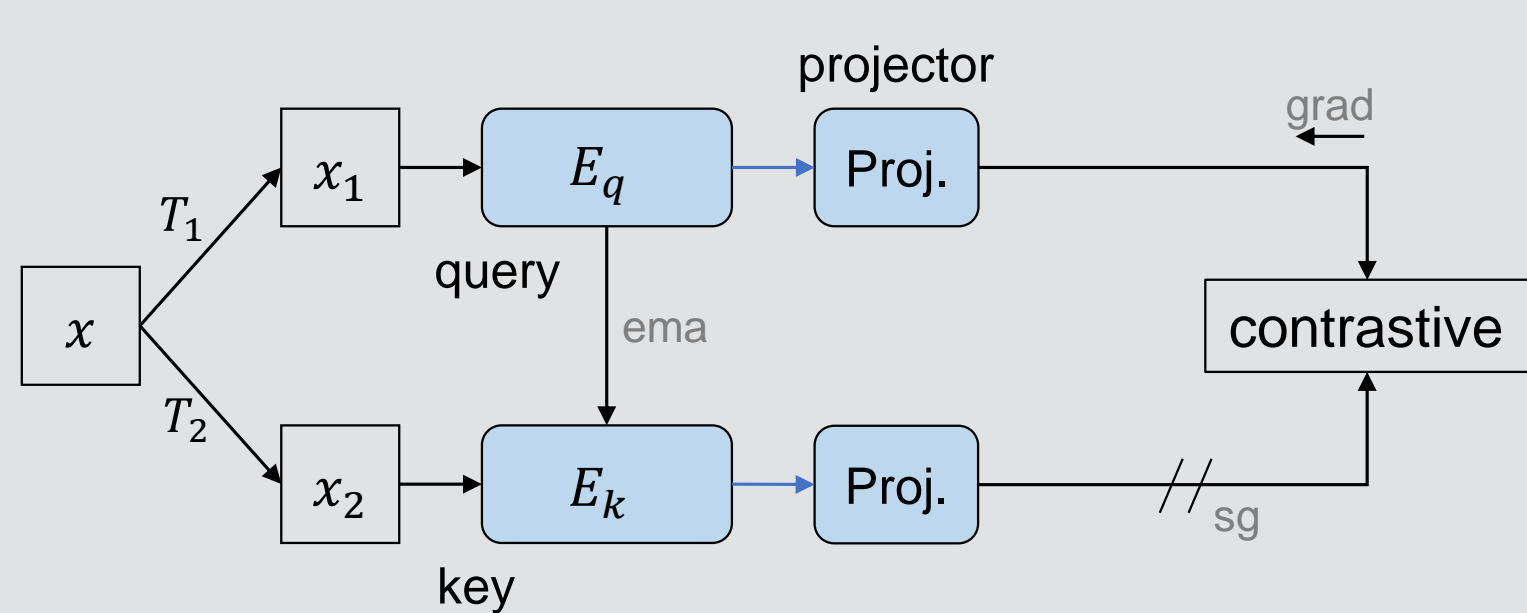
The Pipeline of Self-supervised Learning

- **Self-supervised pre-training**: the model learns generic data representations.
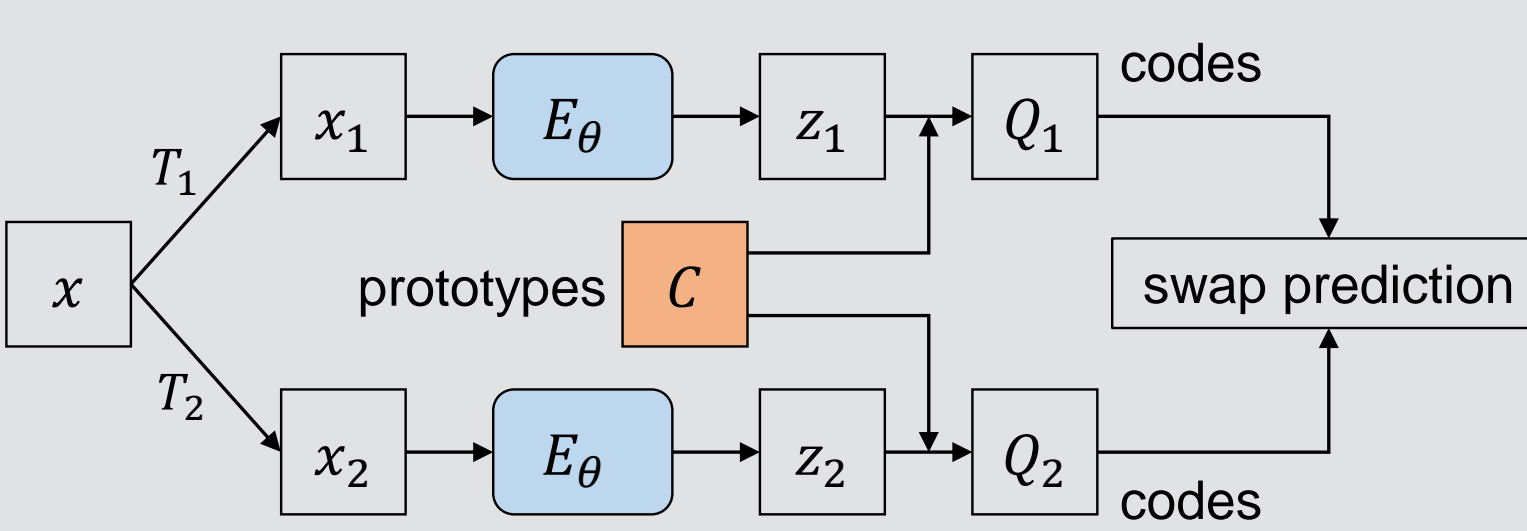- **Supervised downstream tasks**: the pre-trained model gets transferred to downstream applications.
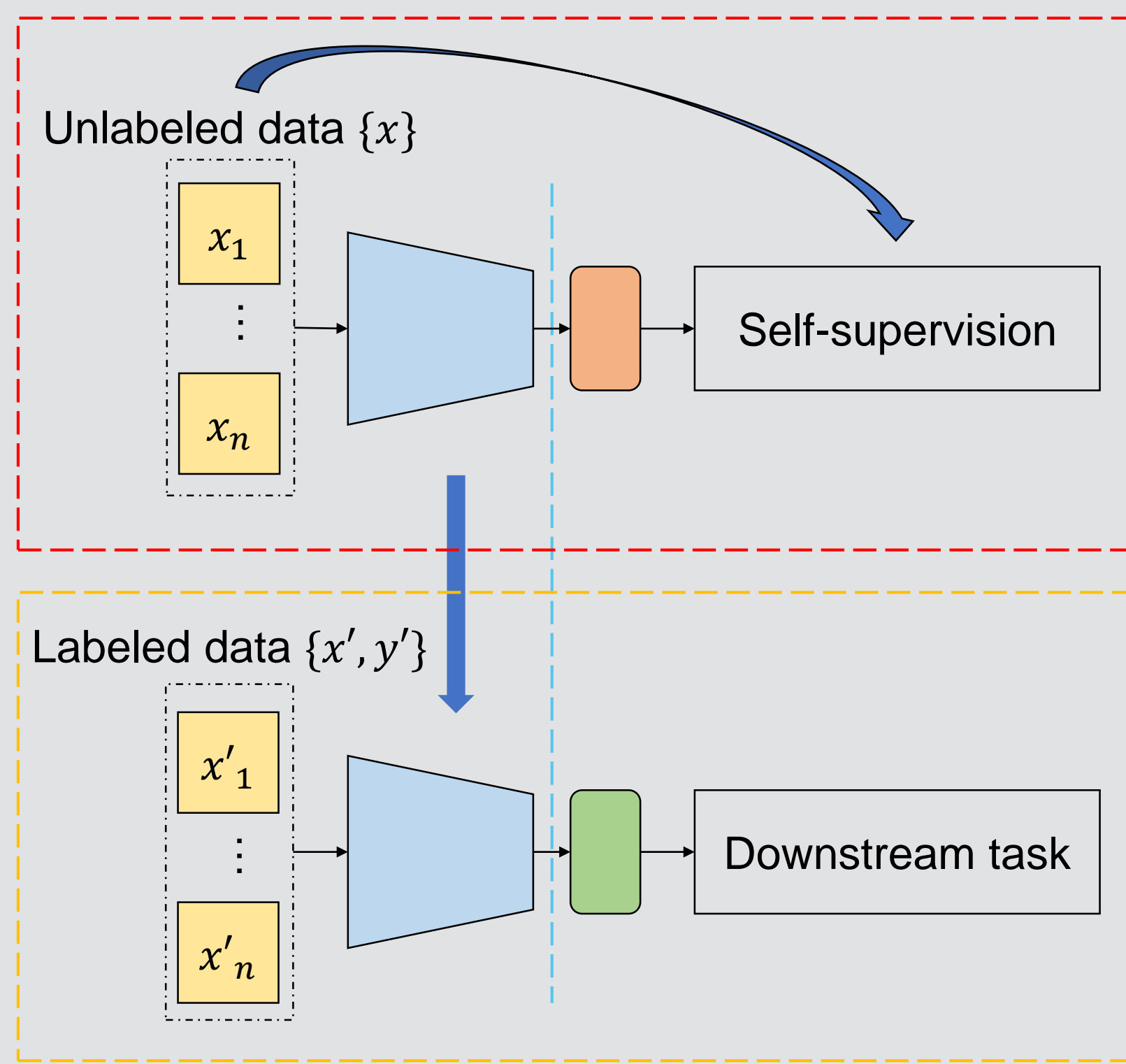
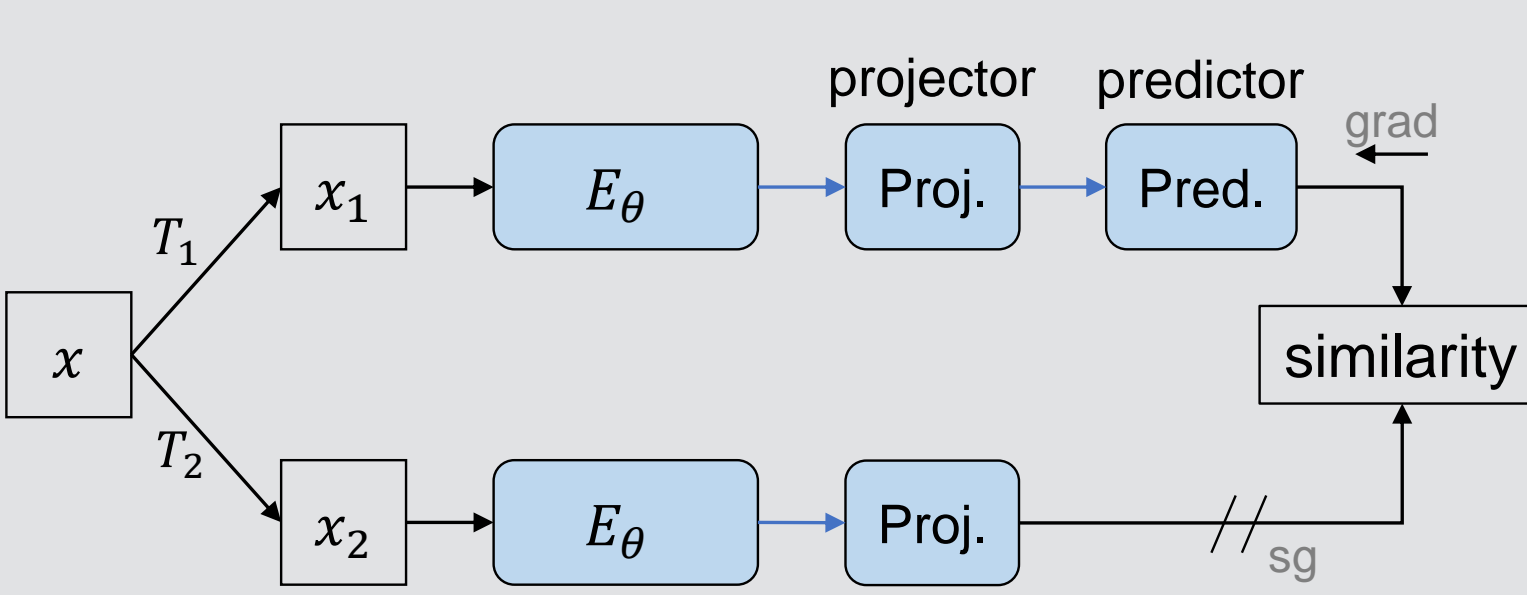

Four Modern Self-supervised Algorithms
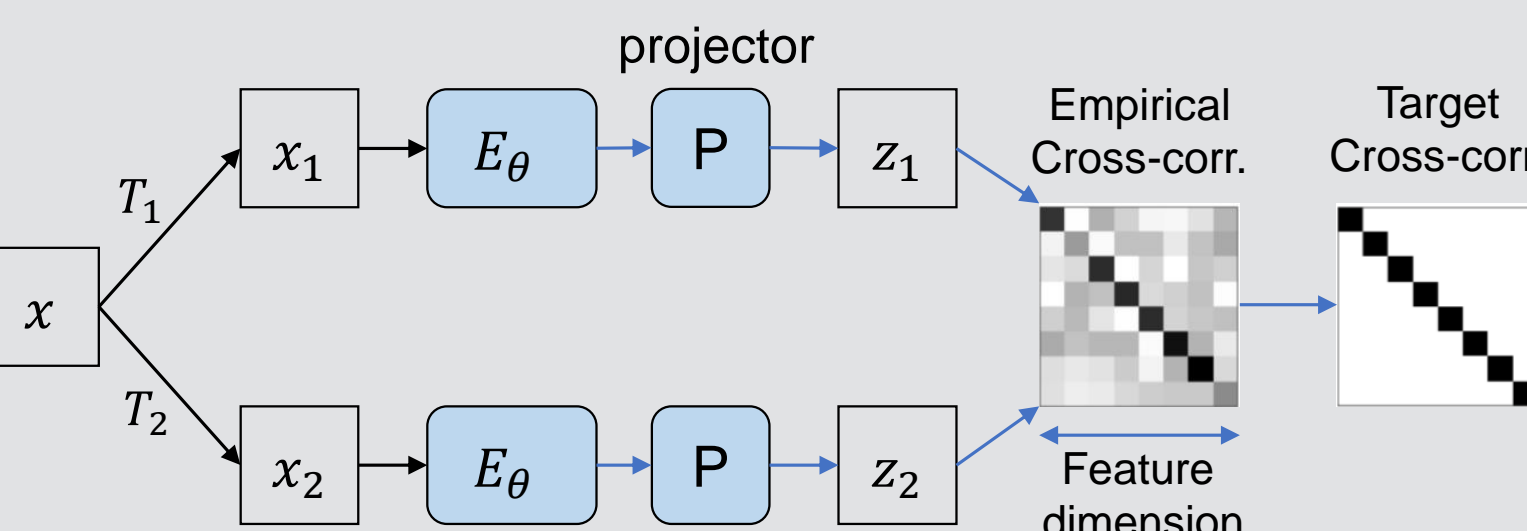
- **Mo**mentum **co**ntrast (MoCo-v2)

- **Sw**apping **a**ssignments between **v**iews (SwAV)

- **Sim**ple **Siam**ese network (SimSiam)

- **Barlow Twins**

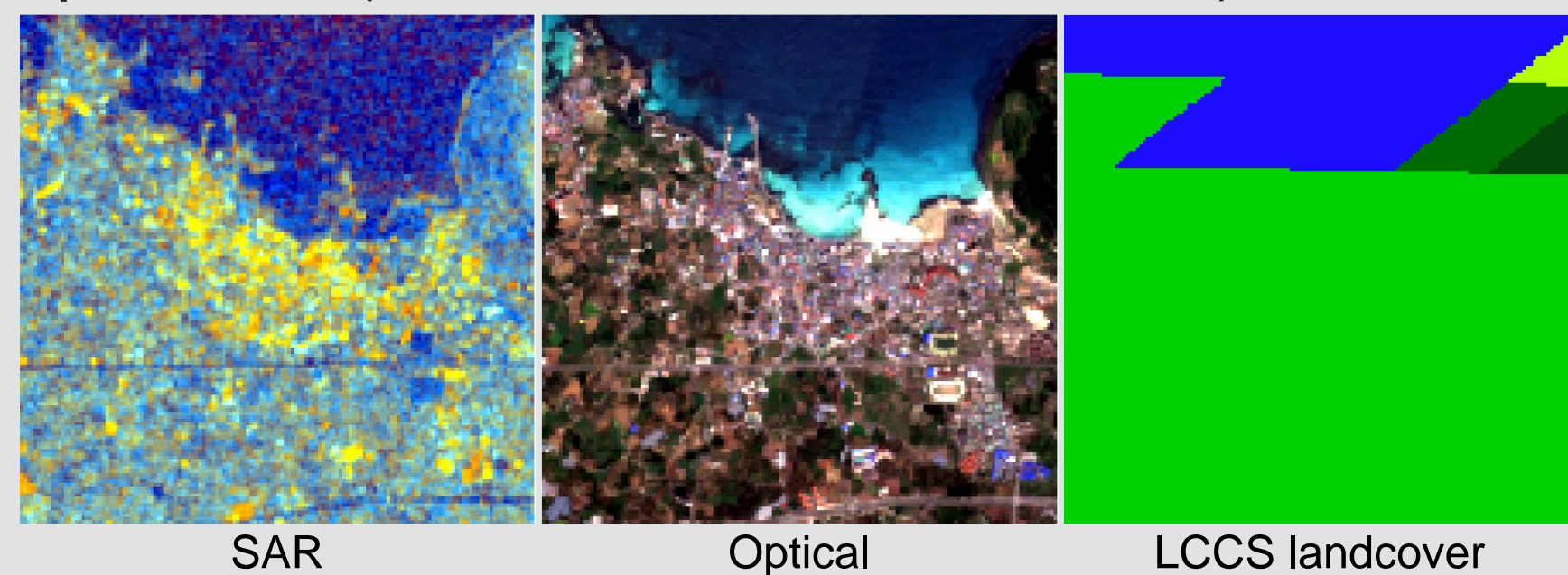## Four Popular Remote Sensing Datasets

- BigEarthNet: multi-label landcover classification, ~600k Sentinel-2 patches.

permanently irrigated land, sclerophyllous vegetation, beaches, dunes, sands, estuaries, sea and ocean

non-irrigated arable land, fruit trees and berry plantations, agro-forestry areas, transitional woodland/shrub

- SEN12MS: landcover classification, ~180k Sentinel-1/2 patches. (We use scene labels here.)

SAR   Optical   LCCS landcover

- So2Sat-LCZ42: local climate zone classification, ~400k Sentinel-1/2 patches.

compact high-rise

SAR   Optical   LCZ label
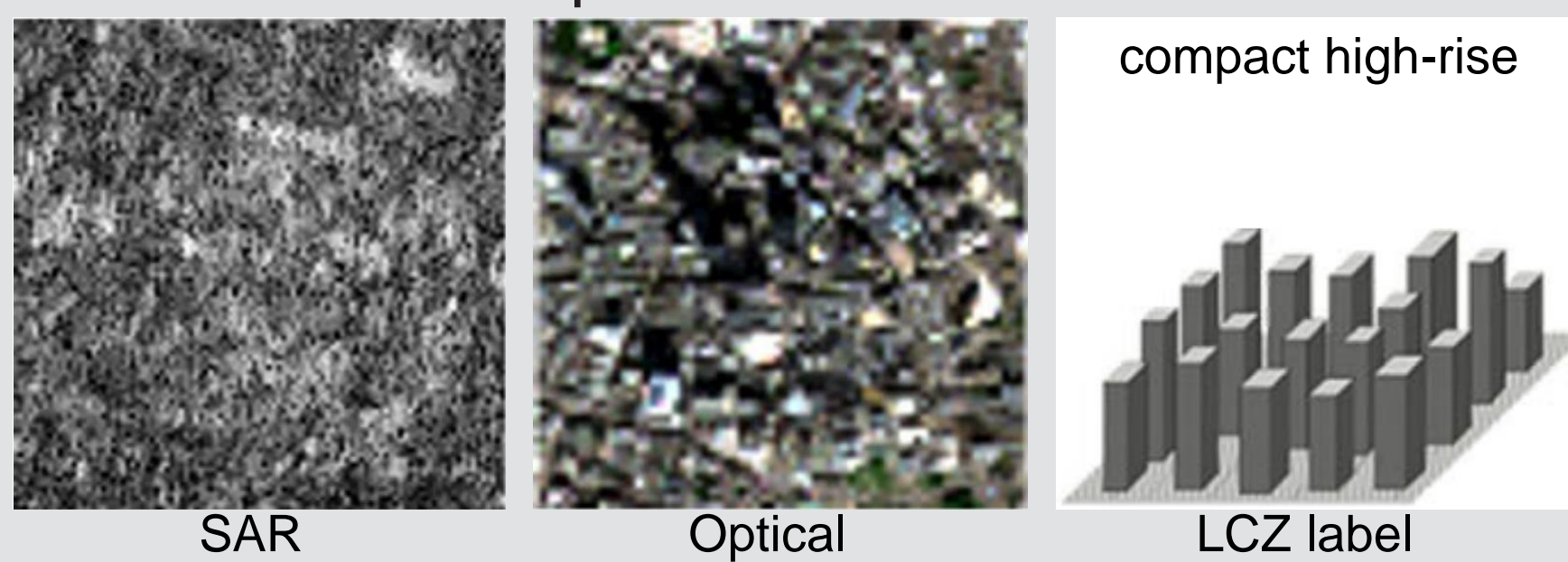
- EuroSAT (transfer only): landcover classification, 27k Sentinel-2 patches.

Industrial buildings   Residential buildings   Annual crop

Experiments: Representation Visualization

- Pretrain ResNet-18 on BigEarthNet with MoCo-v2 and transfer to EuroSAT.
- T-SNE visualization of EuroSAT feature vectors (without labels).
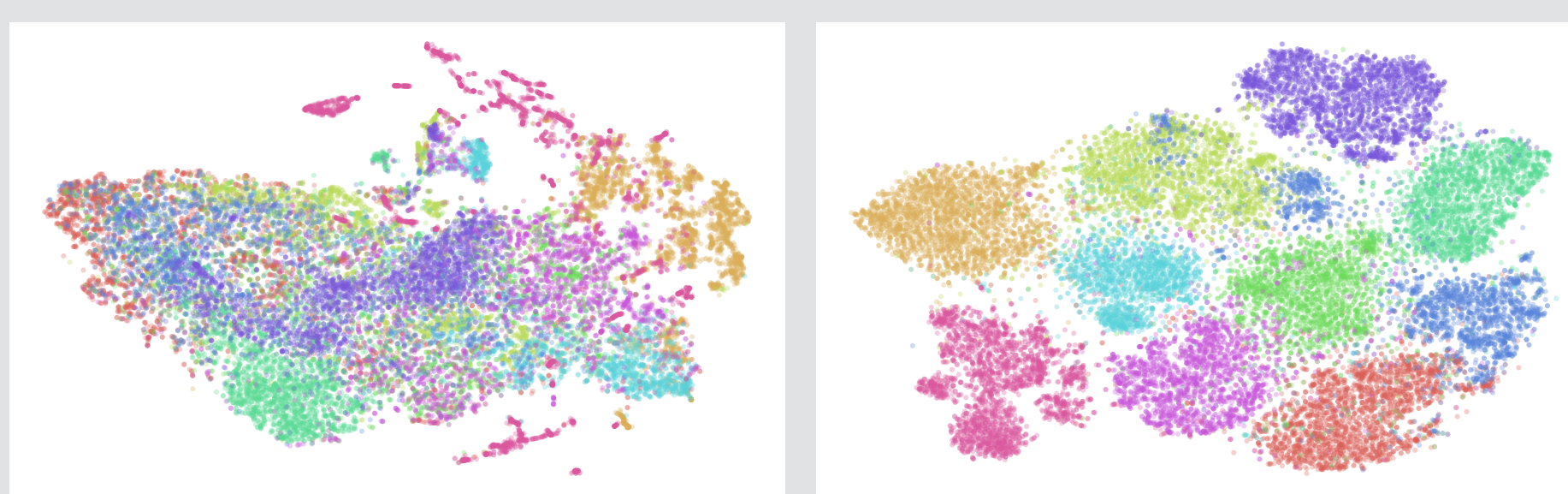
Random initialization *vs* SSL pre-training

Experiments: Baseline results

- Pretrain ResNet-18 on each dataset's training split with each SSL method.
- Evaluate with linear probing (freeze the pre-trained encoder and train only a linear classifier).
- Random initialization and supervised training for comparison.
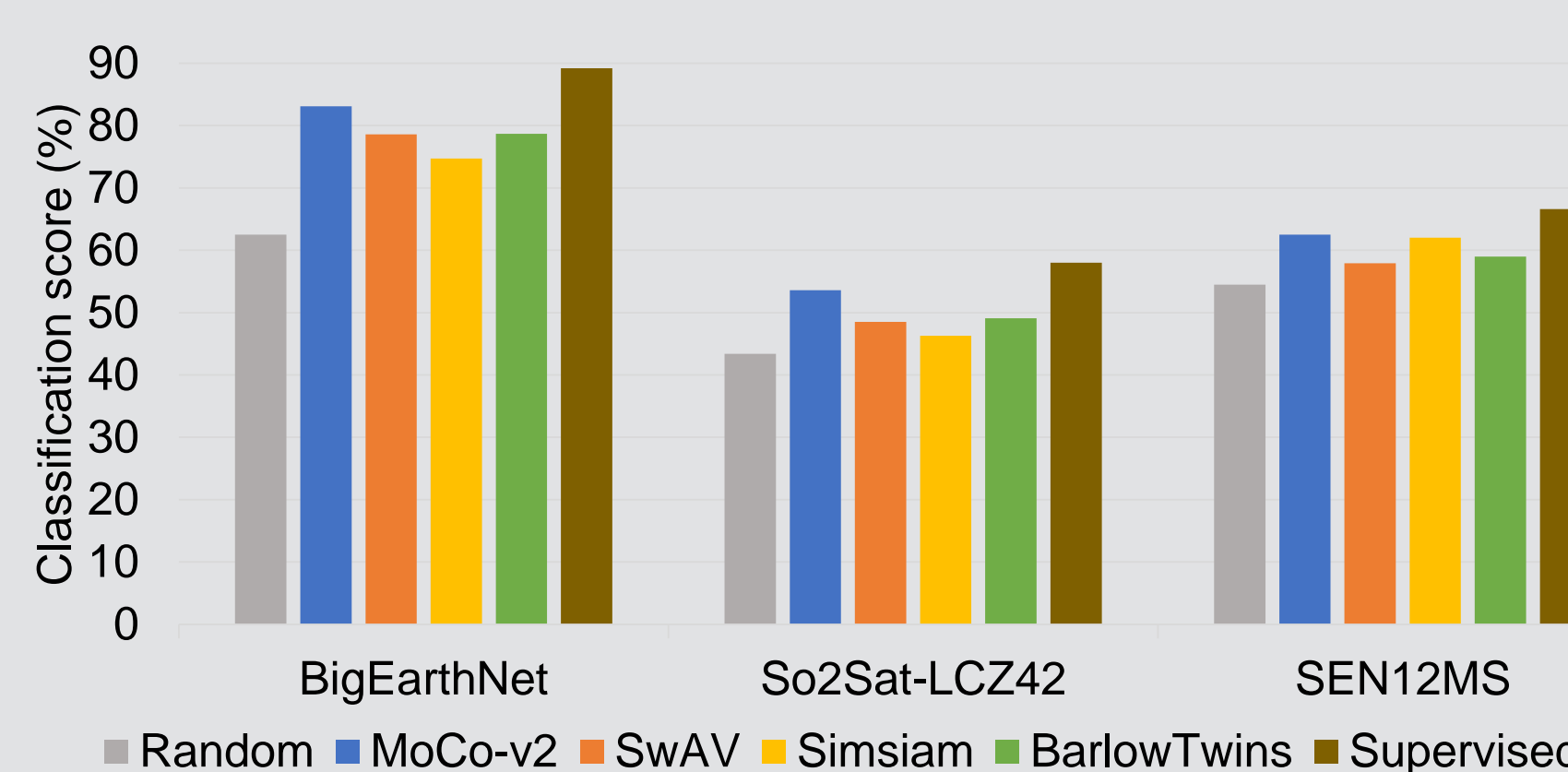
Experiments: Limited Labels & Transfer Learning

- **Label-limited Regime**: the fewer labels available, the bigger the advantage of self-supervised pre-training (pre-train and evaluate on BigEarthNet).
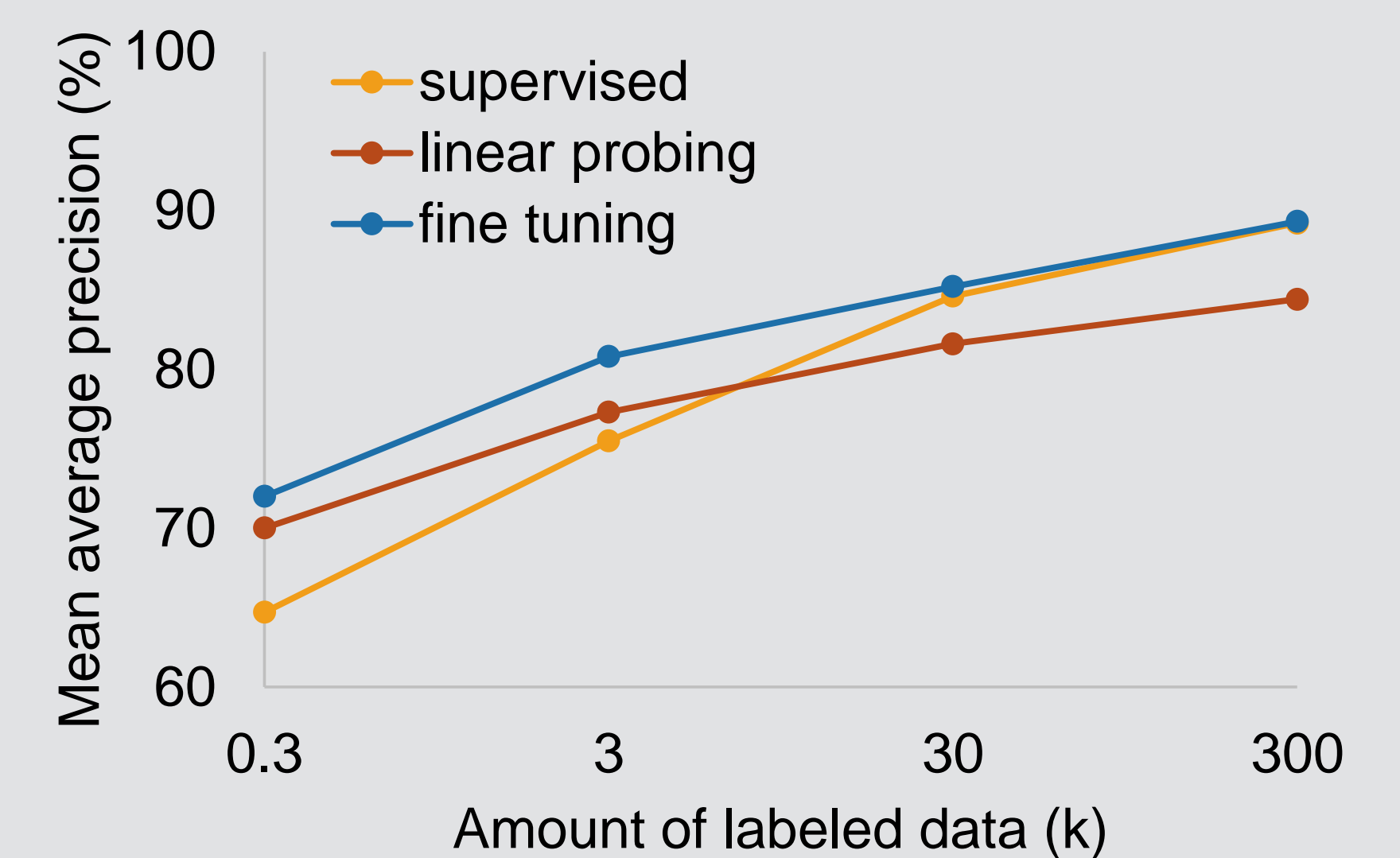

- **Transfer Learning**: self-supervised pre-training transfers well (pre-train on BigEarthNet, transfer to EuroSAT).
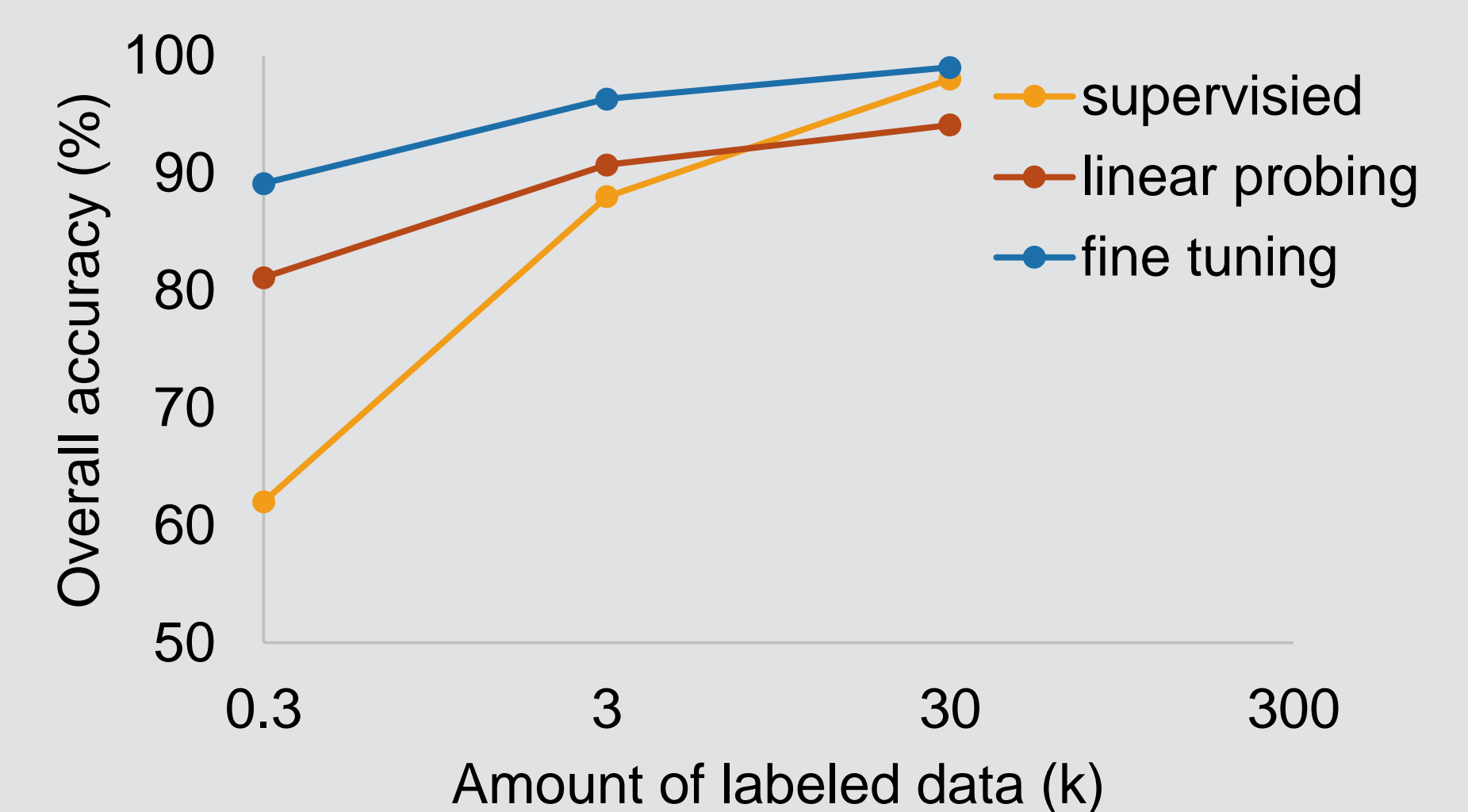
Experiments: Data Augmentation

- **Data augmentations**: ResizedCrop, ColorJitter, GrayScale, GaussianBlur, HorizontalFlip, Channel-Drop.
- Compared to natural images, cropping bears more importance in earth observation, while the other augmentations much less.
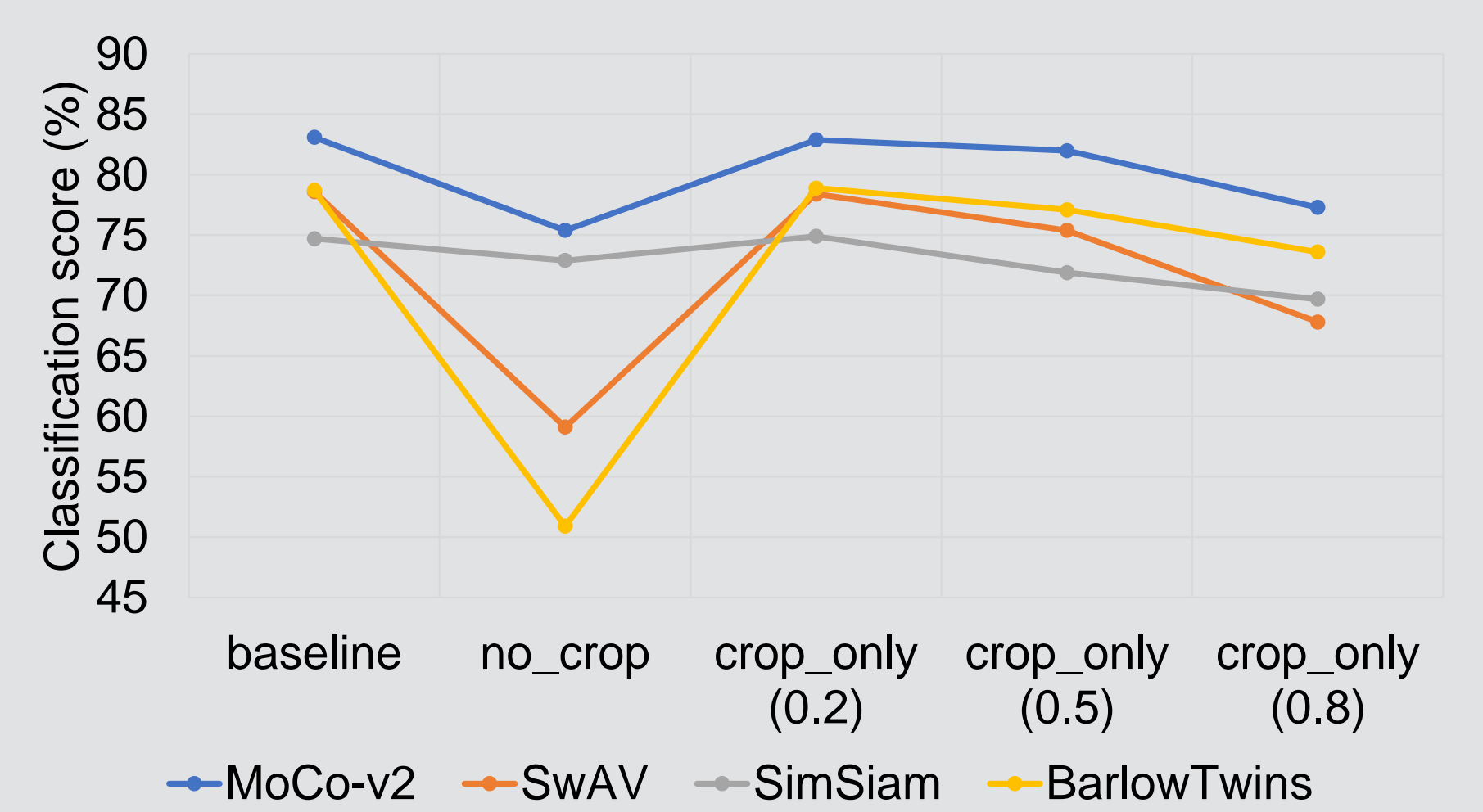
Conclusion

- Self-supervised learning proves to bear huge potential in earth observation data science, reaching comparable or better performance than supervised learning from scratch.
- The importance of data augmentation in spaceborne images is different from natural images, calling for careful design with domain specific knowledges.