Predicting out-of-domain performance under geographic distribution shifts

Haoran Zhang*, Konstantin Klemmer†, Esther Rolf§, David Alvarez-Melis*

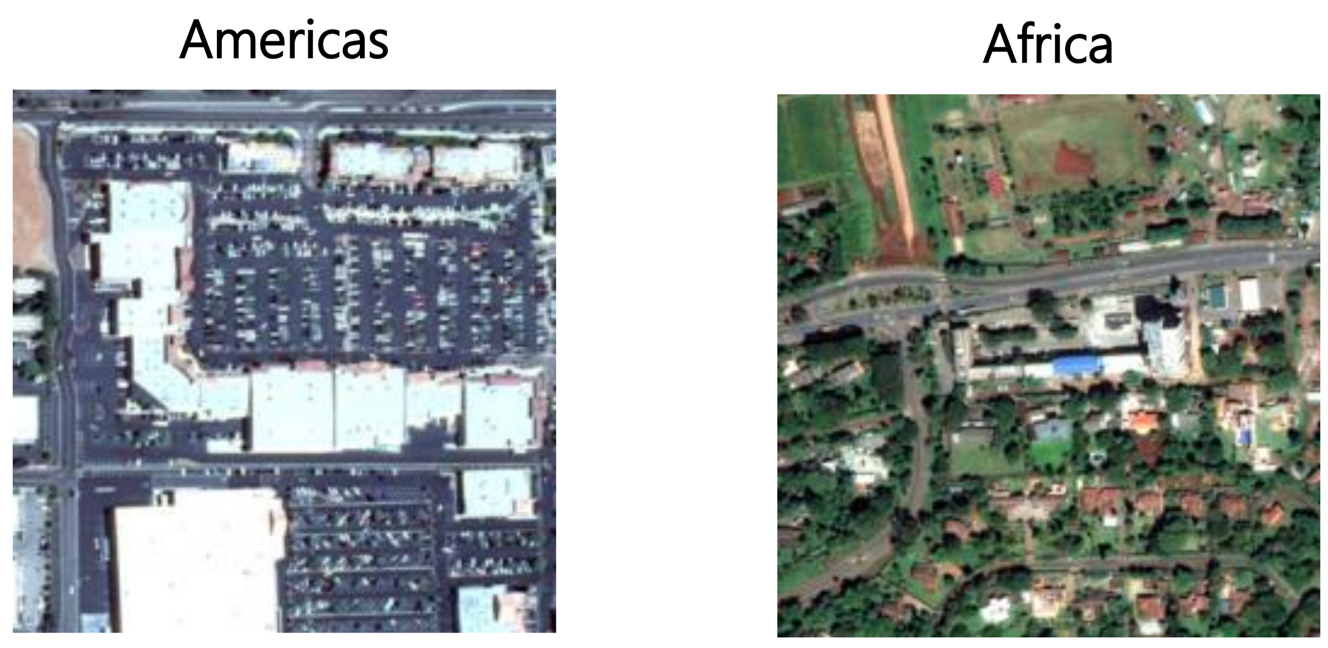
Harvard John A. Paulson School of Engineering and Applied Sciences * | Microsoft † | University of Colorado Boulder §

Motivation:

- Due to gaps in geographical **data availabilities**, **domain adaptation** is commonly used to transfer predictive capabilities trained in data-rich regions to data-poor regions
- In **satellite imaging tasks**, **distribution shifts** across different geographical units are significant challenges:

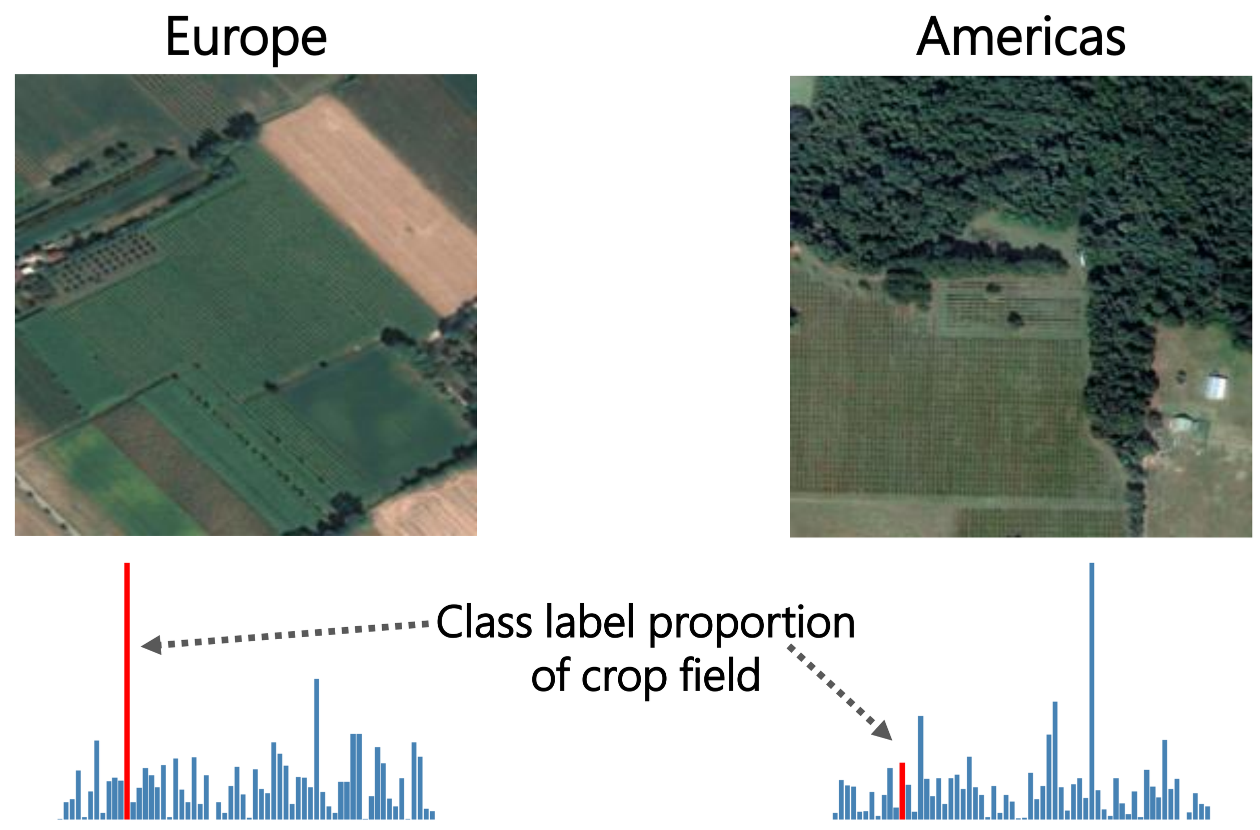
Covariate Shift: Labels stay consistent across domains, but input distribution (e.g., visual patterns) changes

Shopping Mall in ...



Label Shift: Visual appearance of a class remains similar, but class proportions in different domains differ

Crop Field in ...



- These shifts make it less ideal to blindly transfer models to out-of-distribution geographical regions for optimal performance.
- Intuitively, models are more likely to transfer adequately between distributionally similar domains.

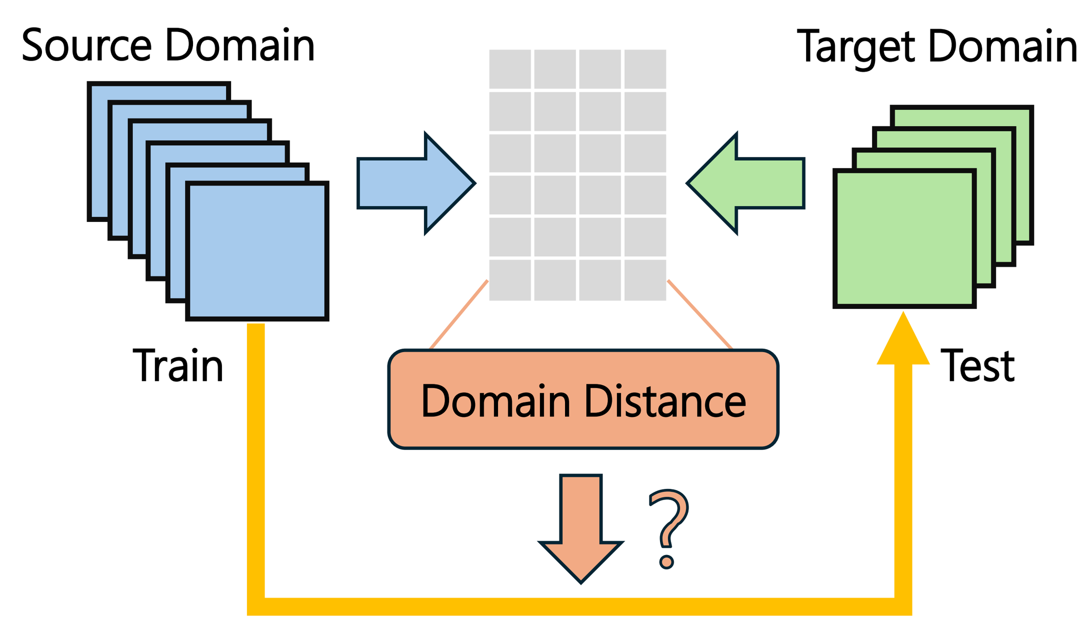
Can we use certain **distance measures** between **domain distributions** to **predict domain adaptation performance** in satellite imaging tasks?

Method:

- We seek a notion of **distance measure** that can serve as reliable **predictors** of **domain adaptation performance**.

Distance Computation:

- **Compute pairwise distance** across domain pairs and **aggregate** through averaging/cost objective optimization.
- **Average Cosine Distance**: angular distance between embedding vectors
- **Wasserstein Distance**: optimal transport based distance between embedding vectors
- **Average Arc Distance**: geographical distance between image locations



Domain Adaptation:

- **Fit** pretrained image model on a **source domain**
- **Evaluate** its **test performance** on every domain (including source domain itself)
- **Finetune** trained model on each other domain as **target domain** in **fewshot** settings, and evaluate test performance in target domain

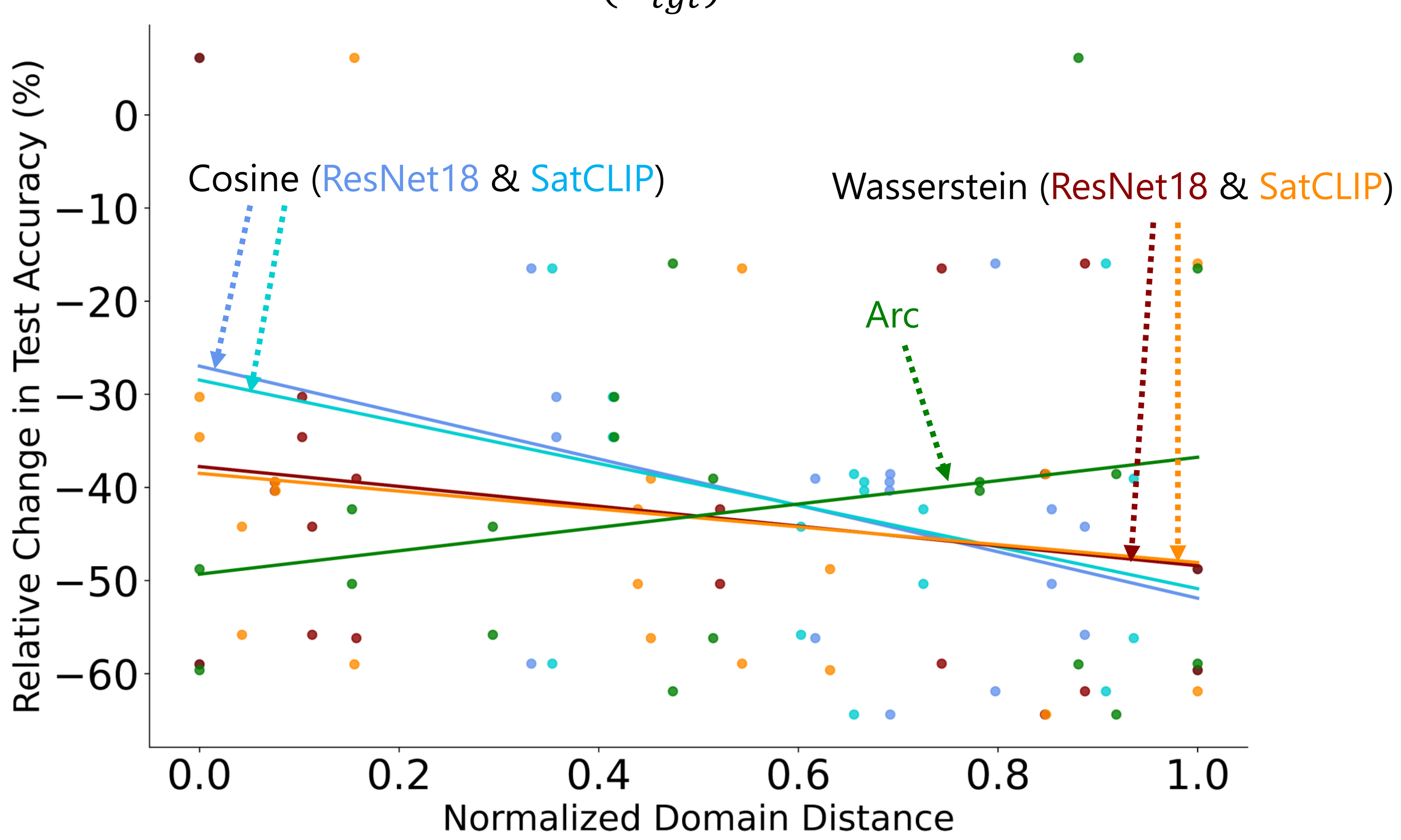
Experiments & Results:

- We conduct experiments and analysis on a variety of models:
  - For domain distance computation, we include image-based models (ResNet, ViT) and location-based models (SatCLIP)
  - For domain adaptation, we train and evaluate ResNet and DenseNet under zeroshot and fewshot transfer settings

Moderate correlation between cosine and Wasserstein distance and effectiveness of domain transfer

To make accuracy values **comparable** across different domain pairs and models, we compute **relative change** in domain adaptation performance

$$\Delta = \frac{\text{acc}(\mathcal{D}_{src} \Rightarrow \mathcal{D}_{tgt}) - \text{acc}(\mathcal{D}_{tgt})}{\text{acc}(\mathcal{D}_{tgt})}$$



Normalized Wasserstein distances of embeddings between continent domains in FMoW-wilds (left: ResNet18, right: SatCLIP)

Asia		0.45	0.15	0.02	1.00
Europe	0.45		0.95	0.00	0.84
Africa	0.15	0.95		0.15	0.97
Americas	0.02	0.00	0.15		0.14
Oceania	1.00	0.84	0.97	0.14	

Asia		0.44	0.05	0.08	1.00
Europe	0.44		0.63	0.00	0.55
Africa	0.05	0.63		0.45	0.85
Americas	0.08	0.00	0.45		0.16
Oceania	1.00	0.55	0.85	0.16	

Domain adaptation performance with DenseNet121 on FMoW-wilds (left: 0-shot, right: 10-shot)

Asia	0.604	0.311 (-48.6%)	0.277 (-49.4%)	0.337 (-44.8%)	0.365 (-34.3%)
Europe	0.269 (-55.6%)	0.605	0.195 (-64.4%)	0.399 (-34.7%)	0.458 (-17.6%)
Africa	0.249 (-58.8%)	0.211 (-65.1%)	0.548	0.251 (-58.9%)	0.332 (-40.3%)
Americas	0.297 (-50.9%)	0.393 (-35.1%)	0.268 (-51.1%)	0.611	0.575 (+3.5%)
Oceania	0.141 (-76.7%)	0.169 (-72.1%)	0.115 (-79.0%)	0.226 (-63.0%)	0.556

Asia	0.604	0.349 (-42.3%)	0.306 (-44.2%)	0.370 (-39.4%)	0.467 (-15.9%)
Europe	0.300 (-50.4%)	0.605	0.281 (-48.8%)	0.399 (-34.6%)	0.464 (-16.5%)
Africa	0.267 (-55.8%)	0.244 (-59.6%)	0.548	0.268 (-56.2%)	0.341 (-38.6%)
Americas	0.360 (-40.4%)	0.422 (-30.3%)	0.334 (-39.1%)	0.611	0.590 (+6.1%)
Oceania	0.230 (-61.9%)	0.249 (-58.9%)	0.195 (-64.4%)	0.250 (-59.0%)	0.556