Change-Aware Sampling and Contrastive Learning for Satellite Images

Utkarsh Mall Bharath Hariharan Kavita Bala

Cornell University

Cornell Bowers C·IS College of Computing and Information Science

CVPR 2023 VANCOUVER, CANADA

Problem

Many existing remote-sensing vision problems require *large amount* of supervision.

Self-supervised learning can alleviate this problem. However, off-the-shelf approaches do not fully utilize the *potential* of the information available in *spatio-temporal satellite images*.

Goal: Can we leverage *spatio-temporal structure* unique to satellite images in self-supervised learning?

Contributions

We present a novel **self-supervised** approach for contrastive learning on satellite images, leveraging **three properties** unique to them.

Locations may change after a long time.

- Long-term temporal Information: We propose a new loss using *long-term temporal information* in satellite images.

Many locations will have seasonal or no change even after a long time.

- Change Awareness: We present a novel approach to *estimate changes* that can be used to encourage *invariance*.

Many locations are repetitive and contain low information.

- Geographical Sampling: We use an *improved geographical sampling* that provides more informative data for representation learning.

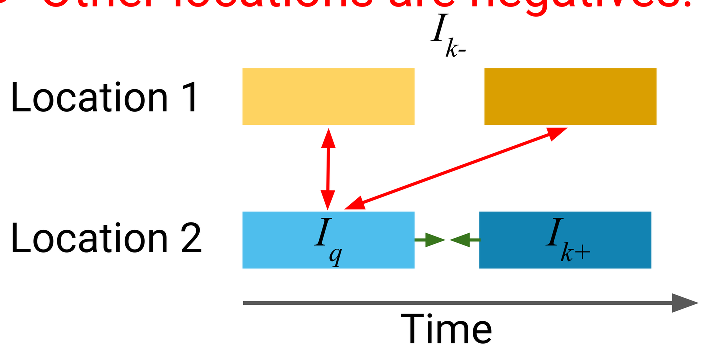
Method

Contrastive Loss and SeCo[1]

Contrastive learning can be used to learn a representation.
- Positive pairs are pulled closer.
- Negative pairs are pushed apart.

$$\mathcal{L} = -\log \frac{\boxed{exp(f(I_q)\cdot f(I_{k+})/\tau)}}{\sum_{I_k \in \mathcal{I}_{k-}+\{I_{k+}\}} \boxed{\exp(f(I_q)\cdot f(I_k)/\tau)}}$$

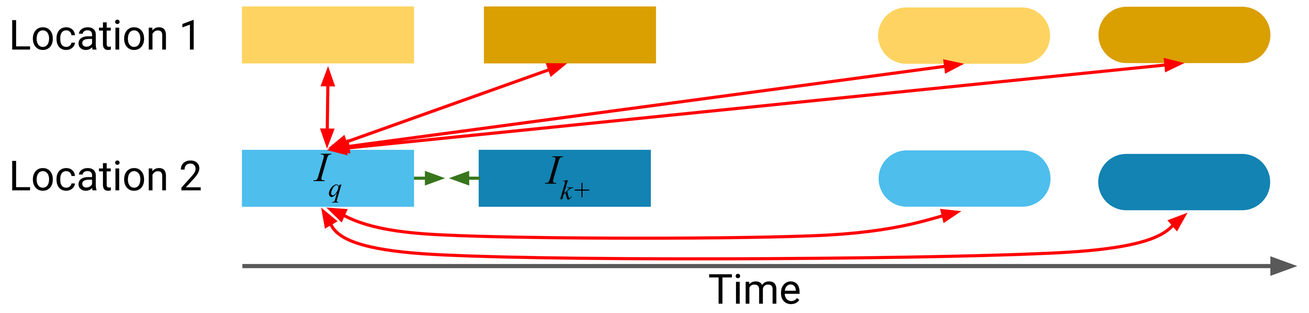
Seasonal Contrast (SeCo):
- Short-term temporal pairs are positives.
- Other locations are negatives.



Long-term temporal contrast

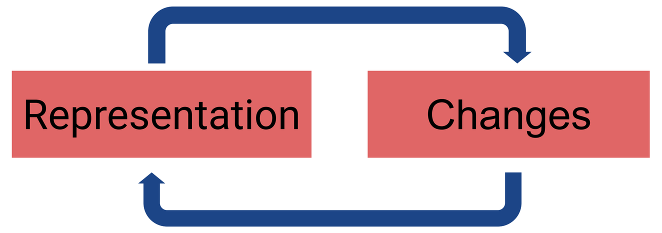
We *additionally* use images with *long time differences* as negatives during contrastive learning.

$$\mathcal{I}_{k-} = \{I_{l_i}^{t_2+\Delta_k} : k \in \{1,2\}\} \cup \{I_{l_h}^{t_j+\Delta_k} : h \neq i, j \in \{1,2\}, k \in \{1,2\}\}$$



Change Awareness

We *only* use images with long time differences as negatives *if locations have changed significantly*.

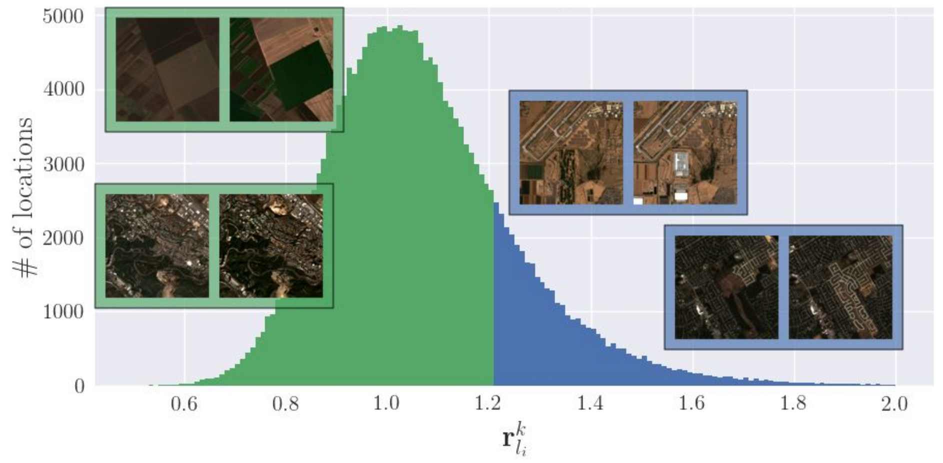


We simultaneously estimate changes and learn representation via-bootstrapping.

Change estimate at an intermediate step k:

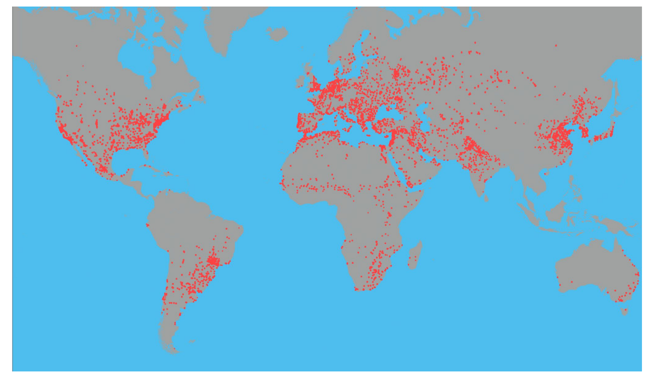
$$\mathbf{r}_{l_i}^k = (1-\beta)R_{l_i}^k + \beta \mathbf{r}_{l_i}^{k-1} \quad R_{l_i}^k = \frac{||f_k(I_{l_i}^{t_1+\Delta_1}) - f_k(I_{l_i}^{t_2+\Delta_1})||^2}{||f_k(I_{l_i}^{t_1+\Delta_1}) - f_k(I_{l_i}^{t_1+\Delta_2})||^2}$$

Histogram of change estimate with the threshold.

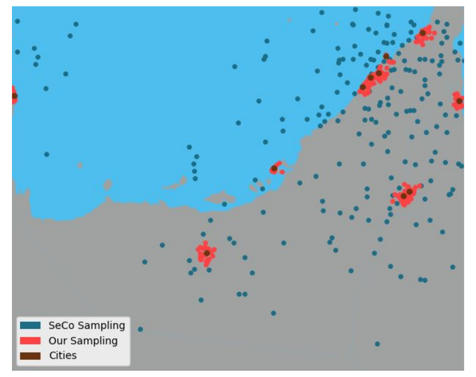


Geographical Sampling
- Gaussian sampling around urban areas.
- Reject samples falling into oceans to avoid repetitive images.

CACo samples



Samples on coastline



Results

Our method performs better at various downstream tasks such as land-cover classification, semantic segmentation, and change detection.

	EuroSat (Acc.)		BigEarthNet (mAP)		OSCD (F1)	Dynamic EarthNet (mIoU)
Pre-training (100k)	ResNet-18	Resnet-50	ResNet-18	Resnet-50	ResNet-18	Resnet-18
Random Init.	64.21	55.32	45.95	45.22	28.91	41.53
ImageNet	86.16	89.08	66.40	71.37	35.30	43.75
Moco v2	87.22	89.75	67.20	72.88	38.21	47.97
GSSL	87.74	90.19	67.36	72.86	44.06	46.77
SeCo	90.05	93.12	67.43	73.42	46.84	46.83
CACo	**93.08**	**94.48**	**69.43**	**73.63**	**50.29**	**50.20**

All our contributions lead to improved performance.

Pre-training (100k)	EuroSat (Acc.)	DynamicEarthNet (mIoU)
SeCo	90.05	46.83
+ geographical sampling	91.42	47.24
+ Longer temporal contrast	92.56	49.68
+ Change Awareness (Ours)	**93.08**	**50.20**

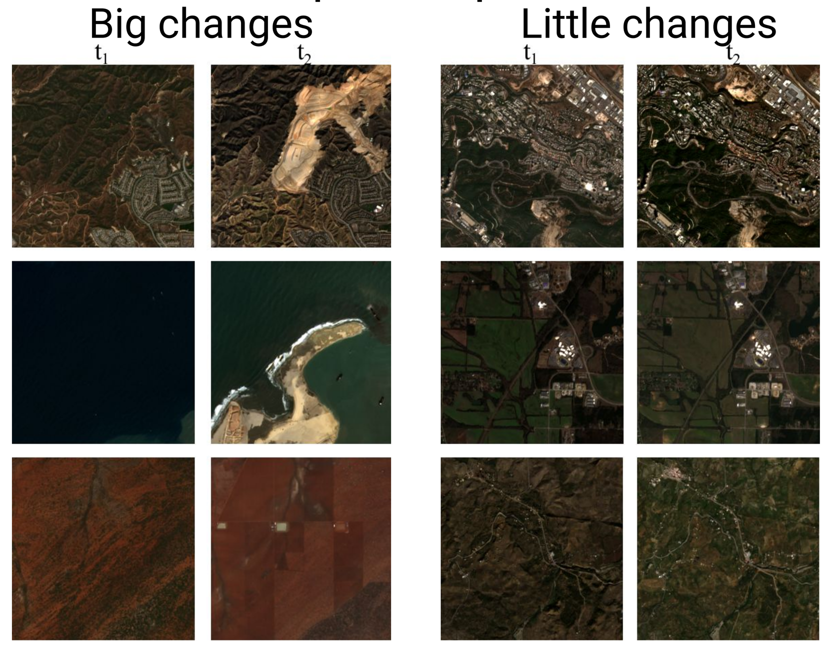
Our ratio estimate is better at understanding changes.

Pre-training (100k)	EuroSat (Acc.)
Long-term distance	91.17
R = Ratio (long-term/short term)	91.98
Expo. Moving Average (R)	**93.08**

Our method generalizes to other Self-supervised frameworks such as SimCLR

Pre-training (100k)	EuroSat (Acc.)	OSCD (F1)
SeCo	87.98	44.05
CACo	**90.83**	**47.50**

Examples of pairs with:



Takeaways

Spatio-temporal satellite images have **numerous unique properties** that can be leveraged to improve self-supervised learning.

All these properties lead to a better representation that is useful for **many downstream remote sensing tasks**.

References

[1] Mañas *et. al.*, Seasonal contrast: Unsupervised pre-training from uncurated remote sensing data. ICCV, 2021

Acknowledgment

This work was funded by NSF (1900783, 2144117, and 2212084) and IARPA (2021-20111000006).