Tracking Climate Change Using Satellites and Artificial Intelligence

Shufan Xia[1], Frankie Chiappetta[2], Margaret Brooks[3], Alex Desbans[4], Neel Gajjar[3], Jules Kourelakos[3], Saad Lahrichi[5], Vaishvi Patel[4], Ada Zhang[3], Edna Zhang[3]

Advisors: Kyle Bradbury[4,6], Jordan Malof[7]

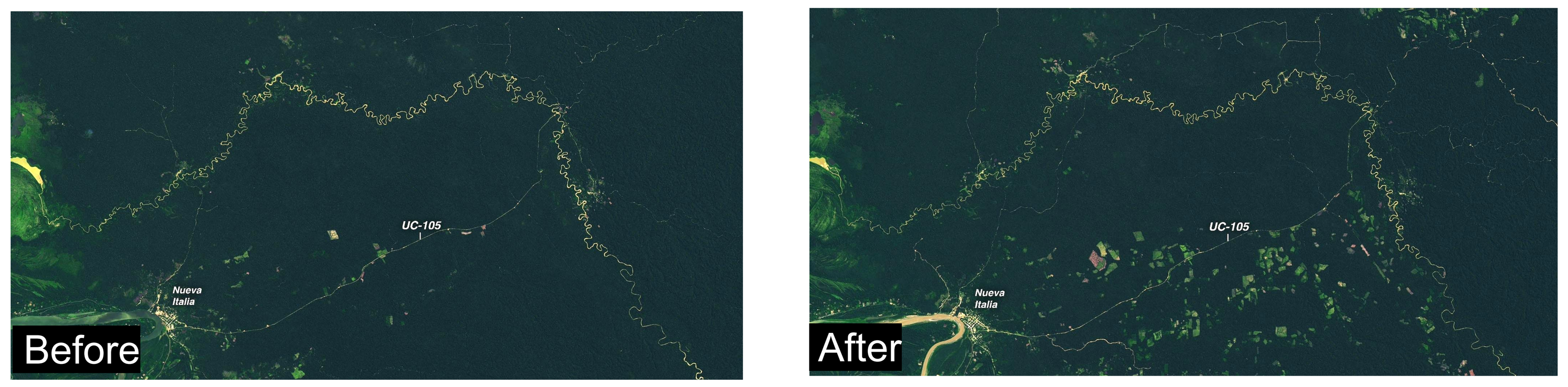
[1] The Graduate School, Duke University, [2] Nicholas School of the Environment, Duke University, [3] Trinity College, Duke University, [4] Pratt School of Engineering, Duke University, [5] Duke Kunshan University, [6] Duke University Nicholas Institute for Energy, Environment & Sustainability [7] Computer Science, University of Montana

Duke UNIVERSITY | BASS CONNECTIONS

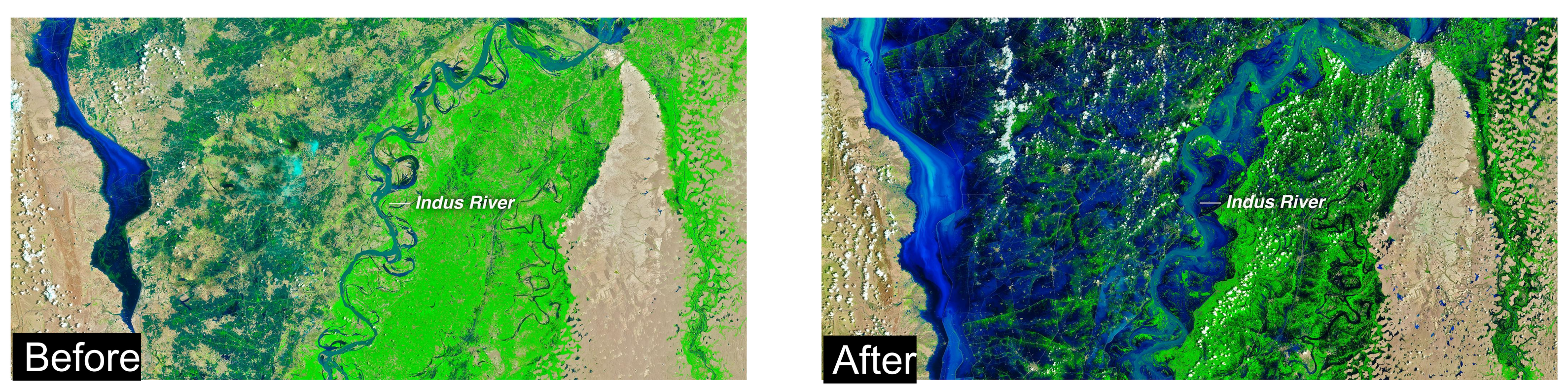
Energy & Environment

1 Motivation

There is growing availability of remote sensing imagery, allowing tracking of climate causes and impacts. Artificial intelligence can help extract information from these images at scale.



Satellite images showing **deforestation** (light green on right) near a road in Peru
https://climate.nasa.gov/images-of-change/?id=812#812-deforestation-near-nueva-italia-peru



Satellite images showing **catastrophic flooding** (blue on right) in Pakistan
https://earthobservatory.nasa.gov/images/150279/devastating-floods-in-pakistan

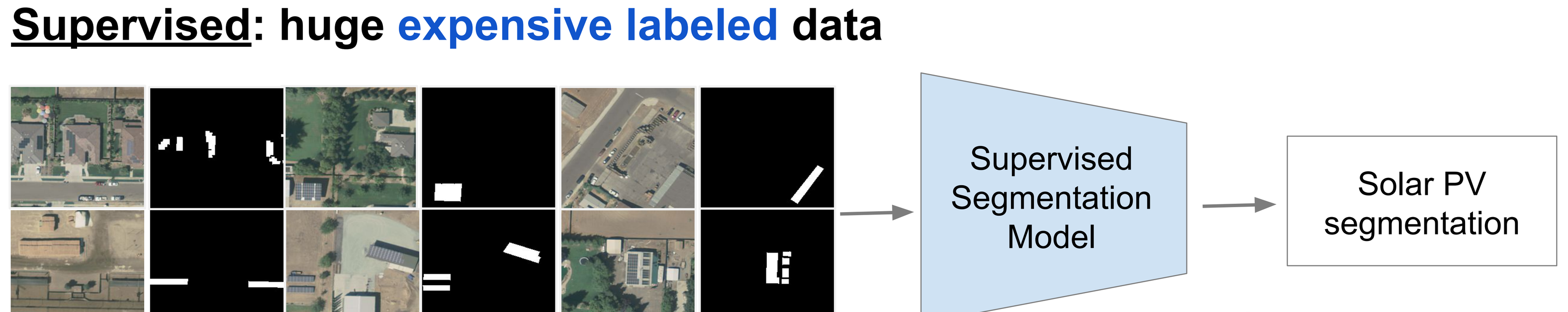
Challenges

- Data with labels (e.g. "river") are often unavailable / expensive
- Supervised learning, which uses labeled data, is difficult to scale up and apply across geographic regions without labeled data
- Existing state-of-the-art pre-trained models (trained on natural imagery) fail to adapt to unique characteristics of satellite images

2 Self-Supervised Learning (SSL)

SSL can use huge amounts of unlabeled data to learn, extracting robust image representations generalizable across geographic domains and tasks.

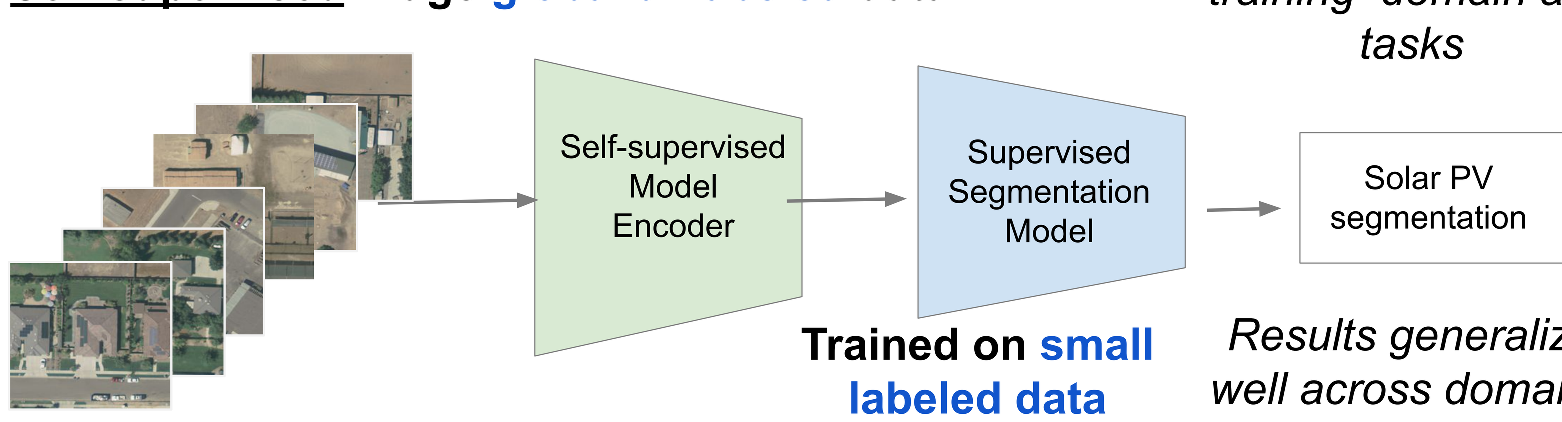
Supervised: huge **expensive labeled** data



(Bradbury et al., 2016)

Results generalize poorly outside training domain and tasks

Self-supervised: huge **global unlabeled** data



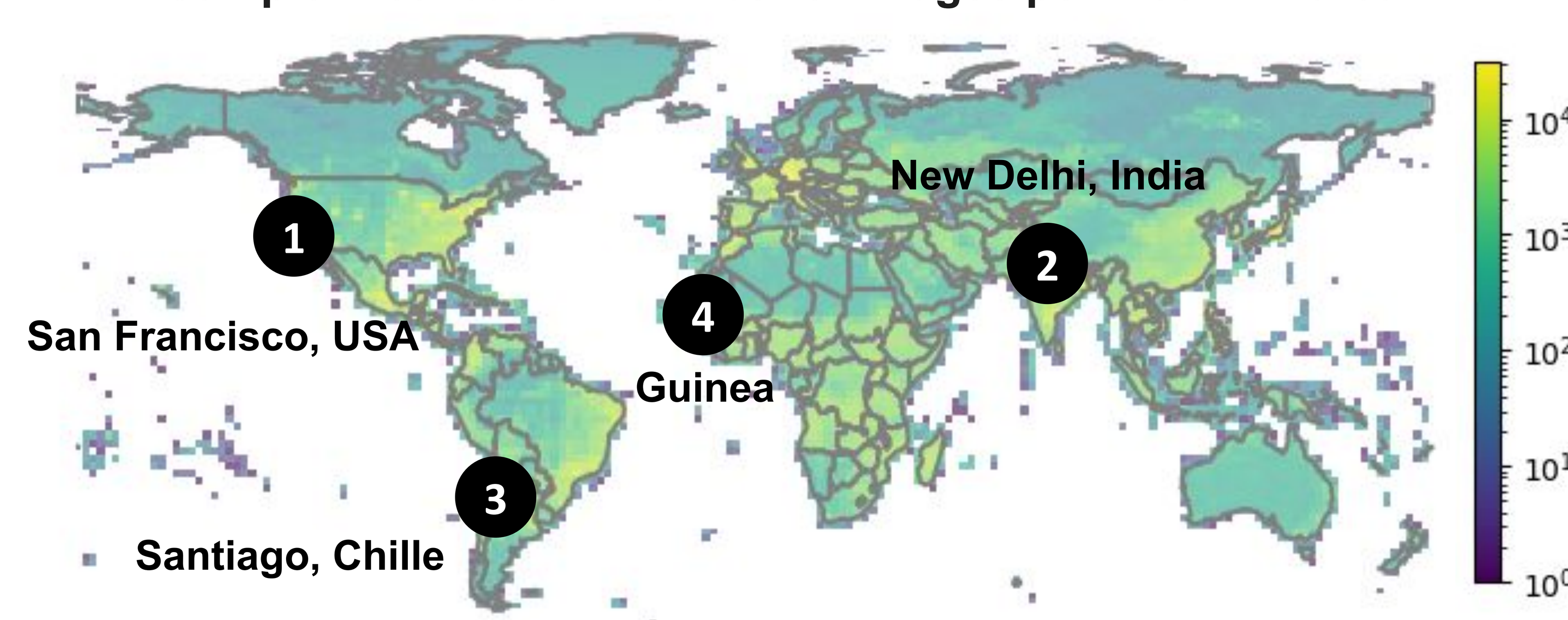
Trained on small labeled data

Results generalize well across domains and tasks

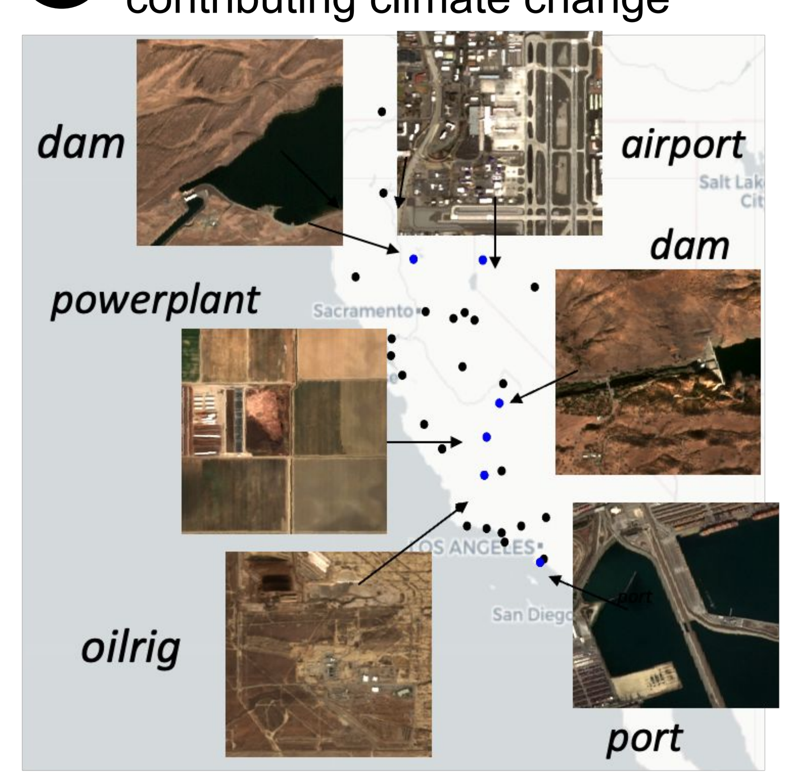
3 Creating a global dataset (GeoNet)

Self-supervised learning model performs best with a large and diverse dataset. We created this **10 million** satellite image dataset, GeoNet, the **first** and **largest** ever to capture geospatial, temporal, and semantic diversity for remote sensing data.

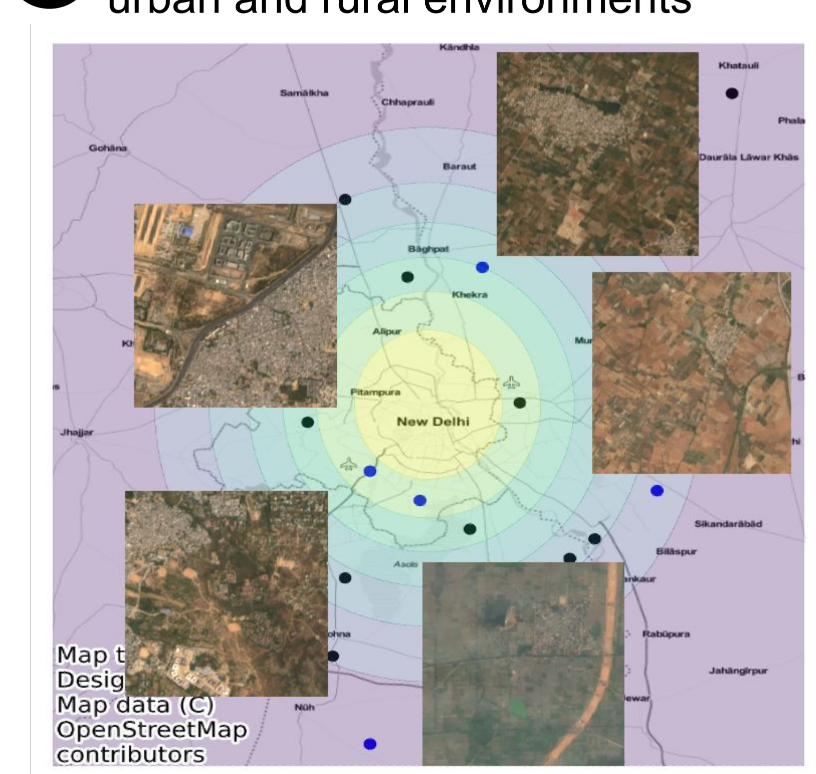
Sample Distribution: number of images per ~ 50 km^2 area



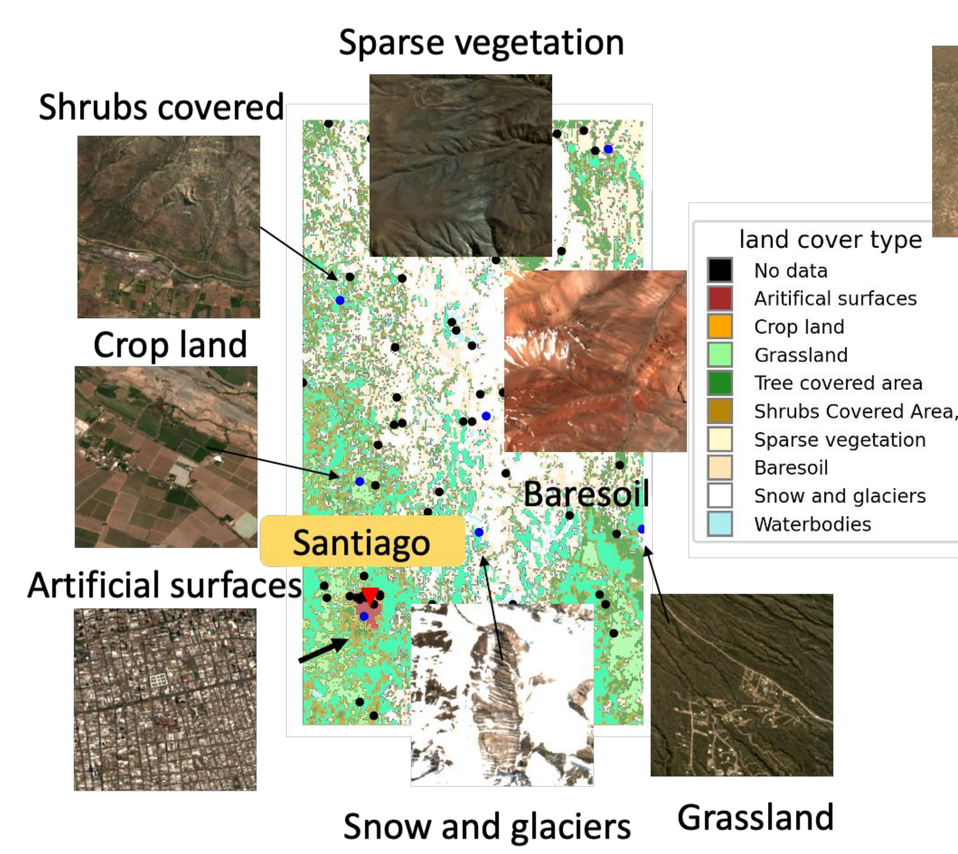
1 Targets built features contributing climate change



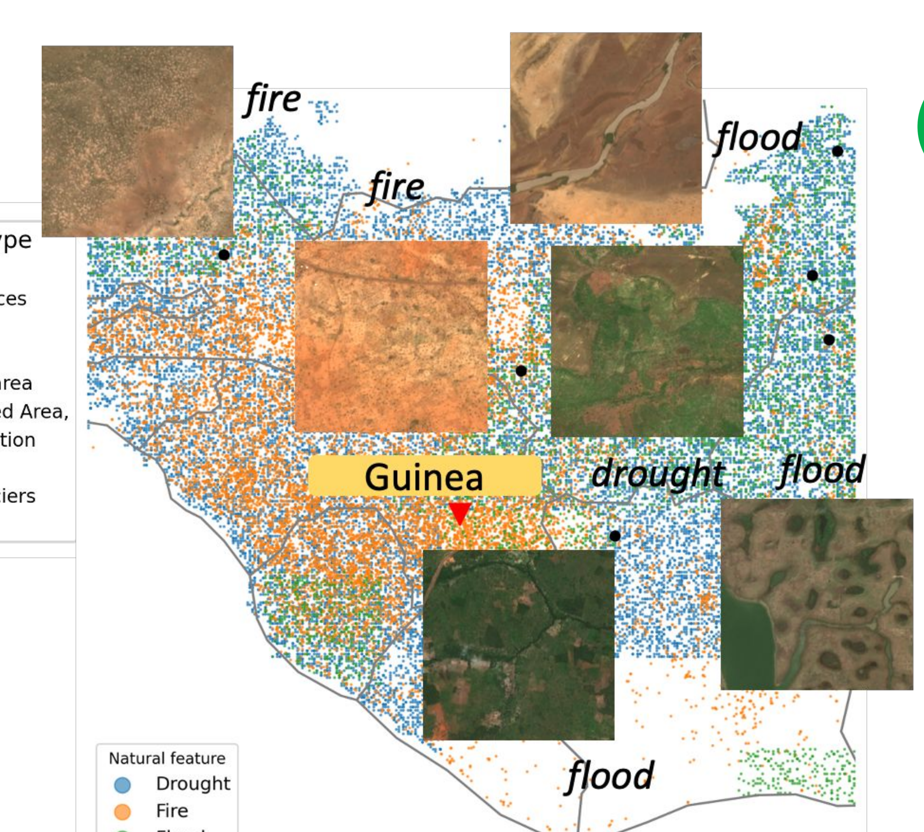
2 Diversely samples from both urban and rural environments



3 Represents diversity in land cover type



4 Samples areas with historical risks of natural disasters

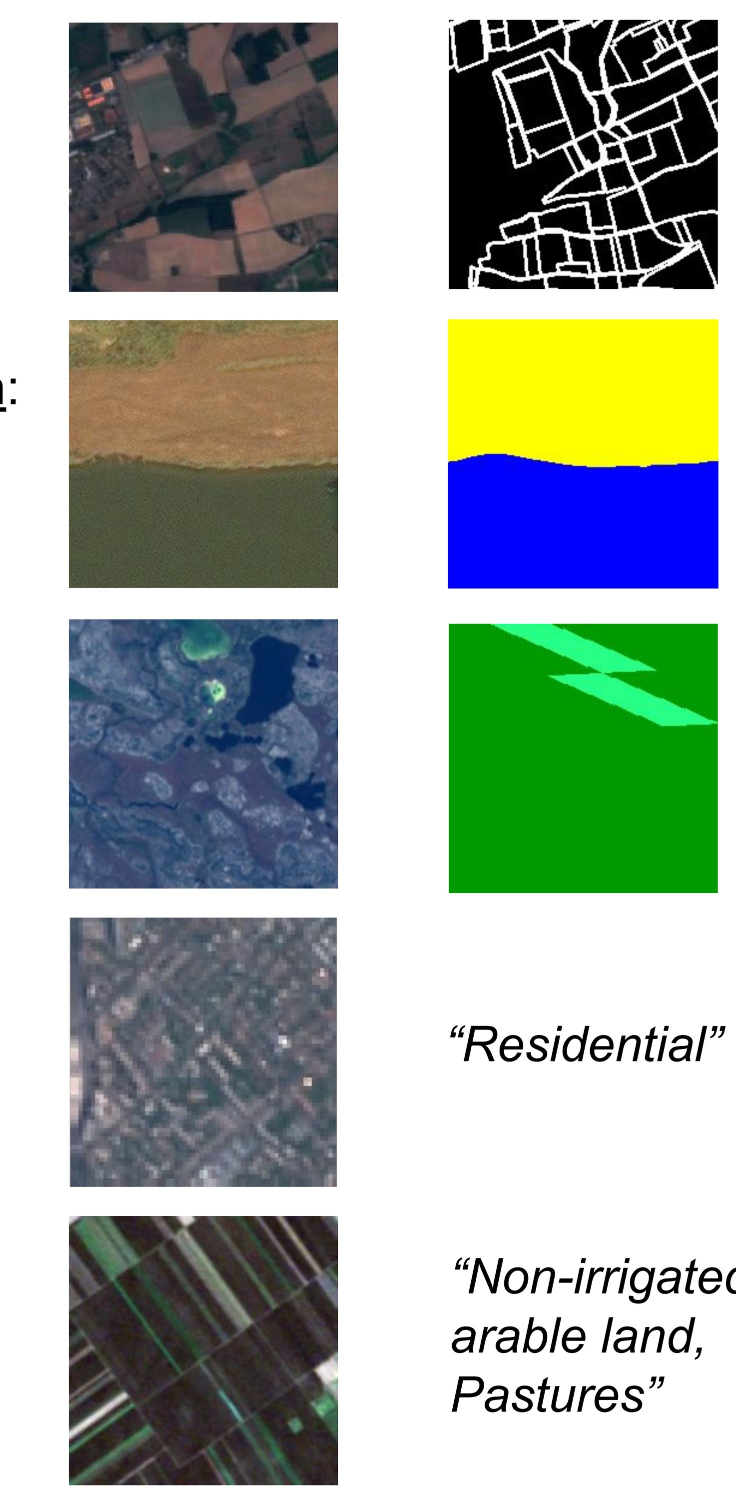


4 Applying SSL to inference tasks

We tested self-supervised pretraining on GeoNet with 5 benchmark datasets, and found our method outperforms competing methods for 3 out the 5 benchmark datasets.

Pretraining using the SSL technique requires no labels, making it a promising and enabling technology for scaling up analyses of global changes to the environment.

Benchmark datasets	Input Image	Prediction
Binary segmentation: SustainBench - field delineation *(Yeh et al., 2021)*		
Multi-class segmentation: Deepglobe - land cover classification *(Demir et al., 2018)*		
Multi-class "weak label" segmentation: SEN12MS *(Schmitt et al., 2019)*		
Multi-class classification: EuroSat - land cover classification *(Helber et al., 2019)*		*"Residential"*
Multi-label classification: BigEarthNet *(Sumbul et al., 2019)*		*"Non-irrigated arable land, Pastures"*



5 For more information and our full results, visit our website

https://bassconnections-edal-22-23.github.io/