**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.

- Geographical Sampling: We use an improved geographical sampling that provides more informative data for representation learning.
- Change Awareness: We present a novel approach to estimate changes that can be used to encourage invariance.
- Long-term temporal information: We propose a new loss using long-term temporal information in satellite images.

**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.

$$\mathcal{L} = - \log \frac{\exp(f(t_i) - f(t_j))}{\sum_{t_k \neq t_i} \exp(f(t_i) - f(t_k))}$$

**Long-term temporal contrast**

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

$$\mathcal{I}_{+} = \{I_{i,j}^{k+1} : k \in [1,2]\} \cup \{I_{i,j}^{k} : k \in [1,2], k \neq i, j \}$$

**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.

**Geographical Sampling**

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

**References**

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

**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.

**Acknowledgment**

This work was funded by NSF (1900783, 2144117, and 2210284) and IARPA (2021-11000006).