High Level Computer Vision

Summer Semester 2023

S3-TSS: Self-Supervision in time for Satellite Images

A novel method of SSL technique in Satellite images

Akansh Maurya (7047939)

Hewan Shrestha (7047533)

Mohammad Munem Shahriar (7002640)

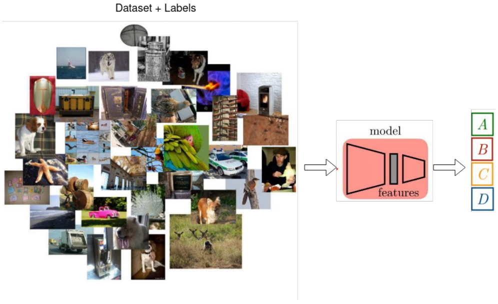
max planck institut informatik | SIC Saarland Informatics Campus | UNIVERSITÄT DES SAARLANDES

Link to this Presentation

Content

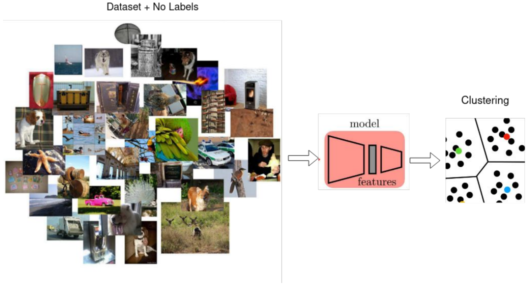
- Supervised vs Unsupervised vs Self-supervision
- General Framework of Self-supervision learning
- Motivation: Problem of existing methods in SSL for satellite images
- Self-supervision in time
- Experiments and results
- Comparison with related work
- Conclusion
- Extension

Supervised vs Unsupervised vs Self-supervision



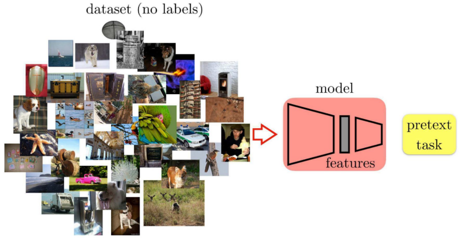
Supervised Learning

- Labelled Dataset
- Train on labelled data to predict or classify
- Ex. Classification, Regression



Unsupervised Learning

- Unlabelled Dataset
- Train on unlabelled data to find a pattern or structure
- Ex. Clustering, KNN



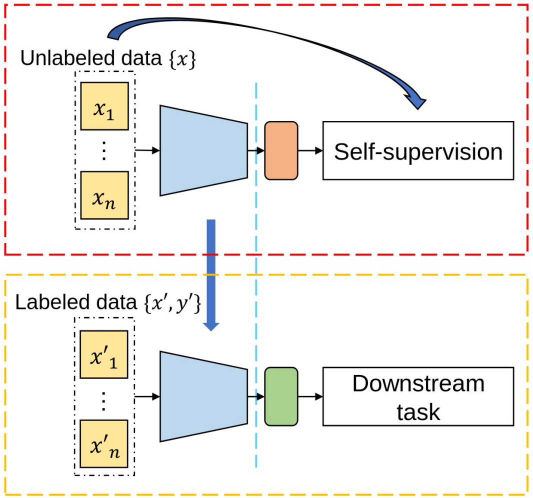
Self-supervised Learning

- Unlabelled Dataset
- Train on unlabelled data to label itself
- Ex. Pretext Task

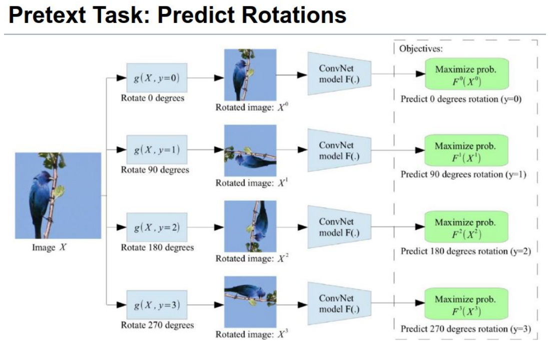
Images: https://openaccess.thecvf.com/content_cvpr_2018/papers/Noroozi_Boosting_Self-Supervised_Learning_CVPR_2018_paper.pdf

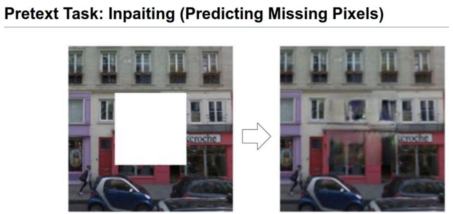
- The performance of Deep Learning methods is very sensitive to the size and quality of training data.
- Annotating a large dataset has its own challenges:
  - Laborious
  - Time Consuming
  - Expensive
  - Prone to human error
- For many application there exist a enormous amount of unlabelled data. Eg. Medical images like Chest X-rays captured on daily basis, satellite images, camera recording etc.
- Self-supervised learning methods aims to utilize the data.

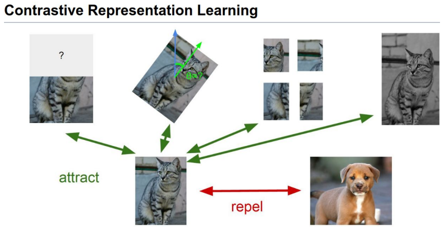
General Framework of Self-supervision learning

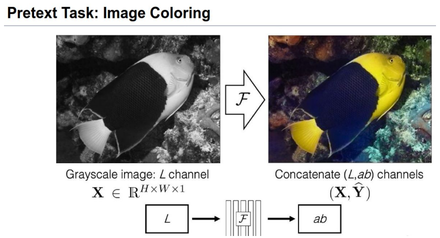


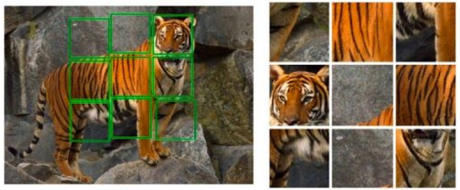
Self Supervised Learning Framework











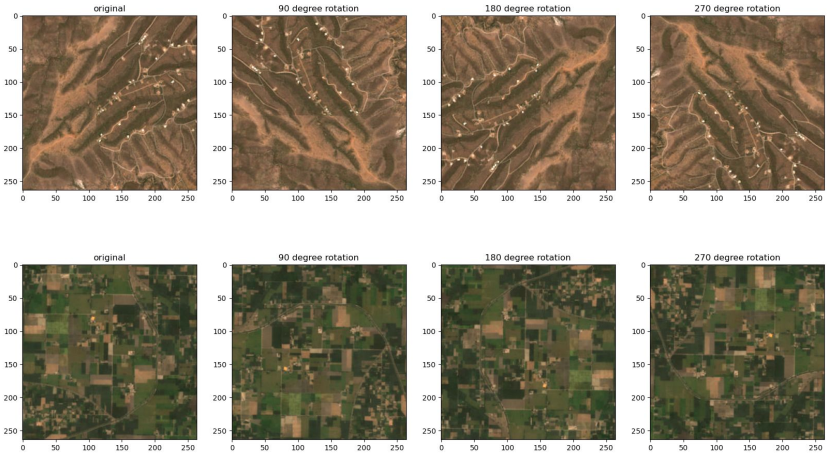
Example of pre-text task and methods

Source: High Level Computer Vision | Bernt Schiele

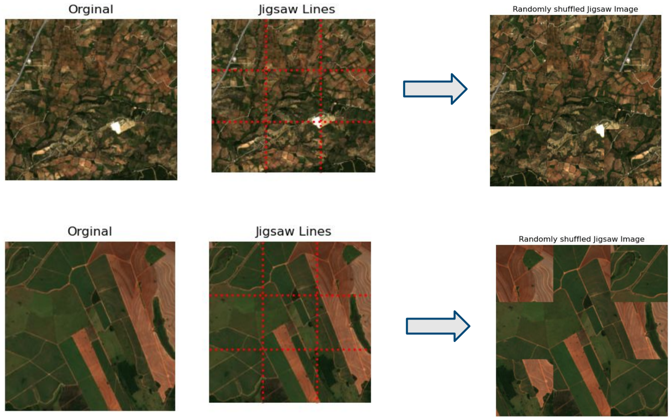
Problem with existing SSL methods for Satellite Images

Performance of Self Supervised methods is majorly depended on:
- Choice of Augmentations
- Methods: Contrastive(MoCo, SimCLR), BYOL, DINO etc.

Rotation



Jigsaw puzzle



Conclusion →
- **Better augmentations are needed for satellite images.**
- **They are very different from natural images.**

Self-Supervision in time

Bauhaus, Saarbrücken, Germany

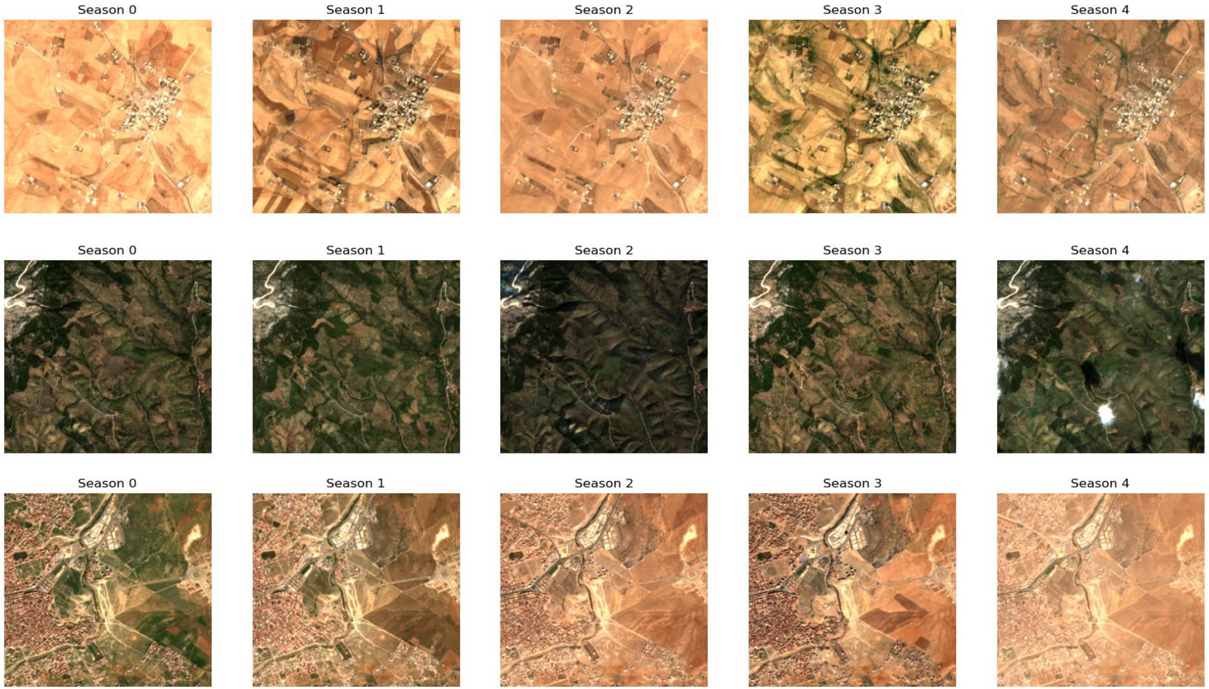
Over of period of time, satellite images goes through (Natural Augmentations):

- Stationary changes:
  - Lightning, solar radiation
  - Fogs, Clouds (Obstruction)
  - Buildings, trees are stationary
  - Weather Condition, seasons
  - Day-night
  - Etc.
- Non-Stationary Changes:
  - Movements of objects like cars
  - Construction activities
  - Etc.
- No Artificial Augmentation will able to replicate these changes.

Can we leverage the Natural Augmentation that happens with satellite images over time?

Images from Planet.com; data curated us.

Self-Supervision in time



Natural Augmentation in time

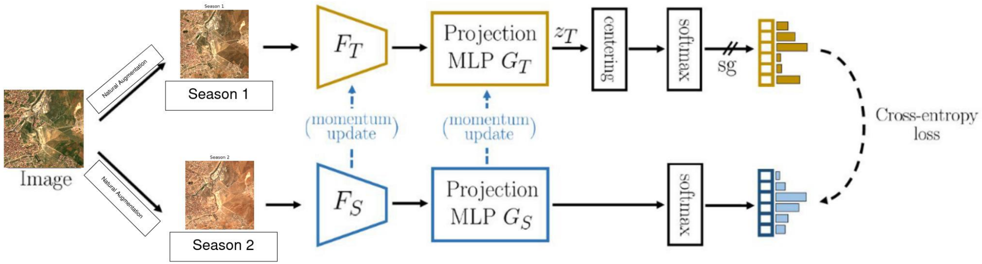
Images from SeCo Dataset:
https://arxiv.org/abs/2103.16607

DINO: Self-Distillation with No Labels

Self-Supervision in time

Our Approach:

- Use State of the Art methods Self supervised method, discussed in Emerging Properties in DINO: Self-**Di**stillation with **No** Labels.
  - The advantage of this SSL method is that it does not require negative example,
- Instead of using Artificial Augmentation, we will be using Natural Augmentations.



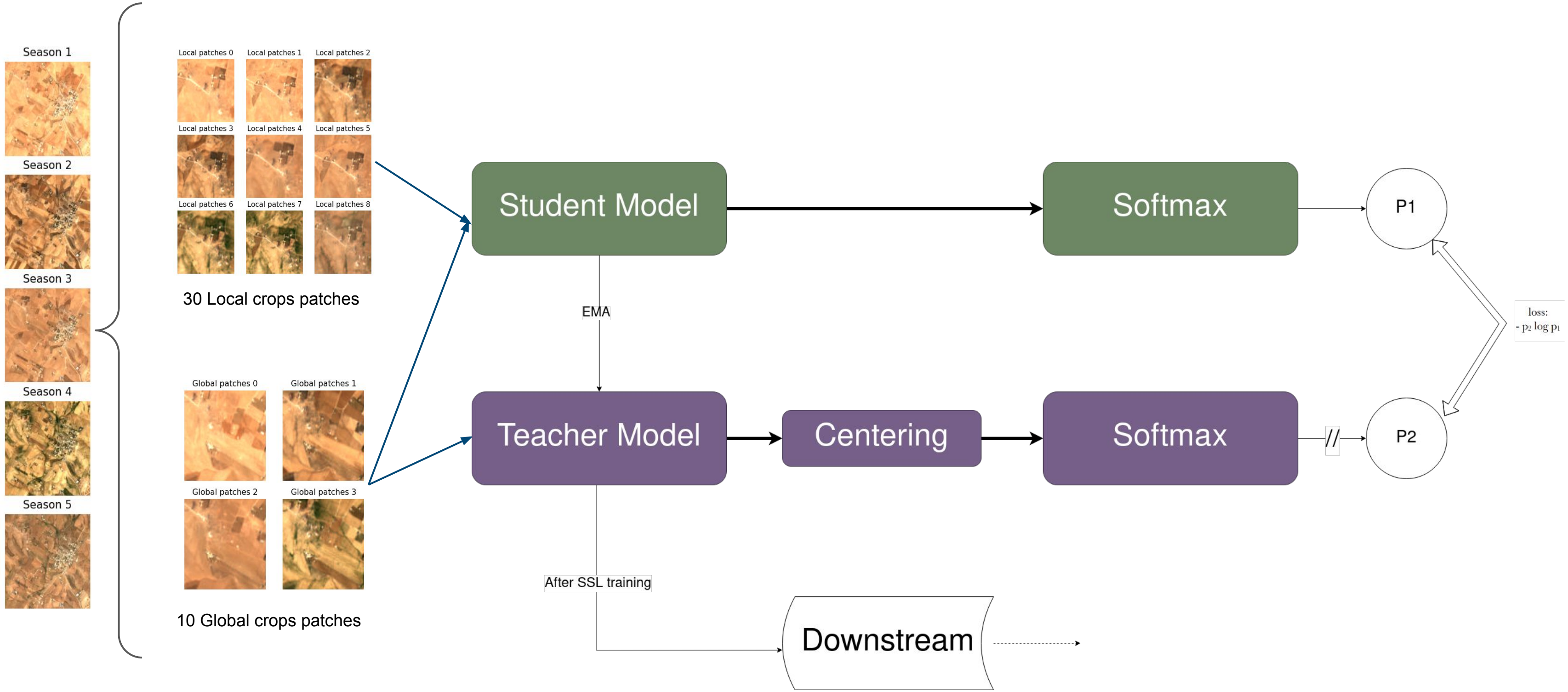
DINO Architecture with Natural Augmentation Modification

Algorithm 1 DINO PyTorch pseudocode w/o multi-crop.

```
# gs, gt: student and teacher networks
# C: center (K)
# tps, tpt: student and teacher temperatures
# l, m: network and center momentum rates
gt.params = gs.params
for x in loader: # load a minibatch x with n samples
    x1, x2 = augment(x), augment(x) # random views

    s1, s2 = gs(x1), gs(x2) # student output n-by-K
    t1, t2 = gt(x1), gt(x2) # teacher output n-by-K

    loss = H(t1, s2)/2 + H(t2, s1)/2
    loss.backward() # back-propagate

    # student, teacher and center updates
    update(gs) # SGD
    gt.params = l*gt.params + (1-l)*gs.params
    C = m*C + (1-m)*cat([t1, t2]).mean(dim=0)

def H(t, s):
    t = t.detach() # stop gradient
    s = softmax(s / tps, dim=1)
    t = softmax((t - C) / tpt, dim=1) # center + sharpen
    return - (t * log(s)).sum(dim=1).mean()
```

Source: https://github.com/facebookresearch/dino

Overview of our method: Self-Supervision in time (S3-TSS)



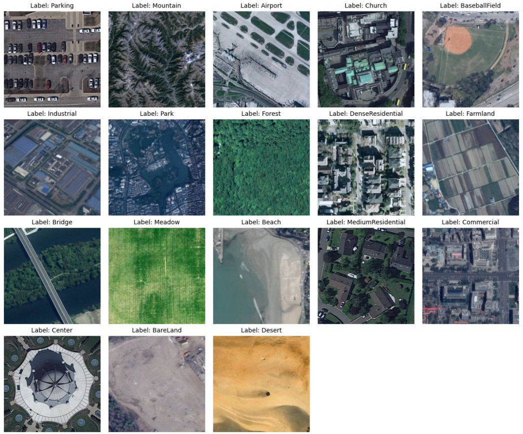
Dataset Descriptions

EuroSAT

Total Samples	27000
Number of Classes	10

AID(Aerial Image Dataset)

Total Samples	5786
Number of Classes	18



Source:
https://zenodo.org/record/7711810#.ZAm3k-zMKEA
https://captain-whu.github.io/AID/

Dataset Descriptions

UCMerced

Total Samples	2100
Number of Classes	21

WHU-RS19

Total Samples	950
Number of Classes	19

Source:
http://weegee.vision.ucmerced.edu/datasets/landuse.html
https://www.kaggle.com/datasets/sunray2333/whurs191

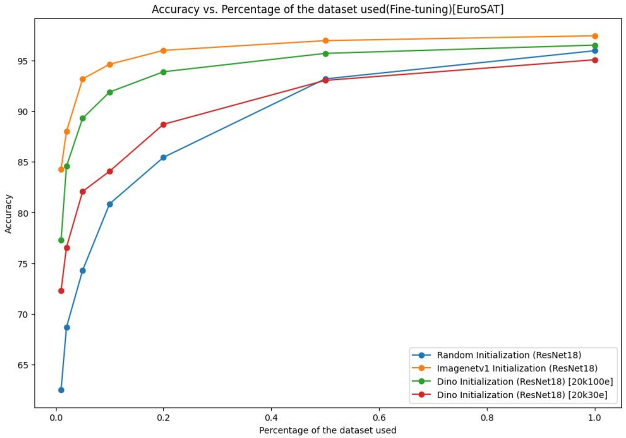
Experiment Setup

- Architecture Used; Resnet-18, Resnet-50.
- Pre-training Dataset: Seasonal Contrast (SeCo)
- Downstream Tasks: Classification
- Downstream Datasets: EuroSAT, AID, UCMerced, WHU-RS19
- Optimizer and Learning rate and majorly all the hyperparameters are constant for all the experiments unless stated.
- Metric:
    - Linear Probing
    - Fine-tuning

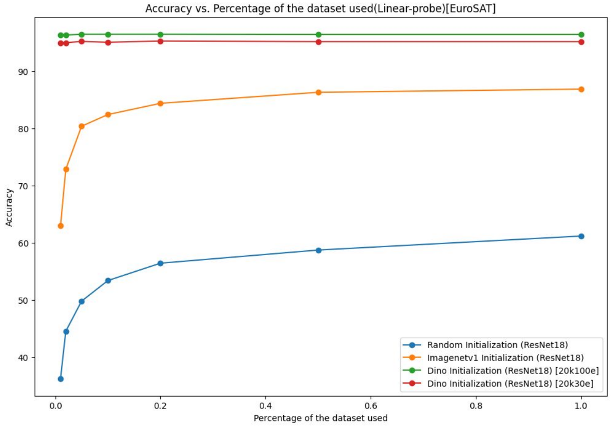
- **Main Question: Can Natural Augmentation perform better than Artificial Augmentation?**

Experiment 1

- Architecture: ResNet18
- Dataset: SeCo-20k(out of 100k)
- Epochs: 30 and 100
- Downstream Datasets: EuroSaT, AID, UCMerced, WHU-RS19
- Metric: Fine-tuning and Linear-probe



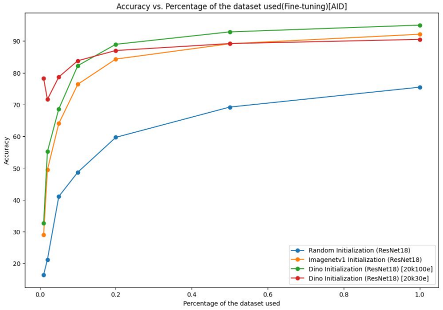
	Percentage	Random_initialization	Imagenetv1	Dino_100_epochs(20k)	Dino_30_epochs(20k)
0	0.01	62.518519	84.296296	77.296296	72.296296
1	0.02	68.666667	88.000000	84.555556	76.555556
2	0.05	74.296296	93.185185	89.296296	82.074074
3	0.10	80.851852	94.629630	91.888889	84.074074
4	0.20	85.444444	96.000000	93.888889	88.703704
5	0.50	93.185185	96.962963	95.703704	93.037037
6	1.00	95.962963	97.444444	96.518519	95.074074

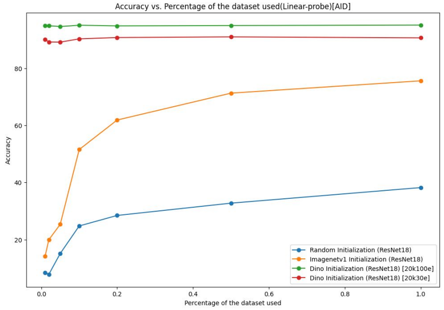


	Percentage	Random_initialization	Imagenetv1	Dino_100_epochs(20k)	Dino_30_epochs(20k)
0	0.01	36.296296	63.000000	96.296296	94.962963
1	0.02	44.555556	72.851852	96.296296	94.962963
2	0.05	49.814815	80.370370	96.444444	95.185185
3	0.10	53.407407	82.407407	96.444444	95.037037
4	0.20	56.444444	84.370370	96.444444	95.259259
5	0.50	58.740741	86.296296	96.407407	95.148148
6	1.00	61.185185	86.851852	96.407407	95.148148

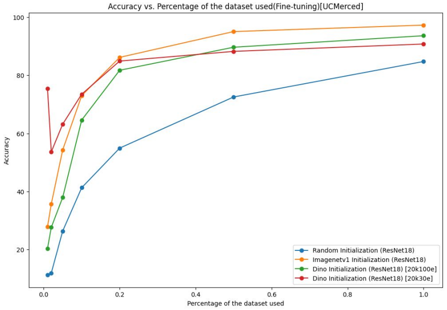
Experiment 1

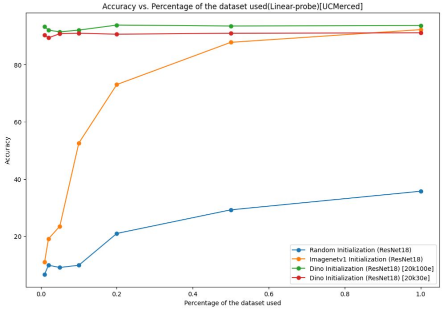
AID Dataset



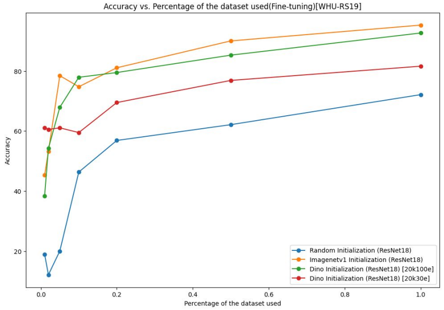


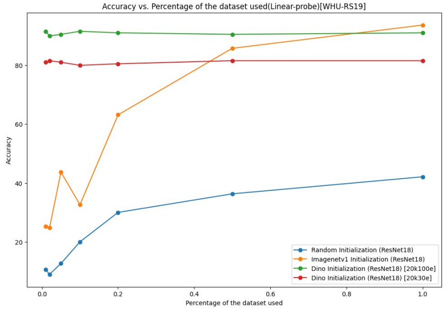
UCMerced Dataset





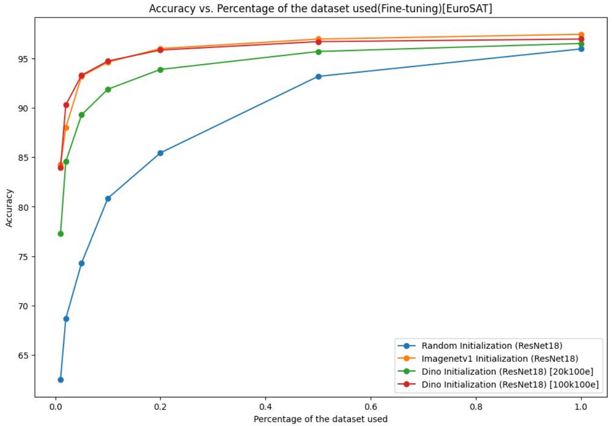
WHU-RS19 Dataset

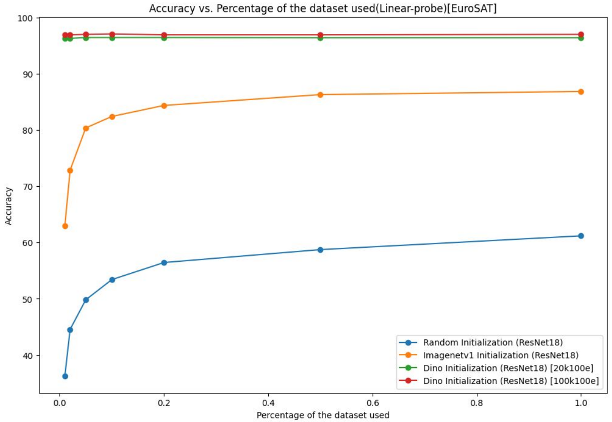




Experiment 2

- Architecture: ResNet18
- Dataset: SeCo-100k
- Epochs: 100
- Downstream Datasets: EuroSaT, AID, UCMerced, WHU-RS19
- Metric: Fine-tuning and Linear-probe



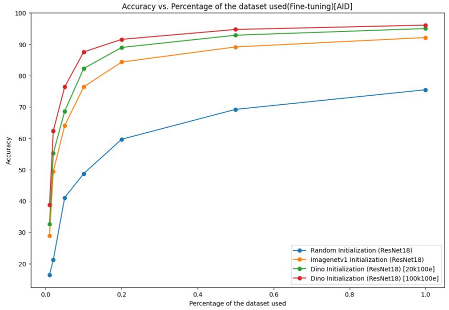


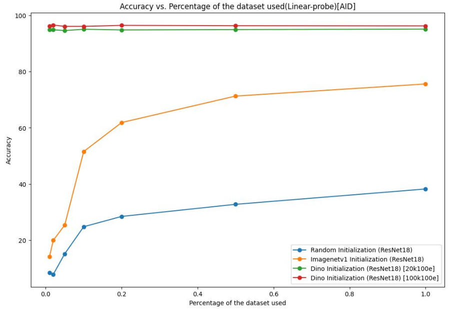
	Percentage	Random_initialization	Imagenetv1	Dino_100_epochs(20k)	Dino_100_epochs(100k)
0	0.01	62.518519	84.296296	77.296296	83.962963
1	0.02	68.666667	88.000000	84.555556	90.296296
2	0.05	74.296296	93.185185	89.296296	93.296296
3	0.10	80.851852	94.629630	91.888889	94.740741
4	0.20	85.444444	96.000000	93.888889	95.851852
5	0.50	93.185185	96.962963	95.703704	96.703704
6	1.00	95.962963	97.444444	96.518519	96.962963

	Percentage	Random_initialization	Imagenetv1	Dino_100_epochs(20k)	Dino_100_epochs(100k)
0	0.01	36.296296	63.000000	96.296296	96.925926
1	0.02	44.555556	72.851852	96.296296	96.925926
2	0.05	49.814815	80.370370	96.444444	97.000000
3	0.10	53.407407	82.407407	96.444444	97.074074
4	0.20	56.444444	84.370370	96.444444	96.925926
5	0.50	58.740741	86.296296	96.407407	96.925926
6	1.00	61.185185	86.851852	96.407407	97.000000

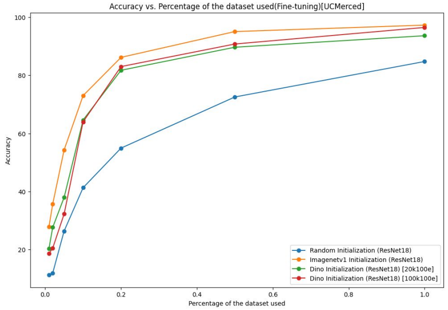
Experiment 2

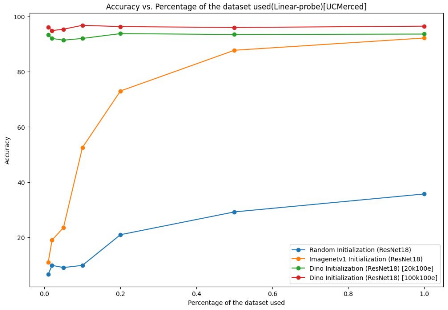
AID Dataset



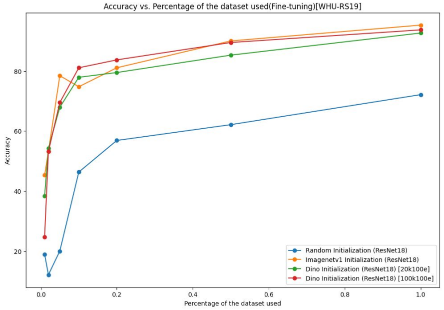


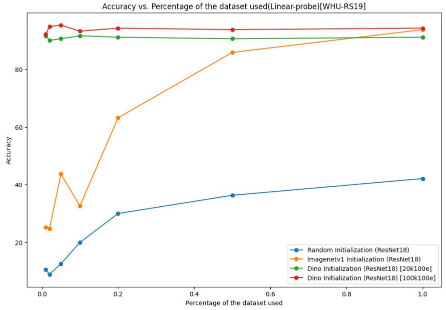
UCMerced Dataset



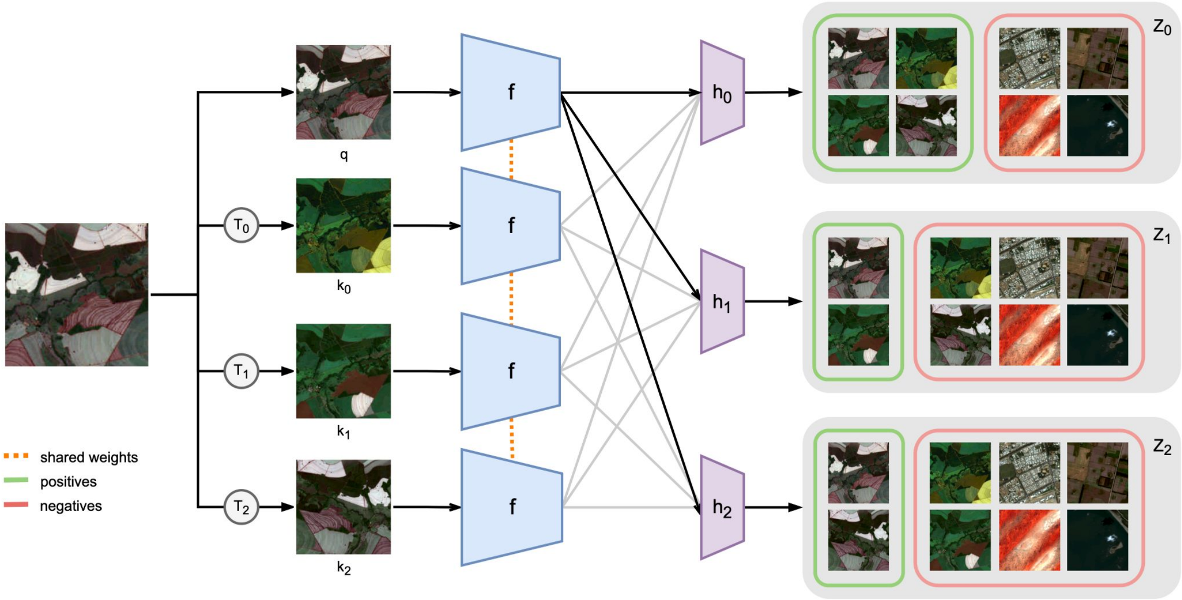


WHU-RS19 Dataset





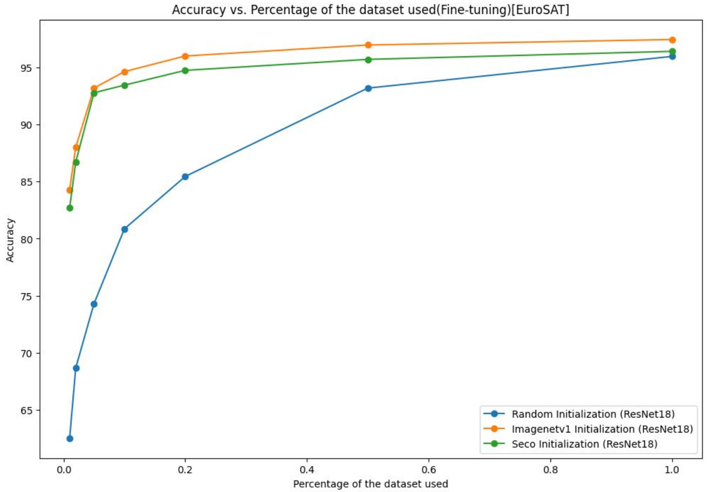
Seasonal Contrast (Baseline)

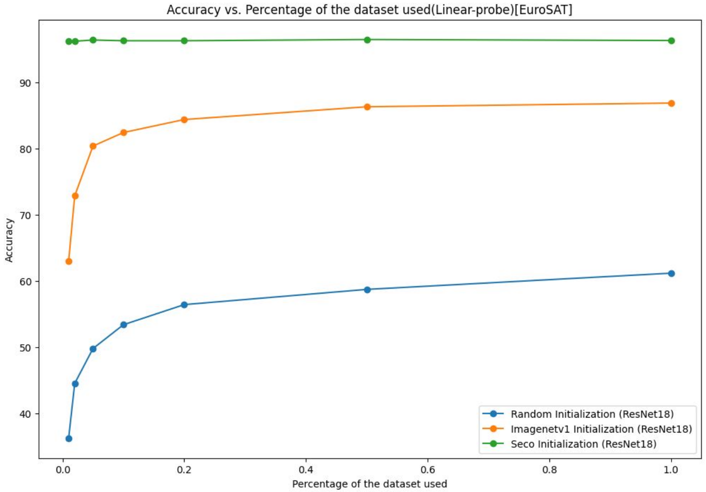


Source:
https://arxiv.org/abs/2103.16607

SeCo Baseline vs ImageNet

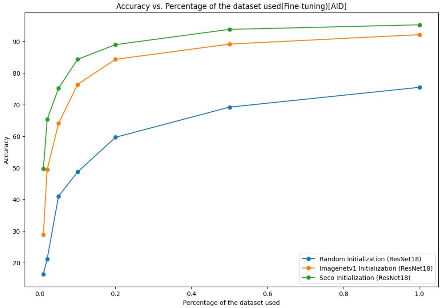
EuroSAT Dataset

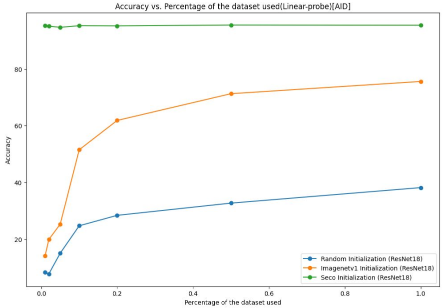




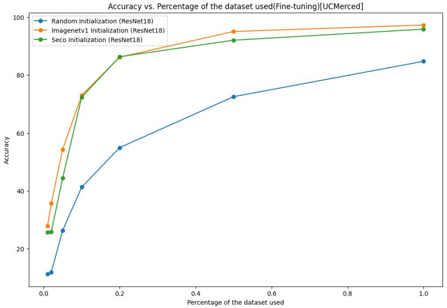
SeCo Baseline on Other Datasets

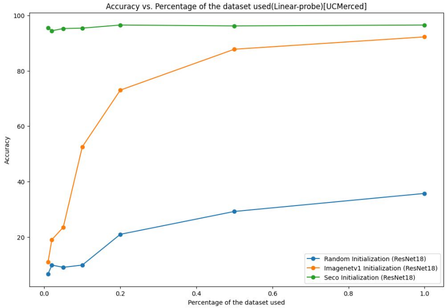
AID Dataset



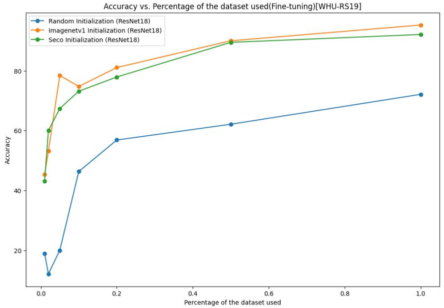


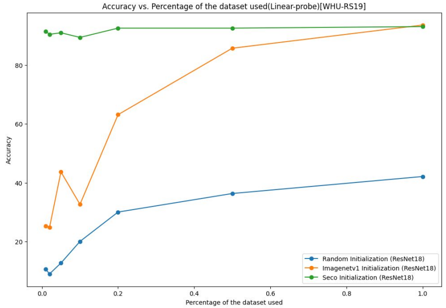
UCMerced Dataset





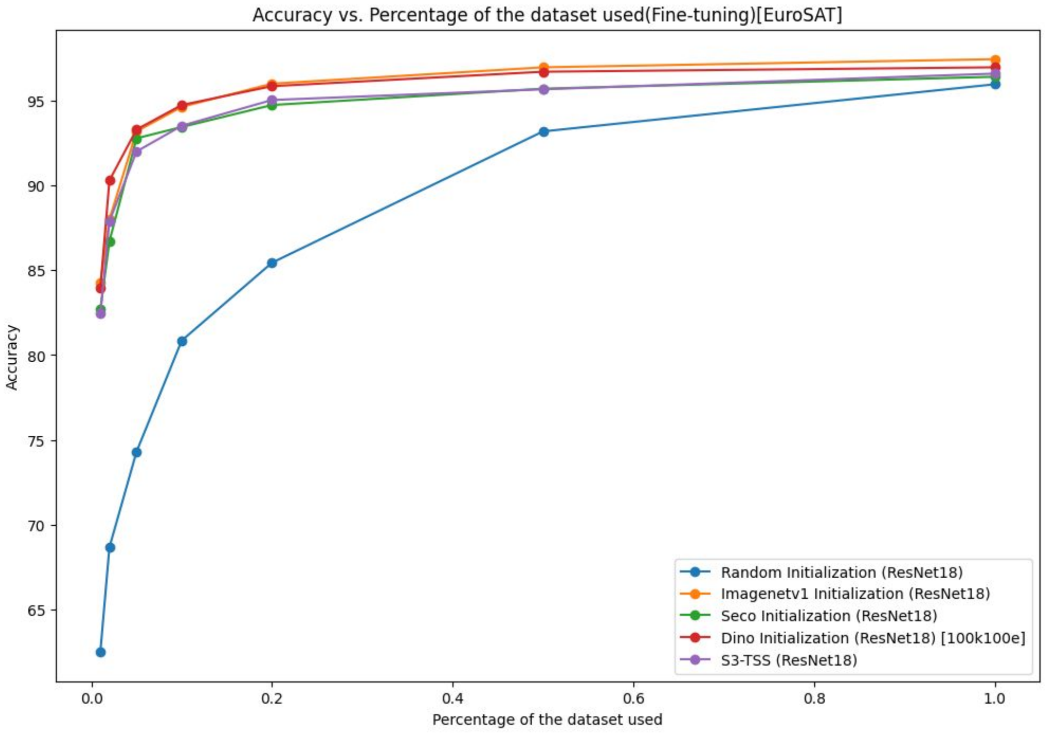
WHU-RS19 Dataset

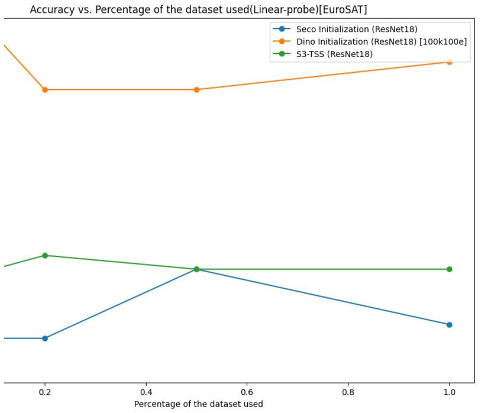




Experiment 3

- Architecture: Re
- Dataset: SeCo-1
- Epochs: 100
- Self-Supervisior
- Downstream Da
  AID, UCMerced,
- Metric: Fine-tun
- Comparison wit





1.00	95.962963	97.444444	96.407407	96.962963	96.592593

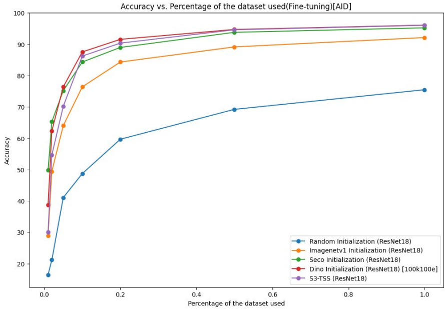
	ation	Imagenetv1	SeCo_initialization	Dino_100_epochs(100k)	S3-TSS
	6296	63.000000	96.222222	96.925926	96.481481
	5556	72.851852	96.185185	96.925926	96.333333
	4815	80.370370	96.370370	97.000000	96.370370
	7407	82.407407	96.259259	97.074074	96.444444
	4444	84.370370	96.259259	96.925926	96.481481
	0741	86.296296	96.444444	96.925926	96.444444
1.00	61.185185	86.851852	96.296296	97.000000	96.444444

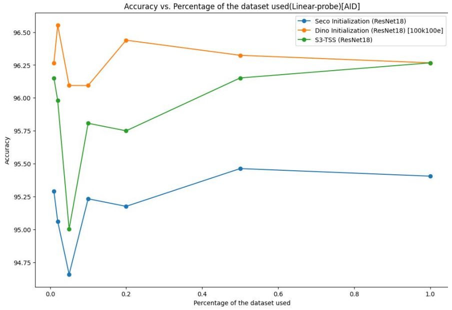
Experiment 3

max planck institut informatik

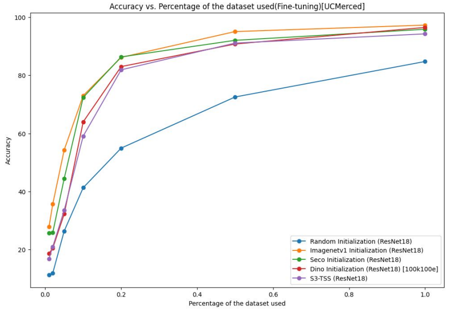
UNIVERSITÄT DES SAARLANDES

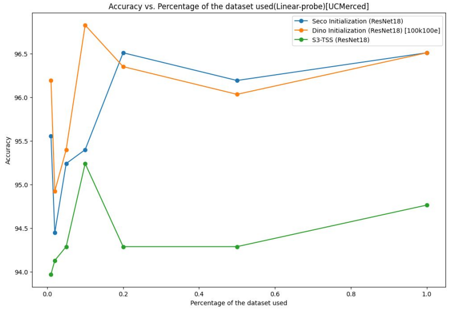
AID Dataset



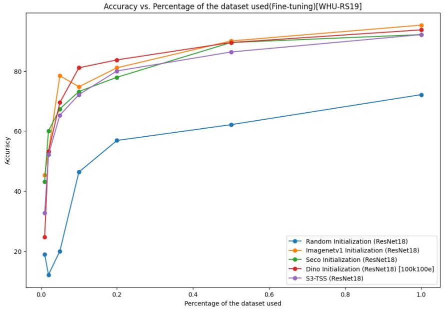


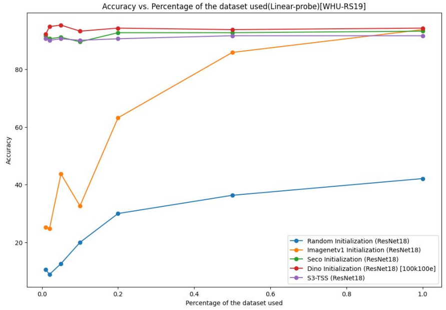
UCMerced Dataset





WHU-RS19 Dataset



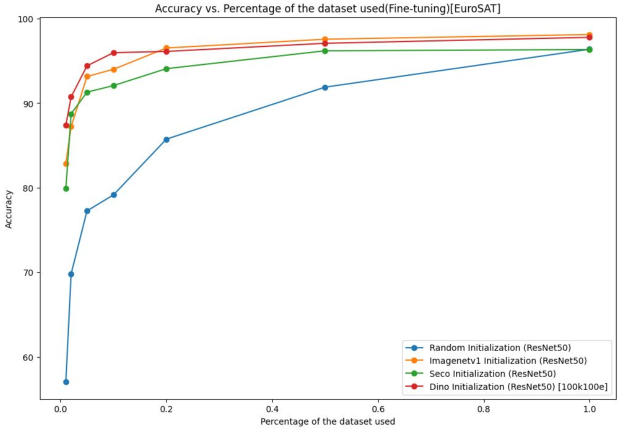


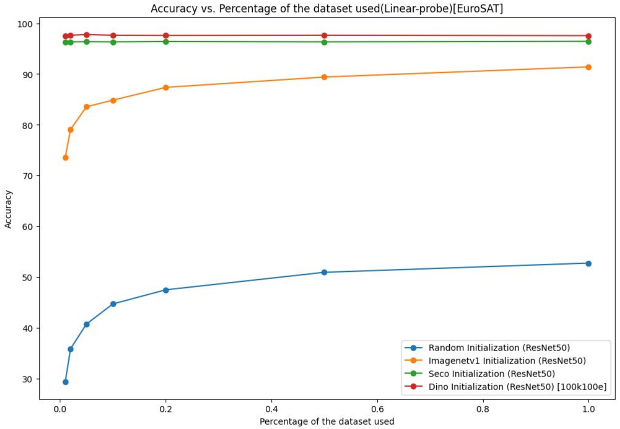
Conclusion

- **Experiment 1**: Concluded that training a self-supervised model with more epochs is better.
- **Experiment 2**: Found that the amount of data for self-supervised learning is crucial; significant improvement observed when moving from 20k to 100k images.
- **Experiment 3**: Introduced S3-TSS, surpassing SeCo without using artificial augmentation, but DINO SSL with artificial augmentation performs better at the cost of computation power.
- **Fine-tuning**: Although we have seen this continued trend that in fine-tuning, for all the datasets, ImageNet initialization performs better as compared to other initialization. We should also note that ImageNet consists of 1 million images, in contrast to us having only 100k images.
- **Linear-Probing**: S3-TSS and DINO initialization achieve superior results compared to ImageNet initialization.

Extending to ResNet50

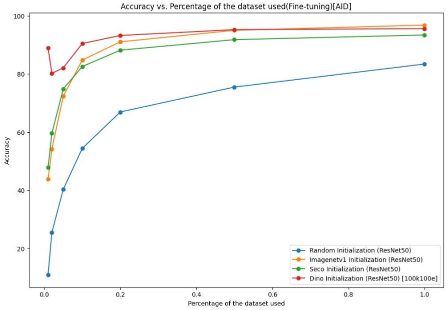
EuroSAT Dataset

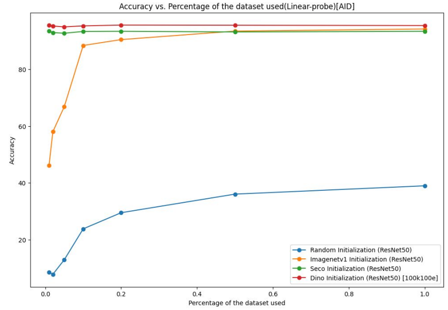




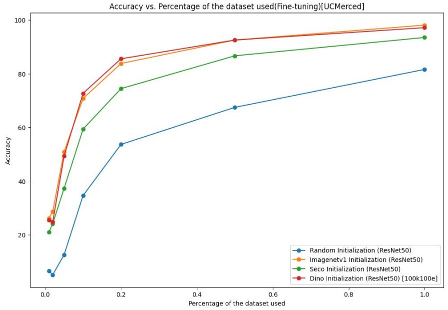
Extending to ResNet50

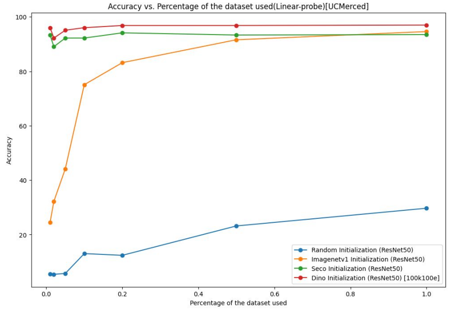
AID Dataset



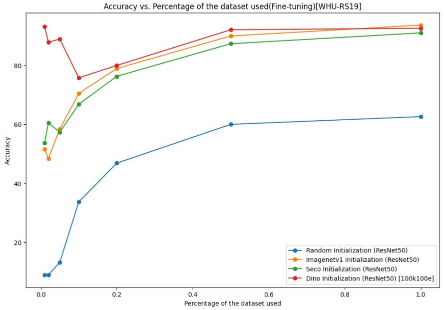


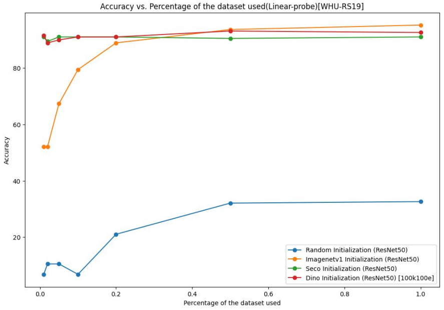
UCMerced Dataset





WHU-RS19 Dataset

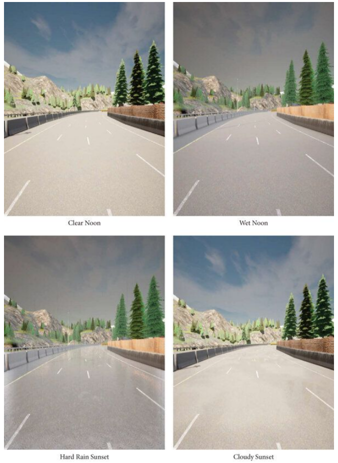




Extension of our work

- In our future work, we want to extend this work to ViT architectures.
- Also, we want to extend to SeCo-1M dataset.
- **Example Application of our work**: Same Road in different weather. Self-driving car models can be trained with this plethora of data with this method.

Method	Arch.	Param.	im/s	Linear	k-NN
Supervised	RN50	23	1237	79.3	79.3
SCLR [12]	RN50	23	1237	69.1	60.7
MoCov2 [15]	RN50	23	1237	71.1	61.9
InfoMin [67]	RN50	23	1237	73.0	65.3
BarlowT [81]	RN50	23	1237	73.2	66.0
OBoW [27]	RN50	23	1237	73.8	61.9
BYOL [30]	RN50	23	1237	74.4	64.8
DCv2 [10]	RN50	23	1237	75.2	67.1
SwAV [10]	RN50	23	1237	**75.3**	65.7
DINO	RN50	23	1237	**75.3**	**67.5**
Supervised	ViT-S	21	1007	79.8	79.8
BYOL* [30]	ViT-S	21	1007	71.4	66.6
MoCov2* [15]	ViT-S	21	1007	72.7	64.4
SwAV* [10]	ViT-S	21	1007	73.5	66.3
DINO	ViT-S	21	1007	**77.0**	**74.5**



Thank You!

Appendix

Discussion of Related Work

- One research paper [1], proposed a multitask learning framework that introduces the combination of self-supervised learning and scene classification tasks.
- This study [2], proposed a self-supervised representation learning technique for change detection in distant sensing after quantifying temporal context by coherence in time.
- On this paper [3], the researchers approached a different effective pipeline named "Seasonal Contrast (SeCo)" which can compile large unlabeled datasets of satellite photos and use self-supervised learning technique for pre-training remote sensing representations.
- Researchers in another study review [4] discussed about latest self-supervised learning developments, mainly for remote sensing.

Content

- Supervised vs Unsupervised vs Self-supervision
- General Framework of Self-supervision learning
- Motivation: Problem of existing methods in SSL for satellite images
- Self-supervision in time
- Experiments and results
- Comparison with related work
- Conclusion
- Extension

References

[1] Zhicheng Zhao, Ze Luo, Jian Li, Can Chen, and Yingchao Piao. When self-supervised learning meets scene classification: Remote sensing scene classification based on a multi-task learning framework. Remote Sensing, 12(20), 2020.

[2] Huihui Dong, Wenping Ma, Yue Wu, Jun Zhang, and Licheng Jiao. Self-supervised representation learning for remote sensing image change detection based on temporal prediction. Remote Sensing, 12(11), 2020.

[3] Oscar Mañas, Alexandre Lacoste, Xavier Giro i Nieto, David Vazquez, and Pau Rodriguez. Seasonal contrast: Unsupervised pre-training from uncurated remote sensing data, 2021.

[4] Yi Wang, Conrad M Albrecht, Nassim Ait Ali Braham, Lichao Mou, and Xiao Xiang Zhu. Self-supervised learning in remote sensing: A review, 2022.