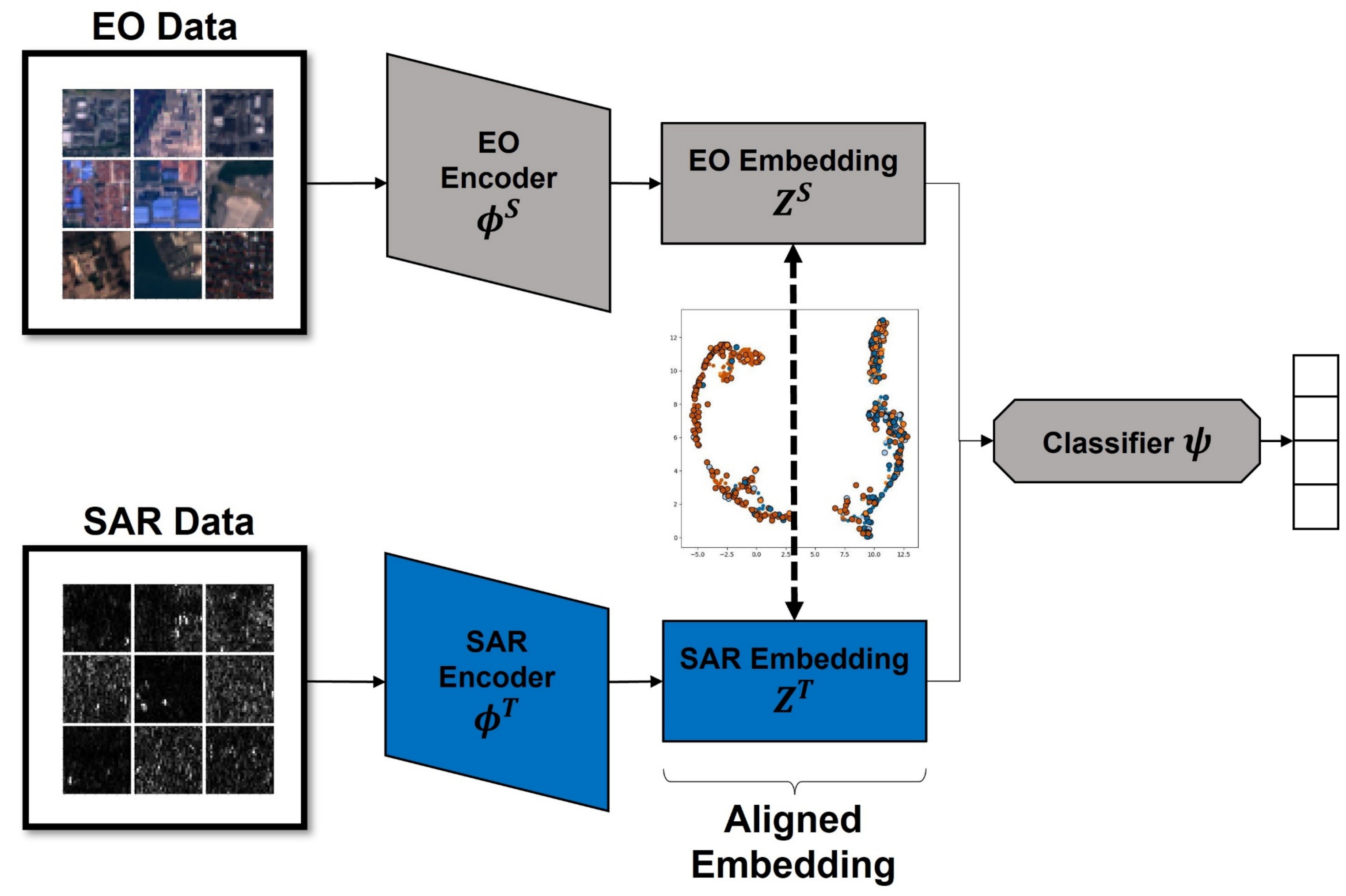
Learning Source Domain Representations for Electro-Optical to SAR Transfer

Boya Zeng, Marcel Hussing, Eric Eaton
{boyazeng, mhussing, eeaton}@seas.upenn.edu
Penn UNIVERSITY of PENNSYLVANIA

Background and Motivation

Scarcity of labeled synthetic aperture radar (SAR) data and label-abundance of electro-optical (EO) data motivate EO to SAR transfer via distribution alignment



Task:
Land use prediction for 4 low-rise classes in the So2Sat dataset
EO to SAR Transfer Framework:

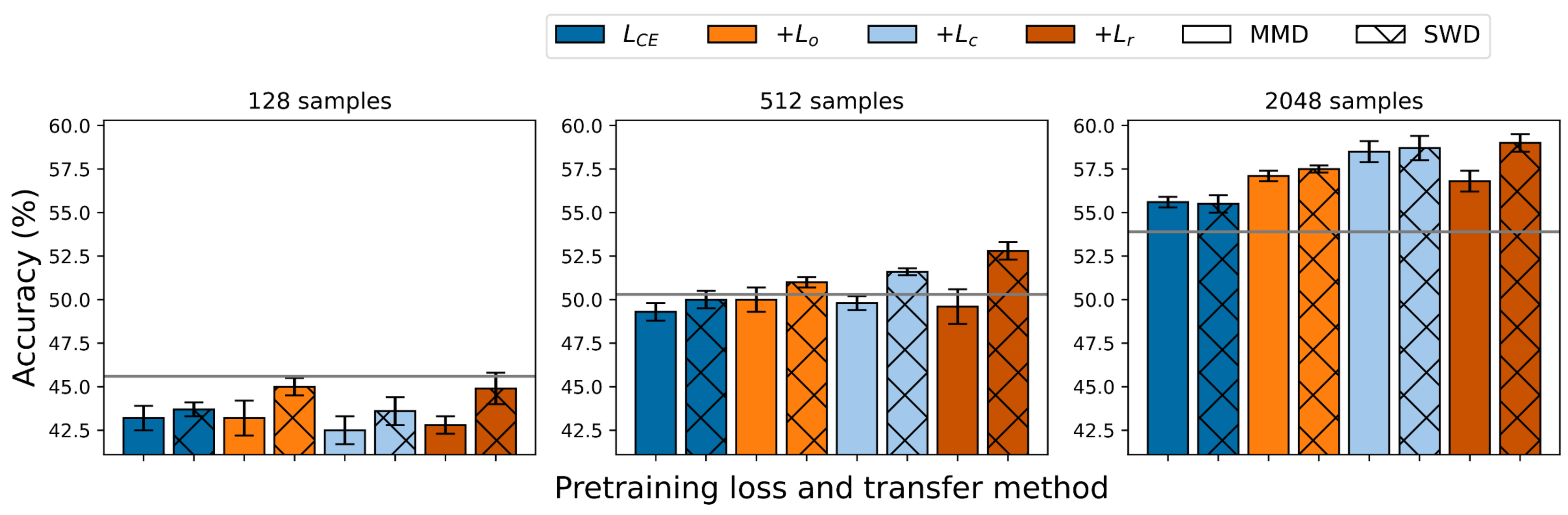
- Y-shaped network: EO and SAR encoders and a shared classifier
- Pretraining EO encoder and shared classifier on EO data
- Freeze shared classifier
- Update SAR encoder to align embedding distributions of EO and SAR data based on metric Sliced Wasserstein Distance (SWD) or Maximum Mean Discrepancy (MMD)

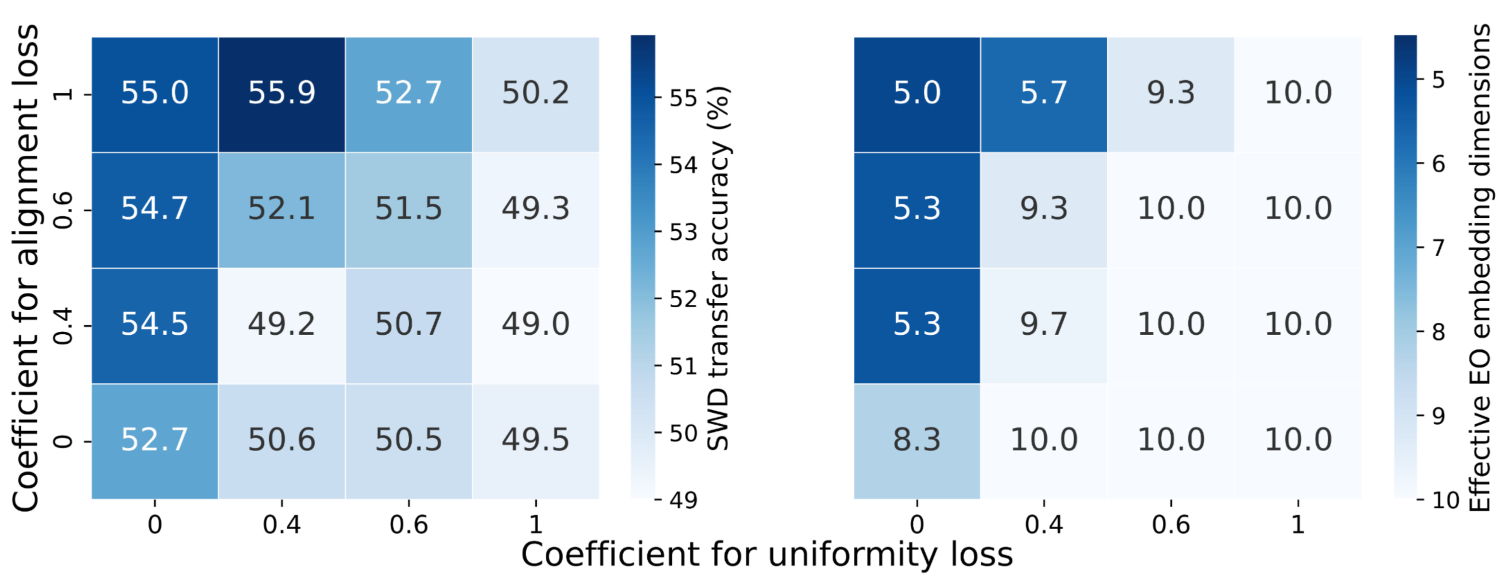
Question: Why does Supervised Contrastive EO pretraining lead to better SAR transfer performance?

Low-rank Source Embeddings Improve Transfer

Dissecting Supervised Contrastive Loss

- Decreasing *uniformity* and increasing *alignment* increases downstream performance
- Increased accuracy is correlated with lower *effective dimension* of the embedding

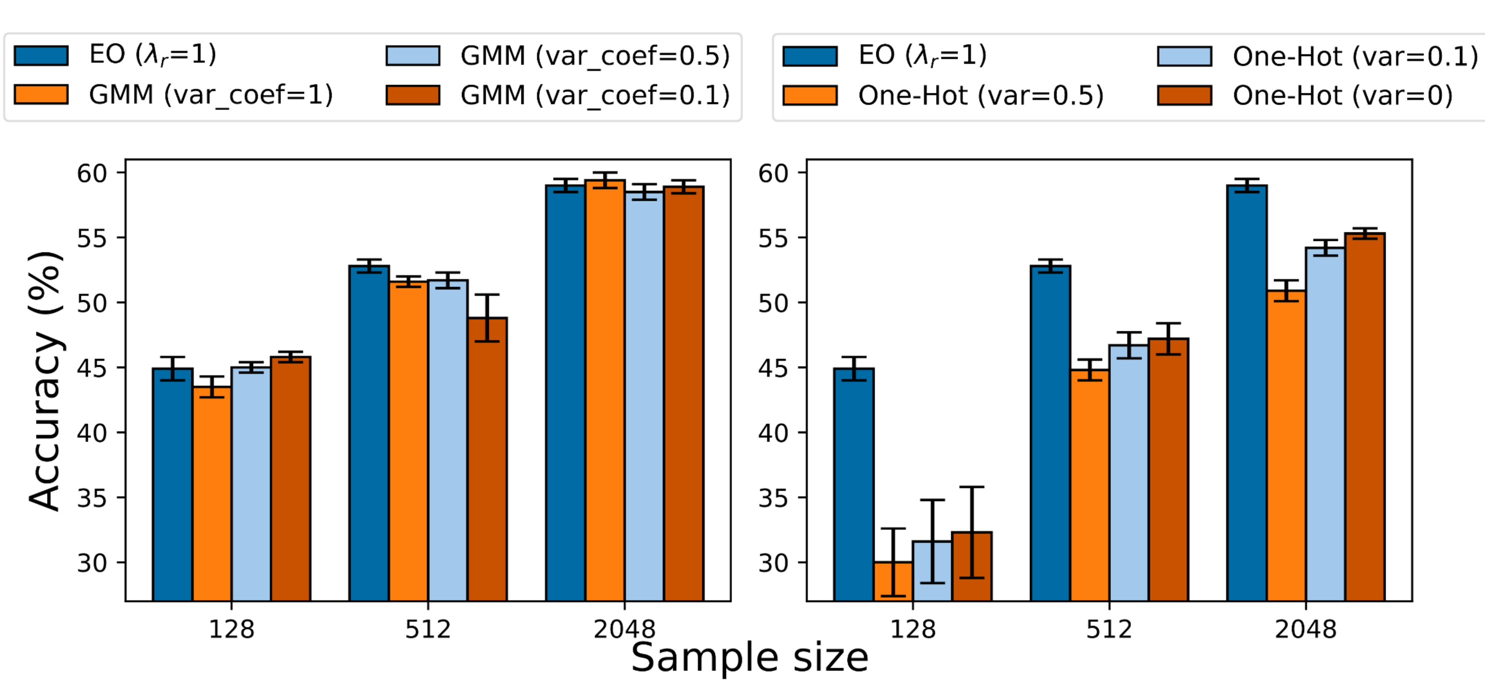




Effective Dimension Reduction as an Optimizable Learning Objective

- Rank Reduction Loss ($+L_r$) that directly regularizes the effective rank of the embeddings outperforms cross-entropy-only baseline (L_{CE}), Supervised Contrastive Loss ($+L_c$), and OLÉ loss ($+L_o$)
- Down-stream accuracy is correlated with low *effective dimension* and *inertia*

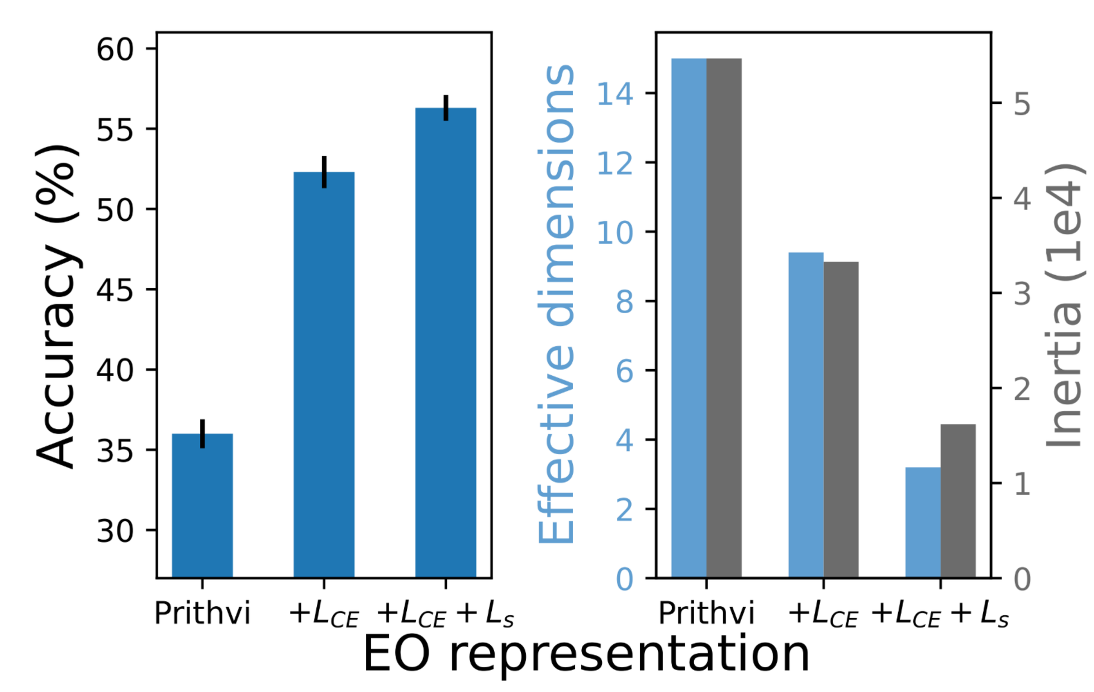
Can We Condense Clusters with Even Simpler Distributions?



- Represent the EO class embedding clusters with Gaussian Mixture Models or One-Hot Vectors and scale the variance
- Such simpler EO embedding distributions with lower *inertia* don't lead to better transfer performance

Upshot: fine-grained information about class distribution and inter-class relations is critical for the source embedding

Can We Align Distribution to a Foundation Model (Prithvi)?



- High *effective dimension* and *inertia* make Prithvi's EO embeddings poor task-specific alignment targets
- Condensing embedding dimension via contrastive learning ($+L_s$) only recovers transfer performance comparable to that of standard CNN

Model (embedding dim.)	MAE (32)	MAE (64)	MAE (128)	MAE (256)
SWD accuracy	52.2$_{\pm 1.1}$	52.9$_{\pm 1.1}$	51.8$_{\pm 0.7}$	51.8$_{\pm 0.5}$
Effective dim.	11.6$_{\pm 1.0}$	18.0$_{\pm 2.0}$	26.8$_{\pm 1.4}$	32.8$_{\pm 2.1}$
Inertia	28108.4$_{\pm 2546.0}$	51677.6$_{\pm 3886.2}$	122952.8$_{\pm 6874.4}$	278275.6$_{\pm 36346.7}$

Transfer performance from Masked Autoencoders trained on So2Sat EO data highlights the importance of EO training data distribution