In-Domain Representation Learning for Remote Sensing

Maxim Neumann, André Susano Pinto, Xiaohua Zhai, and Neil Houlsby

26.04.2020 ICLR - AI4ES

Google Research

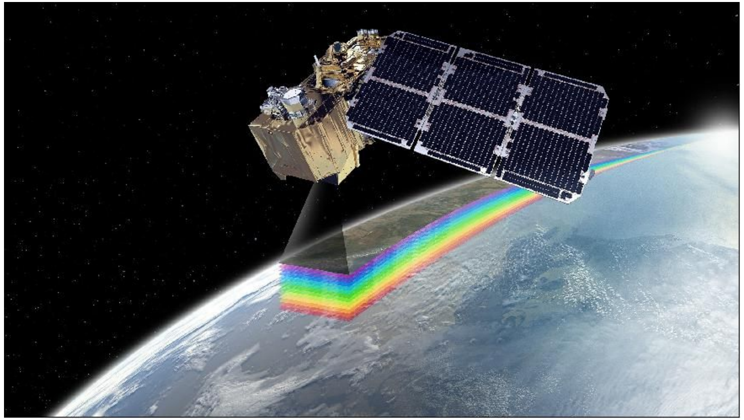
Remote Sensing in Earth Sciences

Earth Observation

- Monitoring, detection, and forecasting of key Earth Science quantities at regional and global scale

Example applications

- Carbon stocks and fluxes
- Deforestation and illegal logging
- Soil moisture monitoring
- Floods, landslides, fires, earthquakes
- . . .



Sentinel-2 Satellite

ESA's Earth observing satellite fleet

Google Research

AI & Representation Learning for Remote Sensing

Data rich

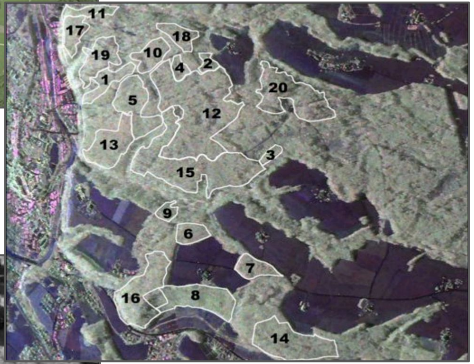
- Huge amounts of satellite data generated daily

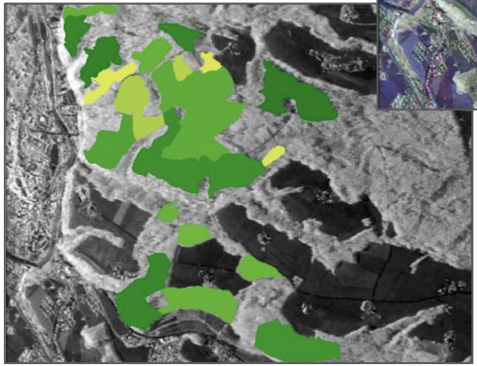
Label poor

- Ground-truth acquisition is expensive and often only limited number of in-situ data is available

→ Representation Learning

Optical and radar imagery over Traunstein Forest, Germany.





Remote sensing of forest structure.
Only 20 "ground truth" measurement labels.

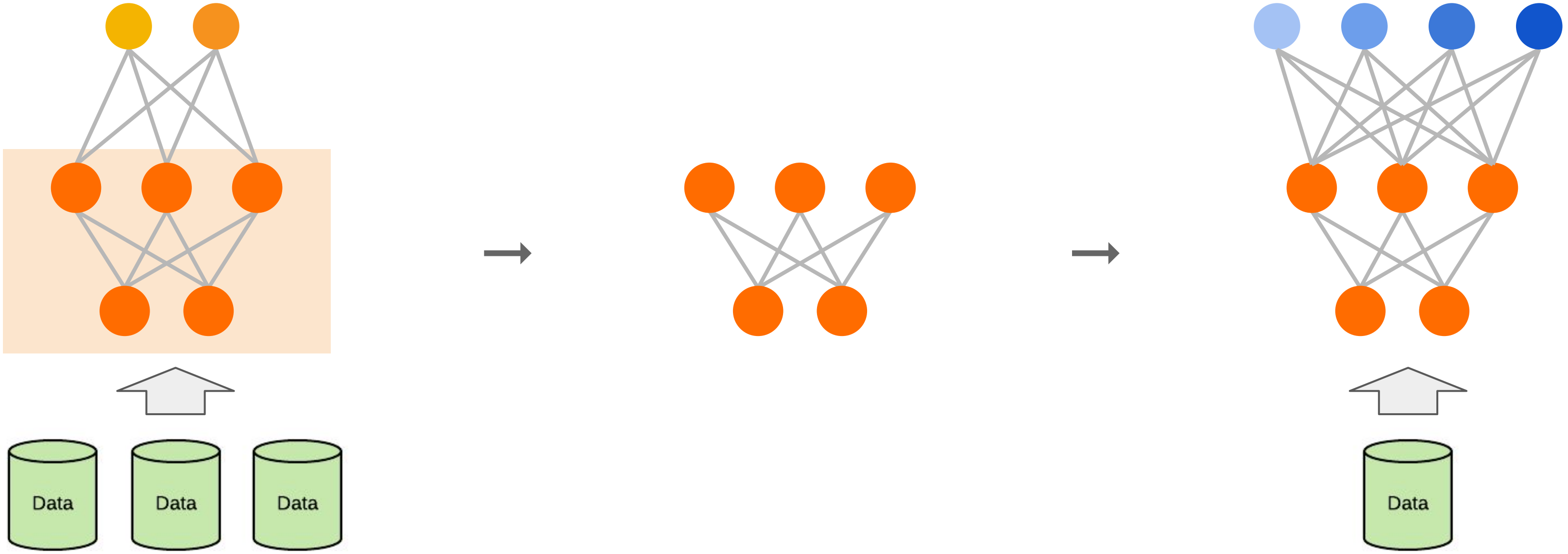
Google Research

Representation Learning

Upstream pre-training
Self/Semi/Un/Fully-supervised

Transferable representation

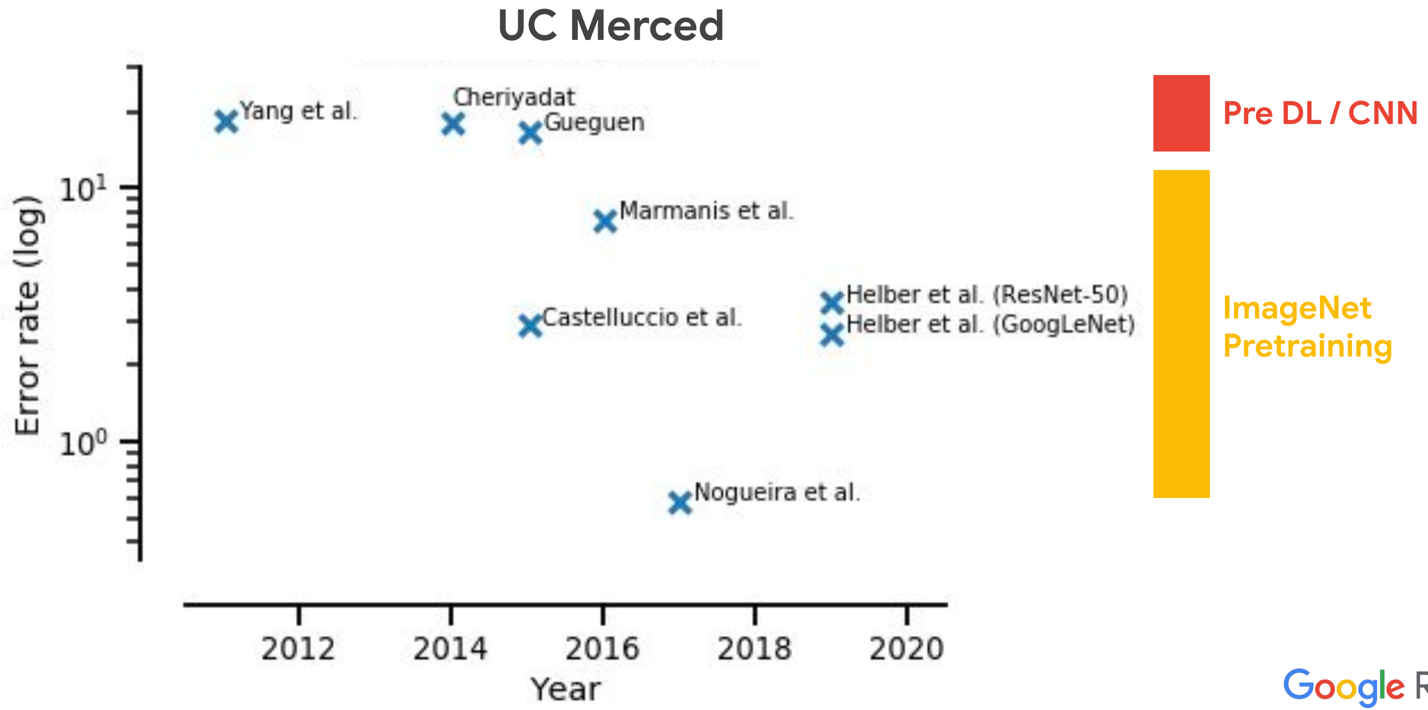
Downstream task
Frozen or fine-tune



Google Research

Representation Learning for Remote Sensing

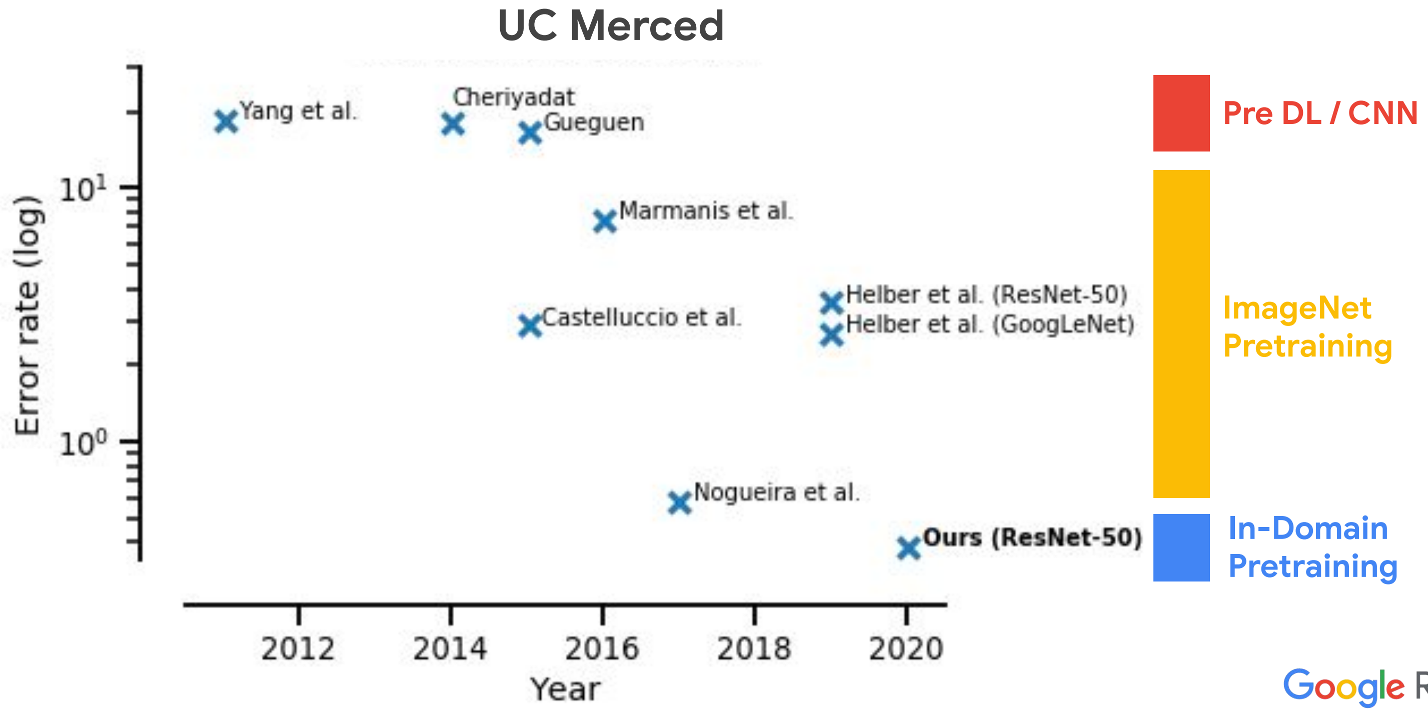
- Basic common approach: Use ImageNet weights



Google Research

Representation Learning for Remote Sensing

- Basic common approach: Use ImageNet weights
- Better: Include *in-domain* remote-sensing data during pre-training



Google Research

Goals

Evaluate representation learning for remote sensing

- Diverse set of data sources and tasks
- Enable faster research iteration by establishing baselines and common evaluation protocol

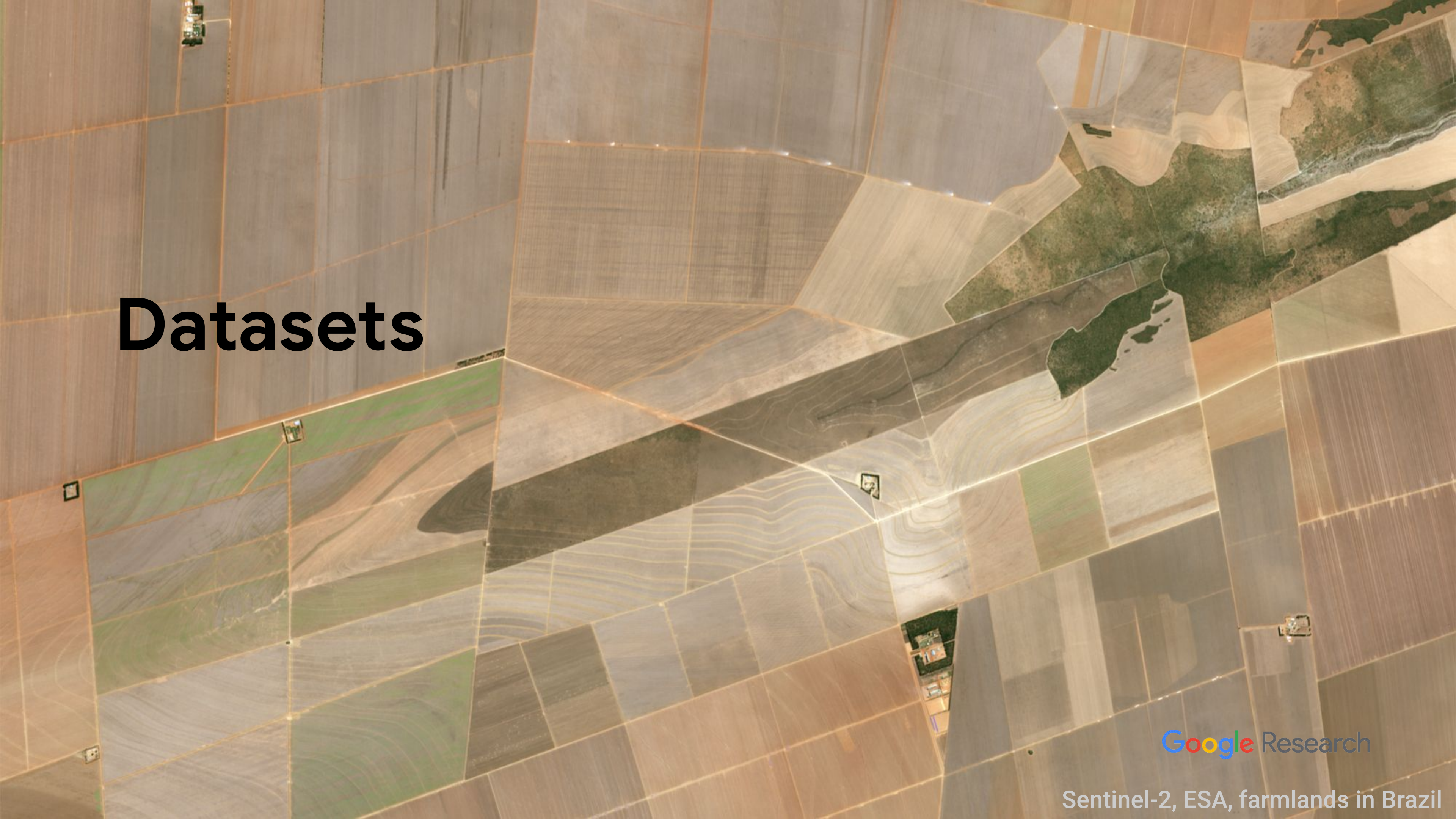
Explore in-domain representation learning

- Basic and efficient approach*:
  - Fully supervised, start from ImageNet, fine-tune entire model

* [A Large-scale Study of Representation Learning with the Visual Task Adaptation Benchmark](https://arxiv.org/abs/1910.04867)
X. Zhai et al. Submitted to ICML 2020. https://arxiv.org/abs/1910.04867

Google Research

Datasets



Sentinel-2, ESA, farmlands in Brazil

Datasets

- 5 datasets:
  - including RGB, multi-spectral, and synthetic aperture radar (SAR) channels.
  - medium and high resolution with variable image sizes - satellite and aerial.
- Train-val-test splits configuration:
  - All except So2Sat: 60%-20%-20% of the original available set.
  - So2Sat: full original train set, validation set is split 25%-75% for val and test splits.

Name	year	Source	Size	Classes	Image size	Resolution	Problem
BigEarthNet	2019	Sentinel-2	590k	43	120x120*	10–60 m	multi-label
EuroSAT	2019	Sentinel-2	27k	10	64x64	10 m	multi-class
RESISC–45	2017	aerial	31.5k	45	256x256	0.2–60+ m	multi-class
So2Sat	2019	Sentinel-1/2	376k	17	32x32	10 m	multi-class
UC Merced	2010	aerial	2.1k	21	256x256	0.3 m	multi-class

UC Merced Land-Use Dataset

"MNIST" of Remote Sensing.

- Aerial
- 100 images x 21 classes
- 256x256 RGB images
- 30 cm resolution

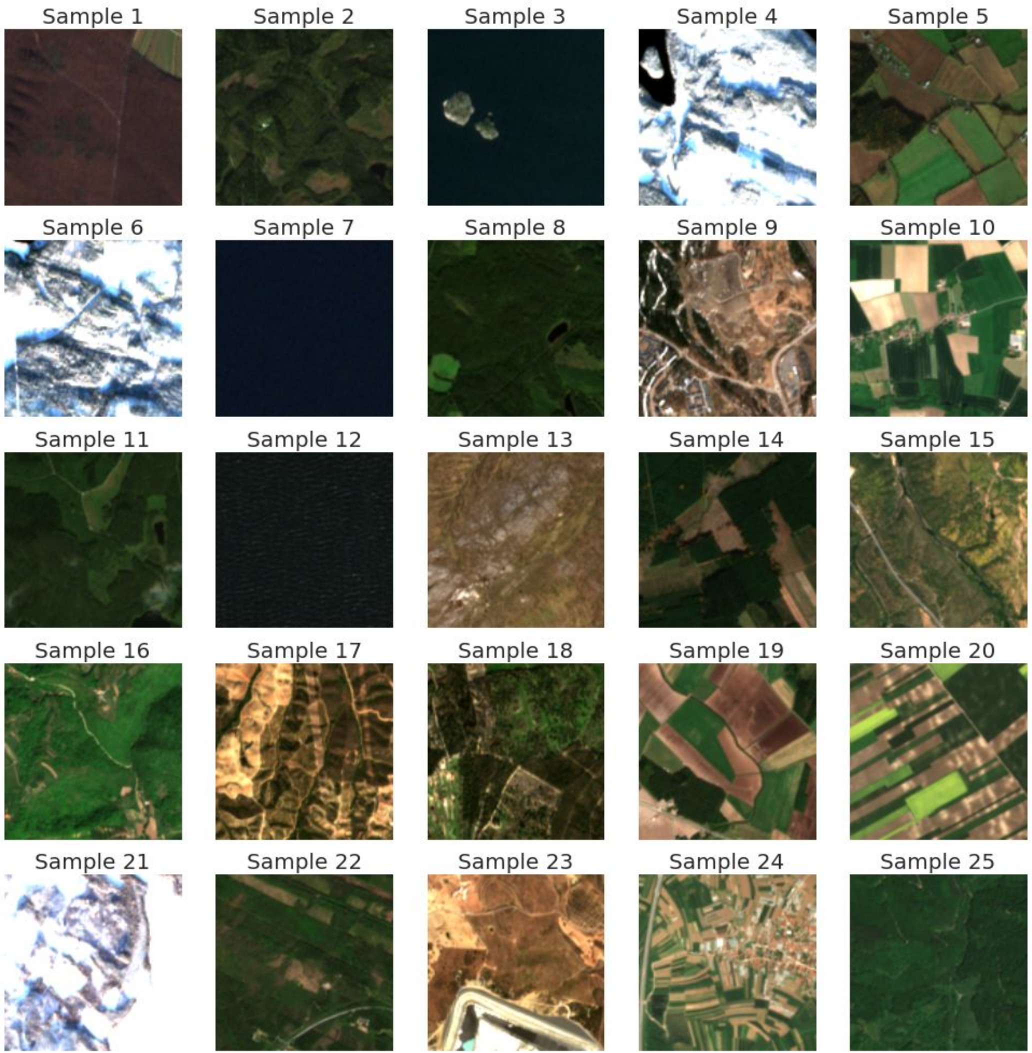
Y. Yang and S. Newsam, “Bag-of-visual-words and spatial extensions for land-use classification,” in Proceedings of the 18th SIGSPATIAL International Conference on Advances in GIS, New York, New York, USA, 2010, p. 270, ACM Press.

BigEarthNet

Large-scale medium resolution dataset.

- Sentinel-2 satellite, including multi-spectral channels
- 490k examples across 43 classes
- Multi-label
- About 12% of the patches are covered by clouds/snow

G. Sumbul, M. Charfuelan, B. Demir, and V. Markl, “BigEarthNet: A Large-Scale Benchmark Archive For Remote Sensing Image Understanding,” in IGARSS, Yokohama, Japan, Jul 2019, pp. 5901–5904.



Experiments and Results

1. Which in-domain data is good for pretraining?
2. Training from scratch vs. ImageNet vs. in-domain
3. Small training sizes regime
4. Comparison to published work

TanDEM-X PolInSAR satellites, DLR

In-Domain Pretraining

Downstream Tasks (1k samples)

Upstream Pretraining Data

	BigEarthNet	**EuroSAT**	**RESISC-45**	**So2Sat**	**UC Merced**
ImageNet	25.10	96.84	84.89	53.69	99.02
BigEarthNet	-	96.45	78.43	50.91	**99.61**
EuroSAT	27.10	-	79.59	52.99	98.05
RESISC-45	**27.59**	**97.14**	-	**54.43**	**99.61**
So2Sat	26.30	96.30	77.70	-	97.27
UC Merced	26.86	96.73	**85.73**	53.52	-

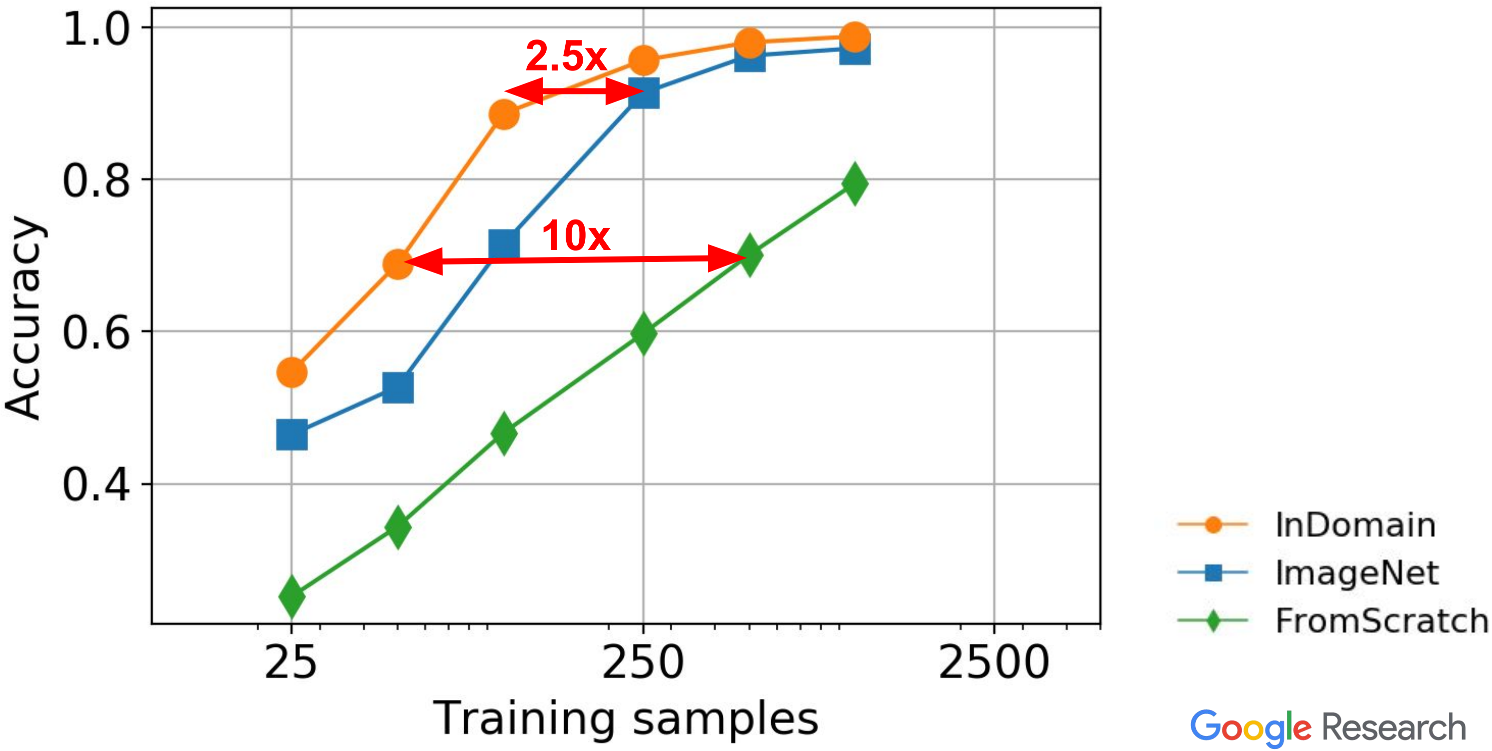
Training from Scratch vs. ImageNet vs. InDomain

Downstream Tasks (100, 1000 samples or Full)

Upstream Source	**BigEarthNet**			**EuroSAT**			**RESISC-45**			**So2Sat**			**UC Merced**		
	100	**1k**	**Full**	**100**	**1k**	**Full**	**100**	**1k**	**Full**	**100**	**1k**	**Full**	**100**	**1k**	**Full**
Scratch	14.5	21.4	72.4	63.9	91.7	98.5	21.4	56.1	95.6	33.9	47.0	62.1	50.8	91.2	95.7
+ImageNet	17.8	25.1	**75.4**	87.3	96.8	99.1	44.9	84.9	96.6	44.9	53.7	63.1	79.9	99.9	99.2
+In-Domain	**18.8**	**27.6**	69.7	**91.3**	**97.1**	**99.2**	**49.0**	**85.7**	**96.8**	**46.4**	**54.4**	**63.2**	**91.0**	**99.6**	**99.6**

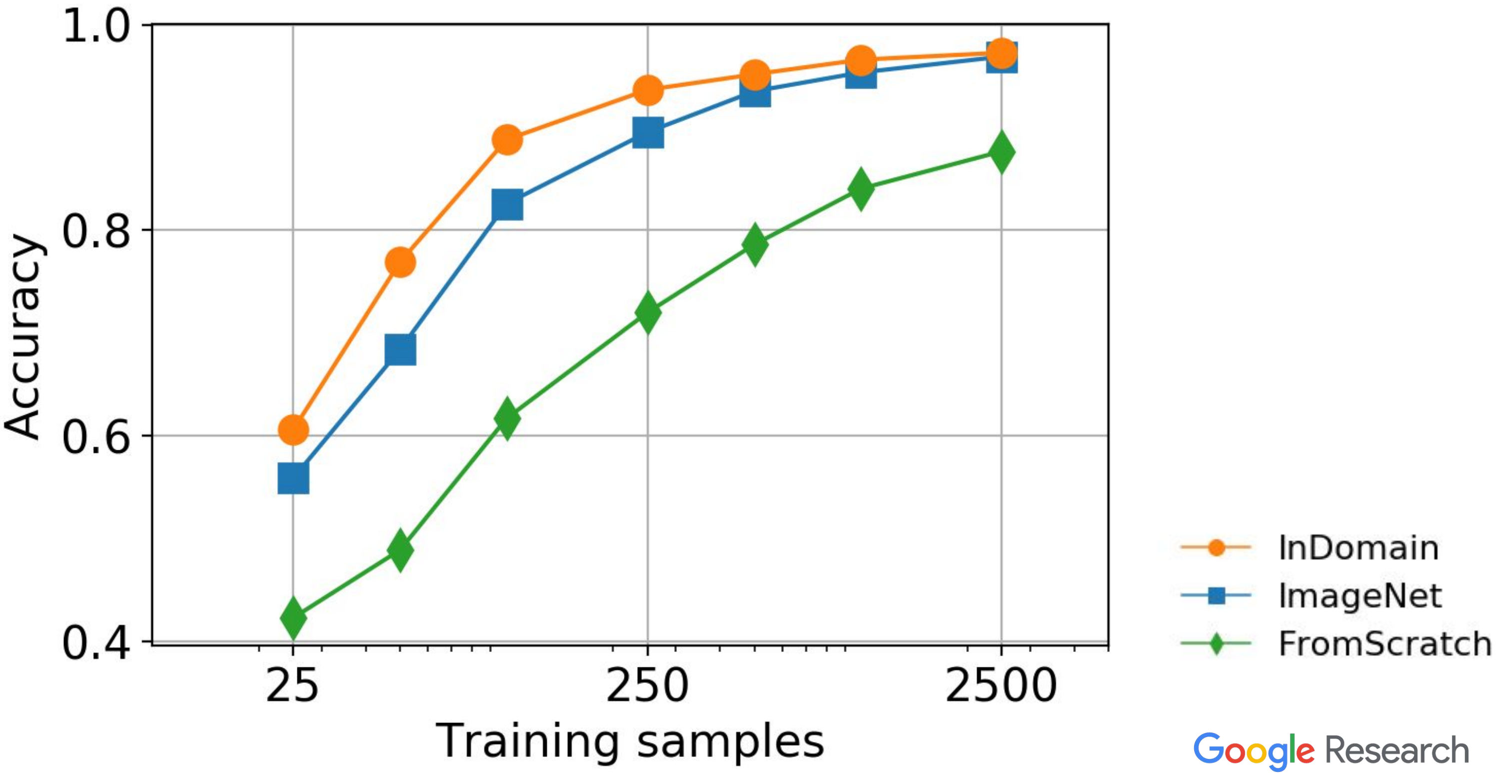
- Accuracy: **From Scratch << ImageNet < InDomain**
- Pretraining **most important at smaller tasks.**

Low number of training samples
UC Merced



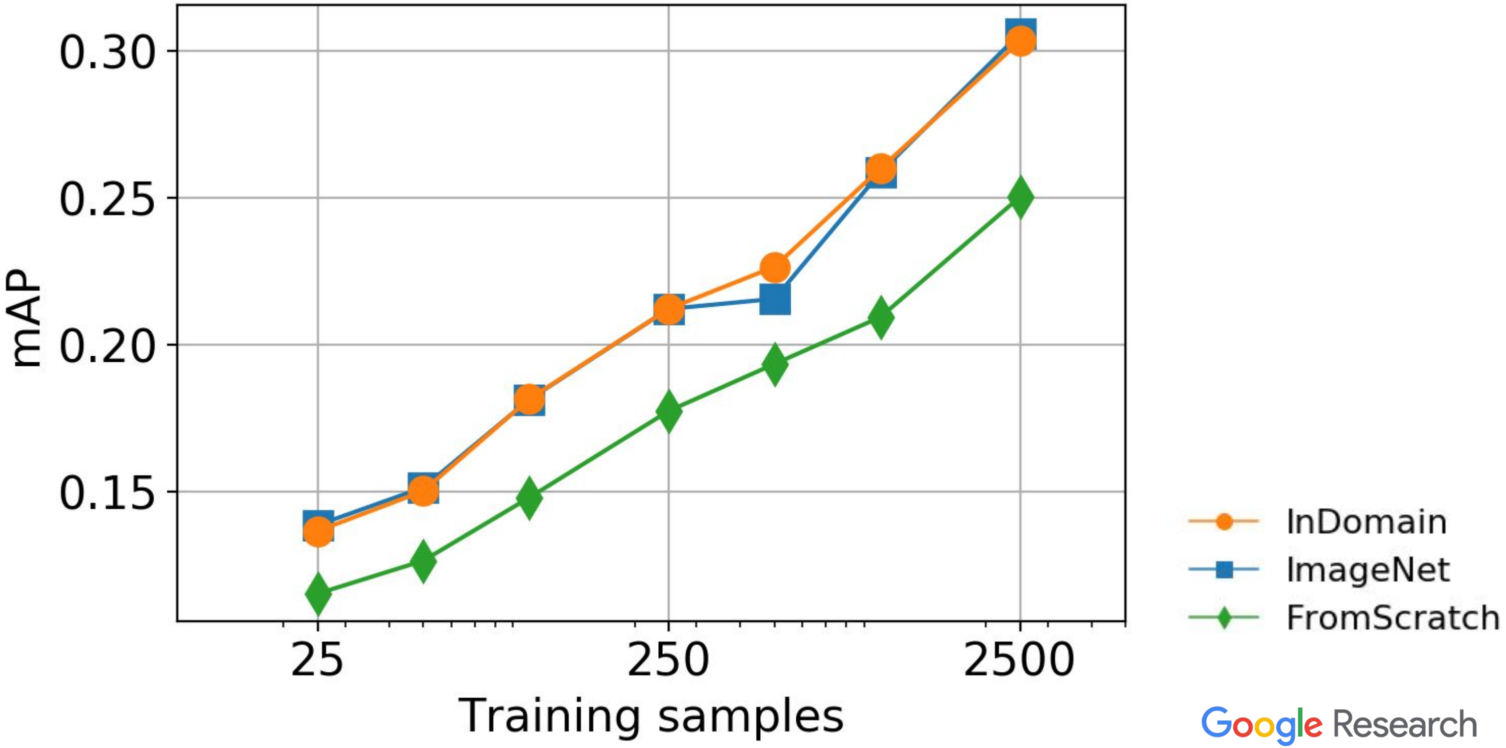
Google Research

Low number of training samples
EuroSAT



Google Research

Low number of training samples
BigEarthNet



Comparison to Published Results

Dataset	Reference	Result
BigEarthNet	Sumbul et al. (2019)	69.93% / 77.1% (P/R)
	Ours	75.36% (mAP)
EuroSAT	Helber et al. (2019)	98.57%
	Ours	**99.20%**
RESISC–45	Cheng et al. (2017)	90.36%
	Ours	**96.83%**
So2Sat	Ours	63.25%
UC Merced	Yang & Newsam (2010)	81.19%
	Marmanis et al. (2016)	92.4%
	Castelluccio et al. (2015)	97.1%
	Nogueira et al. (2017)	99.41%
	Ours	**99.61%**

Conclusion

- Pre-training using in-domain data can improve accuracy
  - especially for tasks with *limited number of samples* with ground-truth information
  - but *depends* as well on the *quality, quantity, and relevance of pretraining data*

- We established benchmarking baselines
  - SOTA on at least 3 of the 5 datasets

- Providing better RS representations helps many downstream tasks

- Open-sourcing links:
  - Datasets in common TensorFlow format:
    - https://www.tensorflow.org/datasets/catalog
  - Train-validation-test splits:
    - https://github.com/google-research/google-research/tree/master/remote_sensing_representations
  - Pretrained representations in TensorFlow Hub:
    - https://tfhub.dev/google/collections/remote_sensing/1

Thank You

In-Domain Representation Learning for Remote Sensing

Maxim Neumann, André Susano Pinto, Xiaohua Zhai, and Neil Houlsby

26.04.2020 ICLR - AI4ES

Google Research

EuroSAT

Medium-sized satellite dataset.

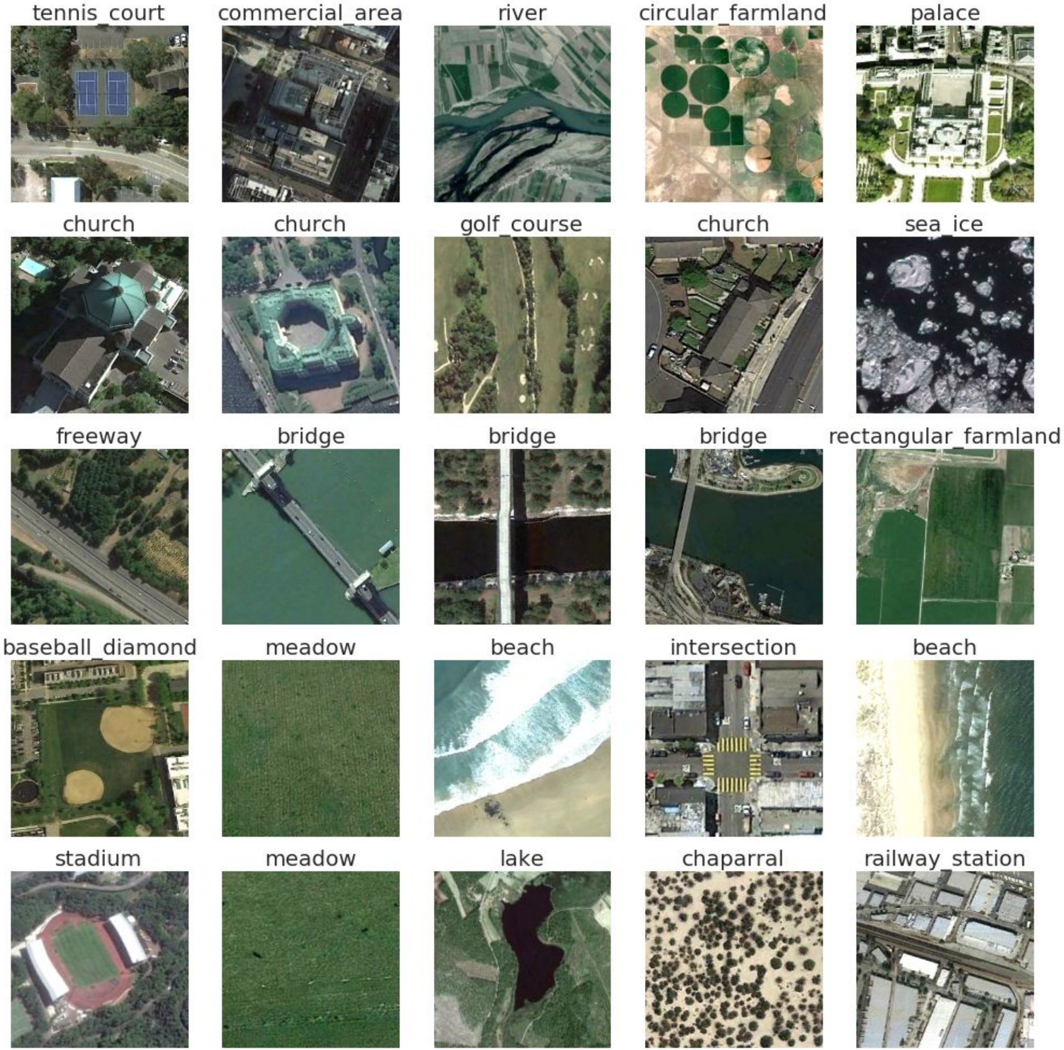
- Sentinel-2 satellite, including multi-spectral channels
- 27k examples over 10 classes
- 10 m resolution per pixel

P. Helber, B. Bischke, A. Dengel, and D. Borth, "Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 12, no. 7, pp. 2217–2226, July 2019.

NWPU RESISC–45

Multi-resolution curated dataset.

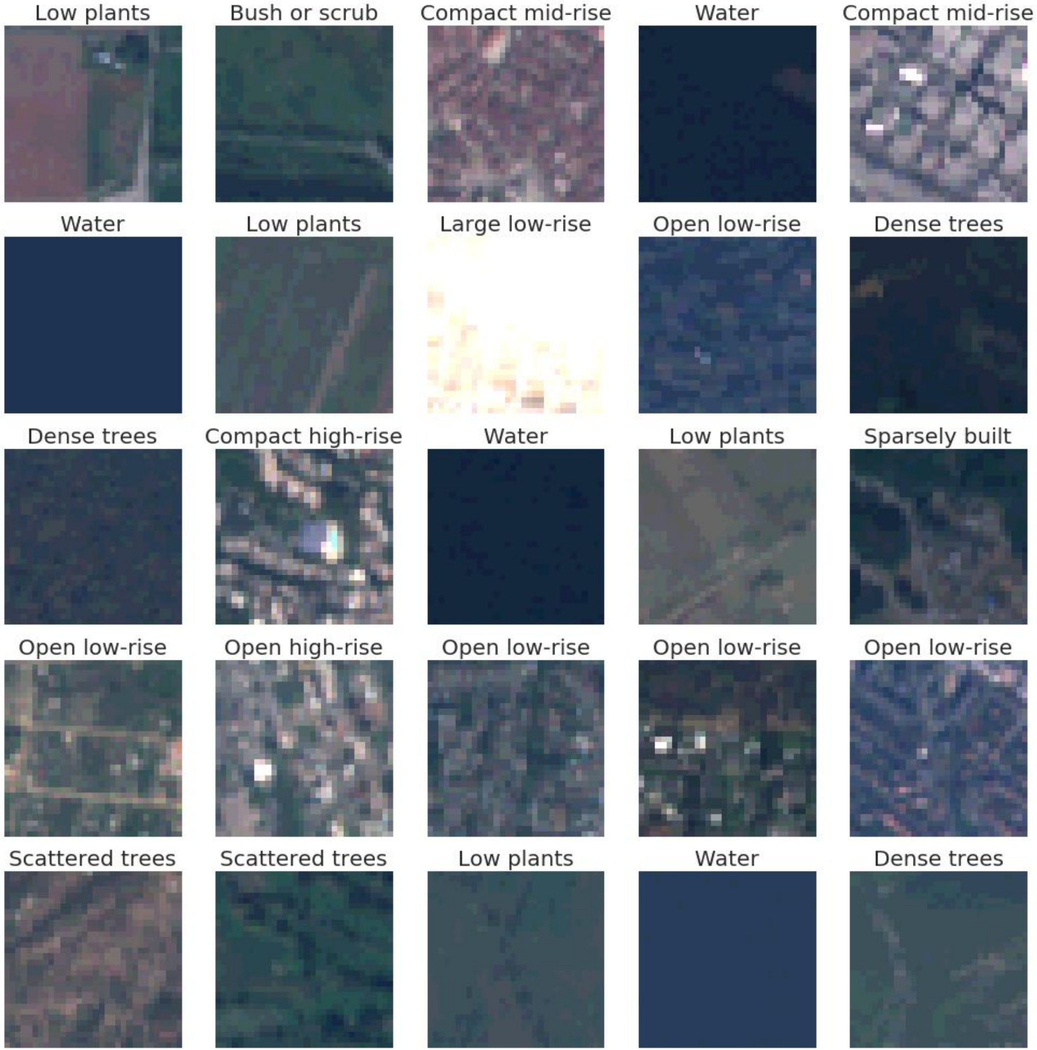
- Source: Aerial
- 31.5k examples over 45 scene classes
- 256x256 image sizes
- Varying spatial resolution from 20 cm to over 60 m



G. Cheng, J. Han, and X. Lu, “Remote sensing image scene classification: Benchmark and state of the art,” Proceedings of the IEEE, vol. 105, no. 10, pp. 1865–1883, Oct. 2017.

So2Sat LCZ-42

Optical, multi-spectral, and synthetic aperture radar (SAR) dataset.



X. Zhu, J. Hu, C. Qiu, Y. Shi, H. Bagheri, J. Kang, H. Li, L. Mou, G. Zhang, M. Haberle, S. Han, Y. Hua, R. Huang, L. Hughes, Y. Sun, M. Schmitt, and Y. Wang, “So2Sat LCZ42,” 2018.