A Distribution Shift Benchmark for Smallholder Agroforestry: Do Foundation Models Improve Geographic Generalization?

Siddharth Sachdeva[1], Isabel Lopez[1], Chandrasekar Biradar[2], David Lobell[1]
Stanford University[1], World Agroforestry Centre[2]
siddsach@stanford.edu

ICLR 2024 ML4RS Workshop

Motivation

- Recent work has reported high accuracy in applying deep learning to individual tree detection
- Performance drops under distribution shifts are common for these approaches

We introduce the first distribution shift benchmark dataset for remote sensing tree detection and ask:
(1) How does performance drop under geographic distribution shift?
(2) Do foundation models improve robustness

Methods

- We perform three types of evaluations:
  (1) Conventional
  (2) Distribution shift
  (3) Few-shot domain adaptation
- Compare performance of a baseline (Faster-RCNN), with computer vision foundation models (SAM, Grounding DINO)
- We report per-zone accuracy metrics and compare performanance of our model with a recent tree cover product released for India

Dataset

- stratify Rajasthan into 8 agro-climatic zones
- sample train & test images for each zone
- annotate individual tree masks for each image
- split In-Distribution (ID) & Out-of-Distribution (OOD) zones for distribution shift evaluation
- do QA using inter-annotator agreement & field inventories

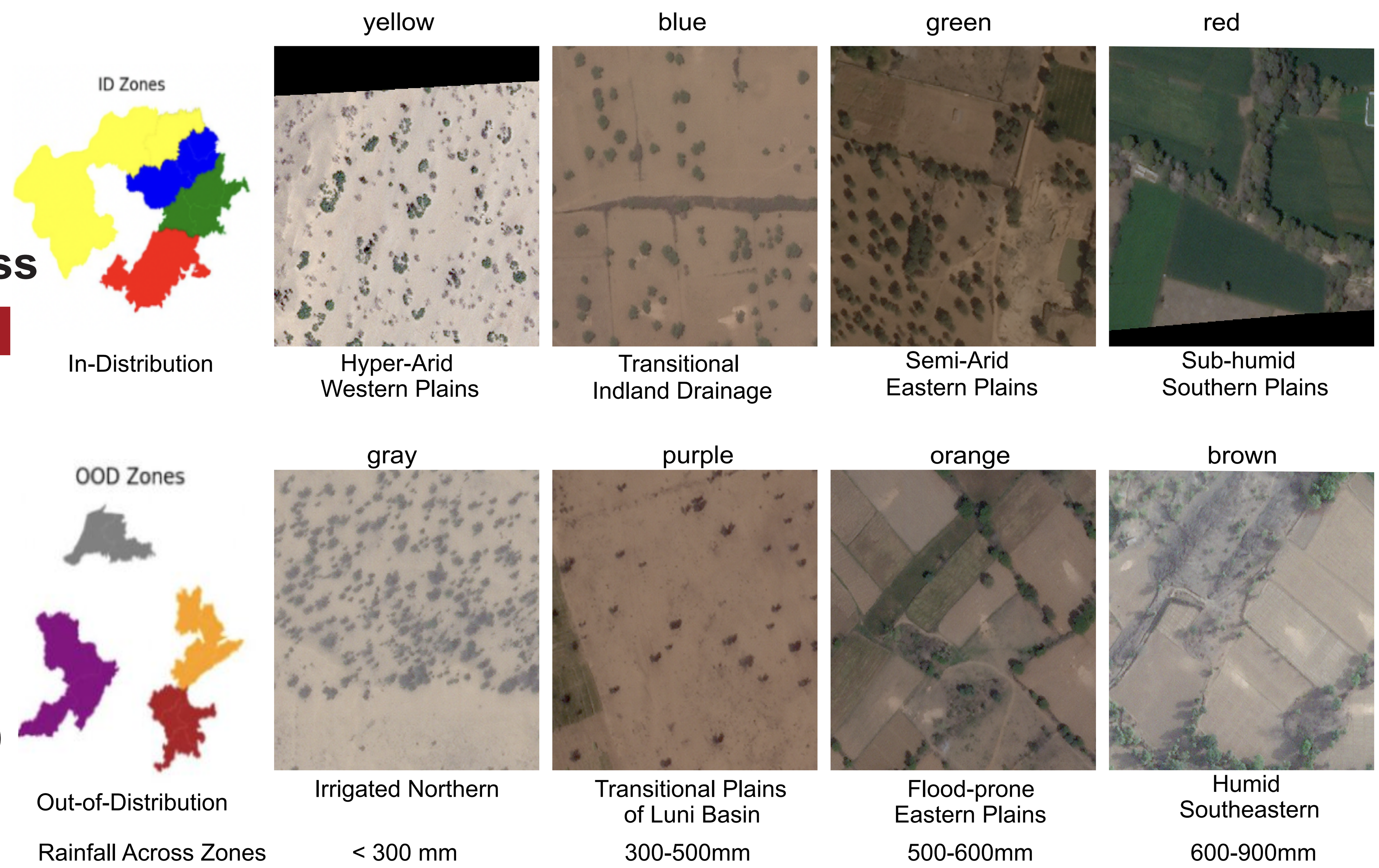

Figure 1. Example Images from different Agro-climatic Zones

Results

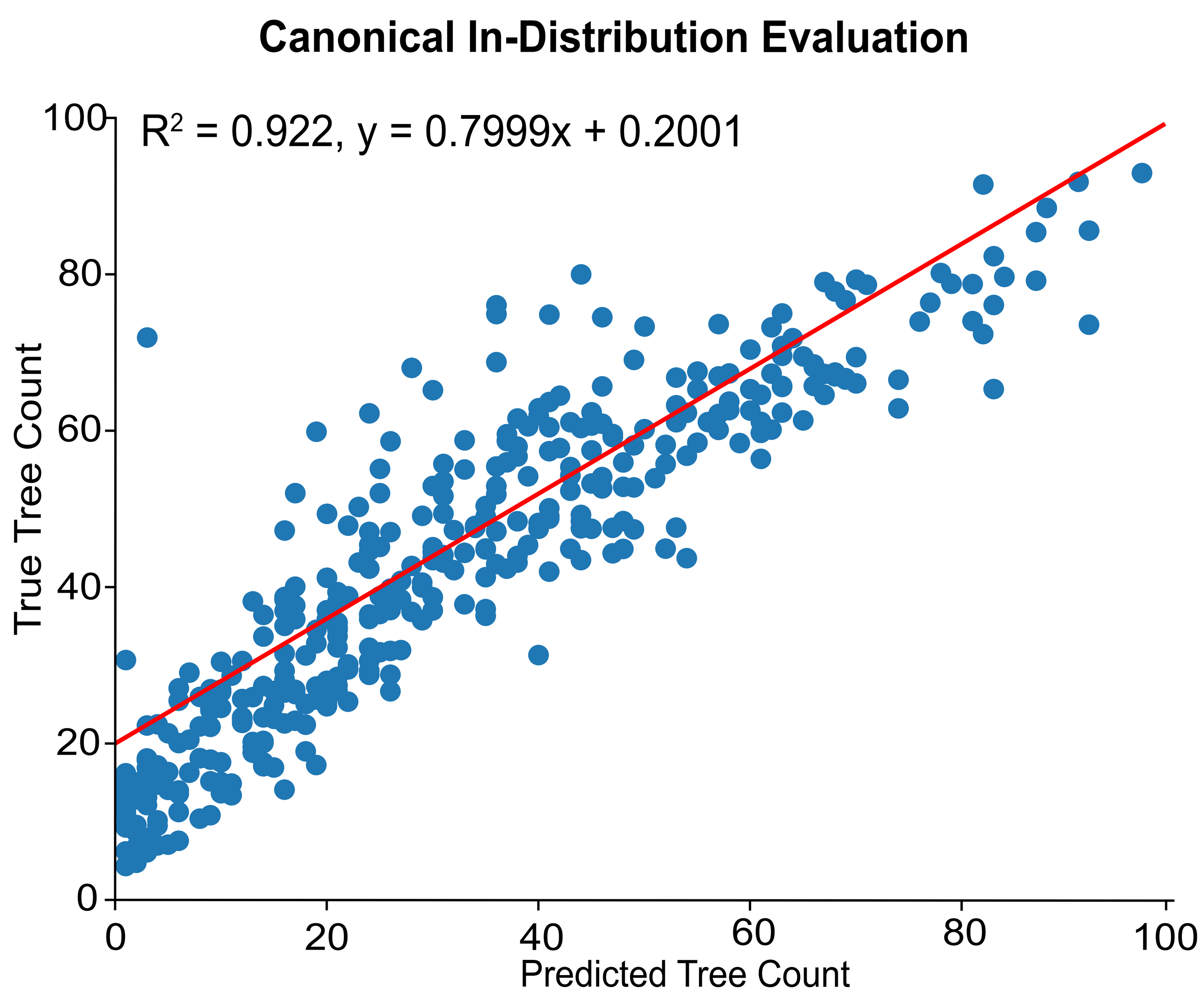

Figure 2. Predicted v.s True Tree count on In-Distribution Test Set, each point is image-level tree count in a 400x400 pixel image with 0.5-m resolution.

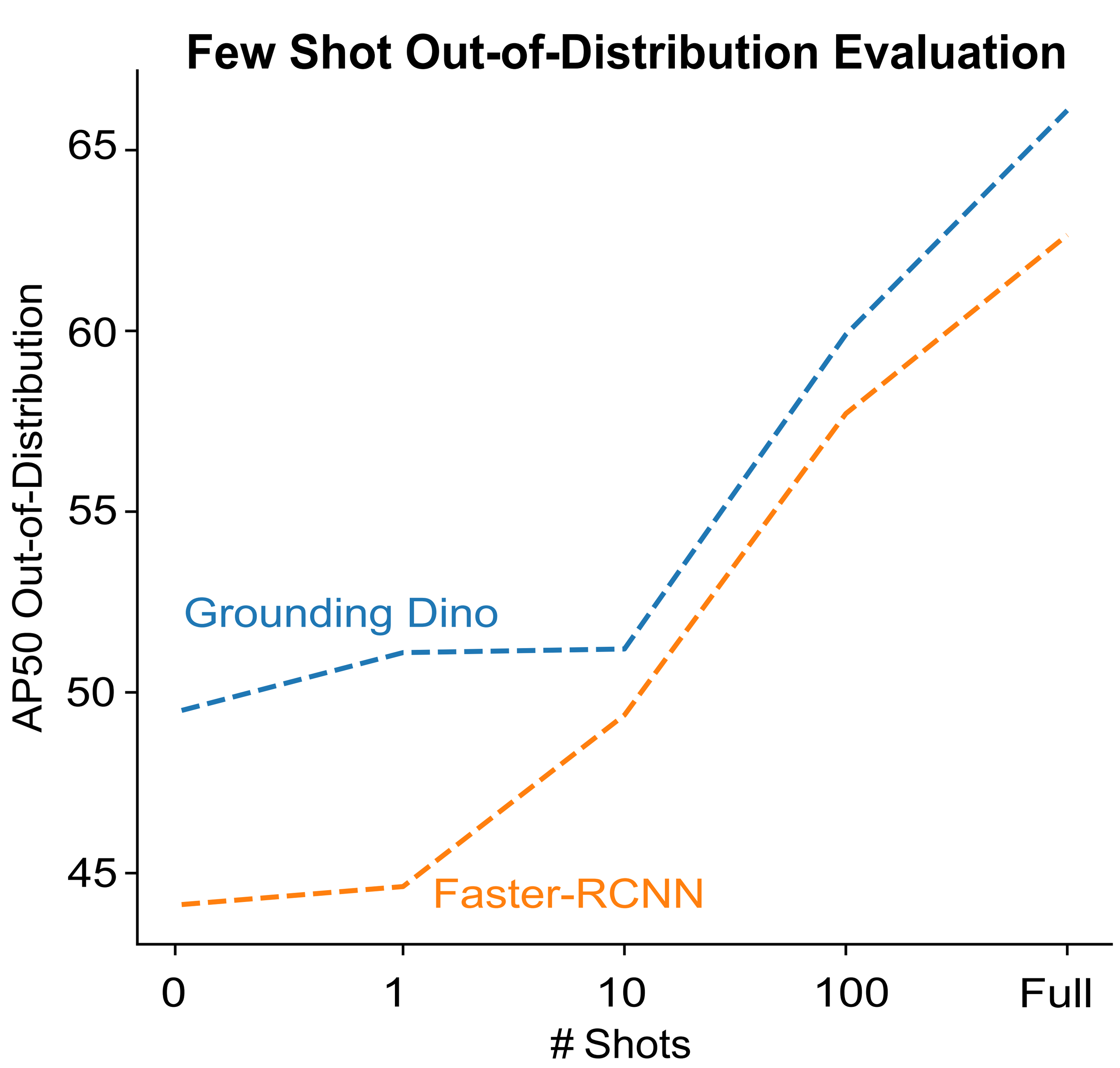

Figure 3. Few-shot OOD Evaluation of Grounding Dino and Faster-RCNN.

Table 1. Distribution Shift Evaluation AP Metrics

Method	Evaluation Type	ID AP	OOD AP
Faster-RCNN	Conventional Evaluation	77.8	63.1
Grounding DINO Full Finetune	**Conventional Evaluation**	**82.1**	**66.7**
Faster-RCNN	Distribution Shift Evaluation	77.8	44.1
SAM Finetune Full Finetune	Distribution Shift Evaluation	78.1	48.5
SAM Finetune Head	Distribution Shift Evaluation	77.7	48.3
SAM Finetune Head then Full	Distribution Shift Evaluation	59.2	41.7
Grounding DINO Full Finetune	Distribution Shift Evaluation	81.4	49.5
Grounding DINO Finetune Head	**Distribution Shift Evaluation**	**81.0**	**50.5**
Grounding DINO Finetune Head then Full	Distribution Shift Evaluation	80.8	48.7

Table 2. Per-zone evaluation of tree count R^2 for our model against best existing data product for tree cover in India

Zone	Grounding Dino R^2	Brandt Product R^2
Hyper-Arid Western Plains	0.862	0.77
Transitional Inland Drainage	0.97	0.806
Semi-Arid Eastern Plains	0.916	0.746
Sub-humid Southern Plains	0.919	0.347
Irrigated Northern	0.818	0.589
Transitional Plains of Luni Basin	0.759	0.619
Flood-prone Eastern Plains	0.771	0.576
Humid Southeastern Plains	0.764	0.326

Conclusion

- Strong performance of our deep learning model in tree detection, similar to recent work and better than best existing product in India.
- Significant drop in performance in OOD agro-ecological zones
- Foundation model based approaches including SAM & Grounding DINO show improvements in both ID & OOD performance, but also exhibit similar performance drops in OOD
- Large variation in accuracy as some areas more inherently difficult.
- With 10 ID examples, baseline performance OOD is similar to foundation model performance.