Large Scale Masked Autoencoding for Reducing Label Requirements on SAR Data

Matt Allen, Francisco Dorr, Joseph A. Gallego-Mejia, Laura Martínez-Ferrer, Freddie Kalaitzis, Raúl Ramos-Pollán, Anna Jungbluth

UNIVERSITY OF CAMBRIDGE

esa Φ-lab

FDL 2023 EUROPE

Introduction

Satellite-based remote sensing is instrumental in the monitoring and mitigation of the effects of anthropogenic climate change. The timeliness and accuracy of interventions based on optical data, which cannot operate at night and is affected by adverse weather conditions, are limited. Synthetic Aperture Radar (SAR) offers an all-weather, day night alternative, but labelled data generation is complex. We apply **masked autoencoding** to **SAR data** covering **8.7% of Earth's land surface**, and show that:

(1) **Label requirements on downstream tasks were reduced, often** by an order of magnitude or more for the same performance.

(2) The performance **gain was greater when tuning on regions outside the pretraining set.**

Synthetic Aperture Radar (SAR)

SAR has two key advantages over RGB imaging:

- **Cloud Penetration** - SAR operates at microwave wavelengths, which can penetrate turbid mediums such as clouds or dust.
- **Day/Night Operation** - SAR sensors illuminate terrain actively, independently of lighting conditions or the day-night cycle.

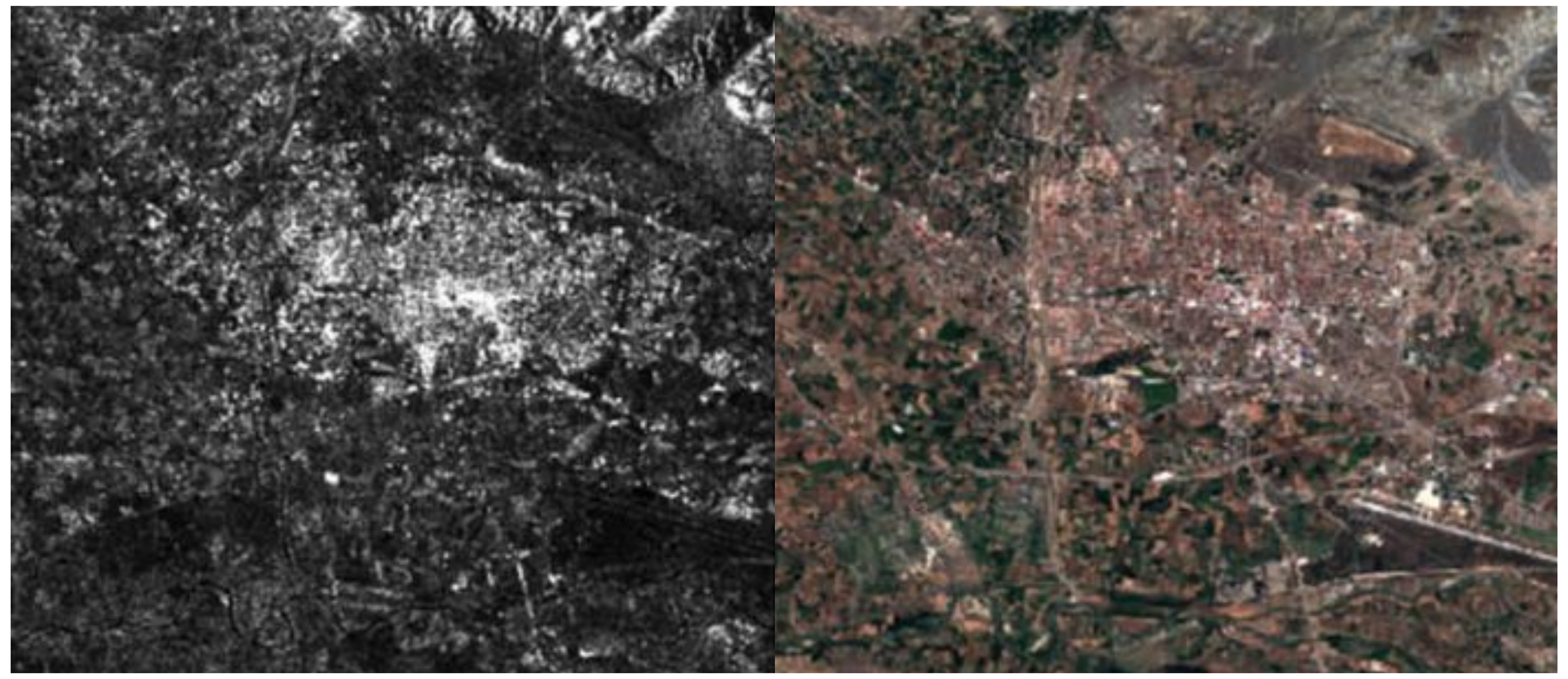

Fig 1: *Polarimetric SAR (Left) and RGB Imagery (Right)*

Data

Our data comprises **4 areas of interest** (AOIs) - China, the Continental United States (CONUS), Europe and South America - divided geographically into training, validation and test sets. The **pretraining set comprises China, CONUS and Europe only.**

Input Data comprises Sentinel-1 GRD [1] polarimetry (amplitude) data, 10m/pixel

Downstream Tasks comprise TERRA MODIS [2] vegetation prediction (regression per-image) and ESA world cover [3] land cover classification (semantic segmentation, 10 classes)

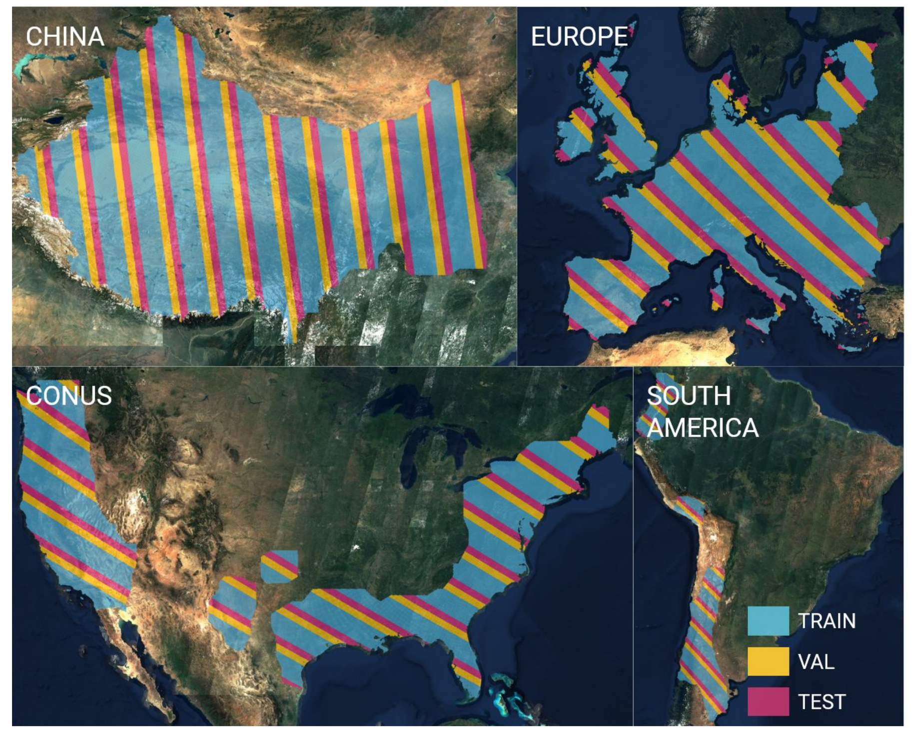

Fig 2: *Train/validation/test split indicated by coloured bands. Training data from China, CONUS and Europe make up the pretraining set. Models were evaluated downstream on data from Europe and South America. Coverage determined by intersection with ARIA GUNW interferograms [4] (unused).*

Model

Masked autoencoder [5] with ViT-B [6] encoder with 12 channels - per-season average of VV, VH and VV-VH polarisations. Patch size 16, 448x448 images.

Results

Performance was **improved by pretraining in all cases**. The fully supervised, randomly initialised models were often **outperformed by a pretrained model on 1 or 10% of the labelled data**.

Results

Vegetation Prediction

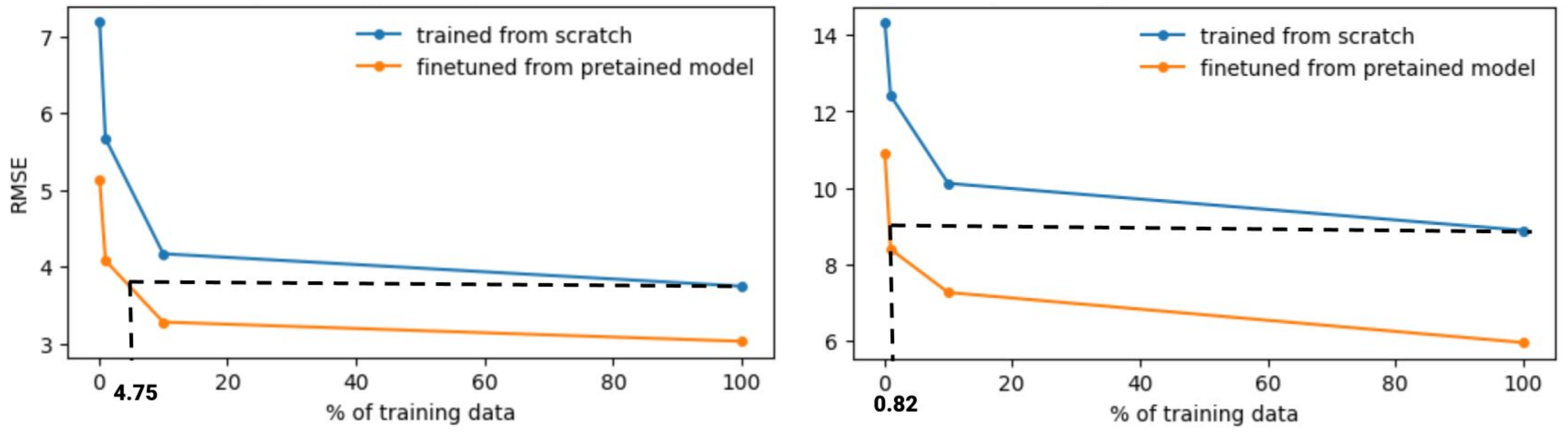

Fig 3: *Labelled data ablation for **vegetation prediction** on **Europe (Left)** and **S. America (Right).** Reported using RMSE.*

Land Cover Classification

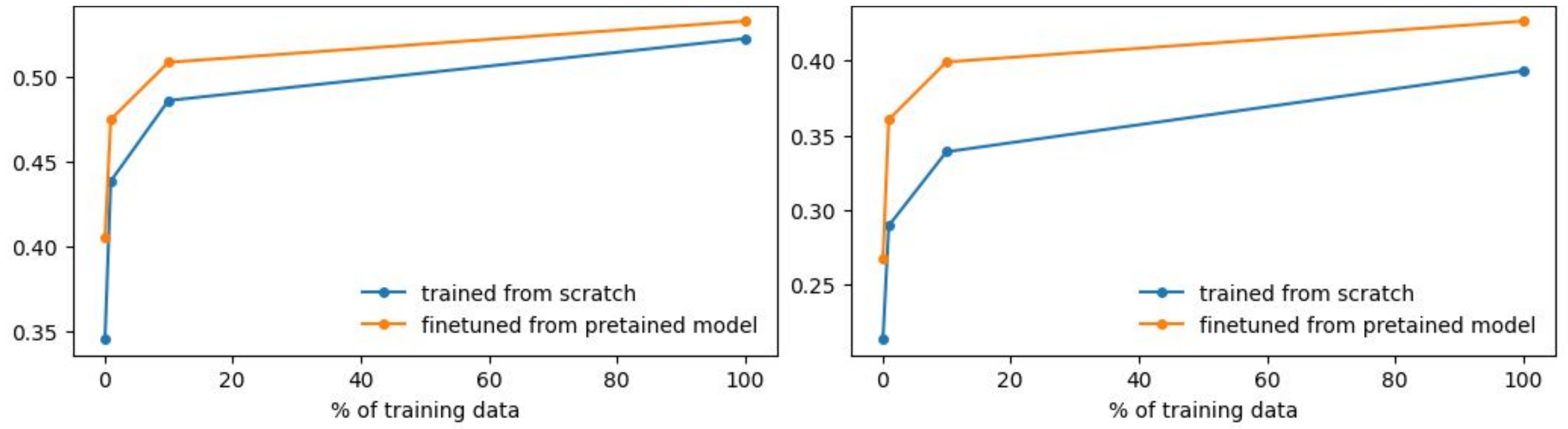

Fig 4: *Labelled data ablation for **land cover classification** on **Europe (Left)** and **S. America (Right).** Reported using mIOU.*

References

[1] https://developers.google.com/earth-engine/datasets/cataog/COPERNICUS_S1_GRD

[2] developers.google.com/earth-engine/datasets/catalog/MODIS_006_MOD44B

[3] developers.google.com/earth-engine/datasets/catalog/ESA_WorldCover_v200

[4] https://asf.alaska.edu/data-sets/derived-data-sets/sentinel-1-interferograms

[5] *Masked Autoencoders are Scalable Vision Learners*, He et al., 2021

[6] *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*, Dosovitskiy et al., 2021

esa Φ-lab · FDL 2023 EUROPE · Google Cloud · NVIDIA · SCAN AI · Centre for Satellite Data in Environmental Science · PASTEUR iSi · AIRBUS · The Danish Meteorological Institute · D-ORBIT · CATAPULT Satellite Applications · UNIVERSITY OF OXFORD · TRILLIUM EUROPE