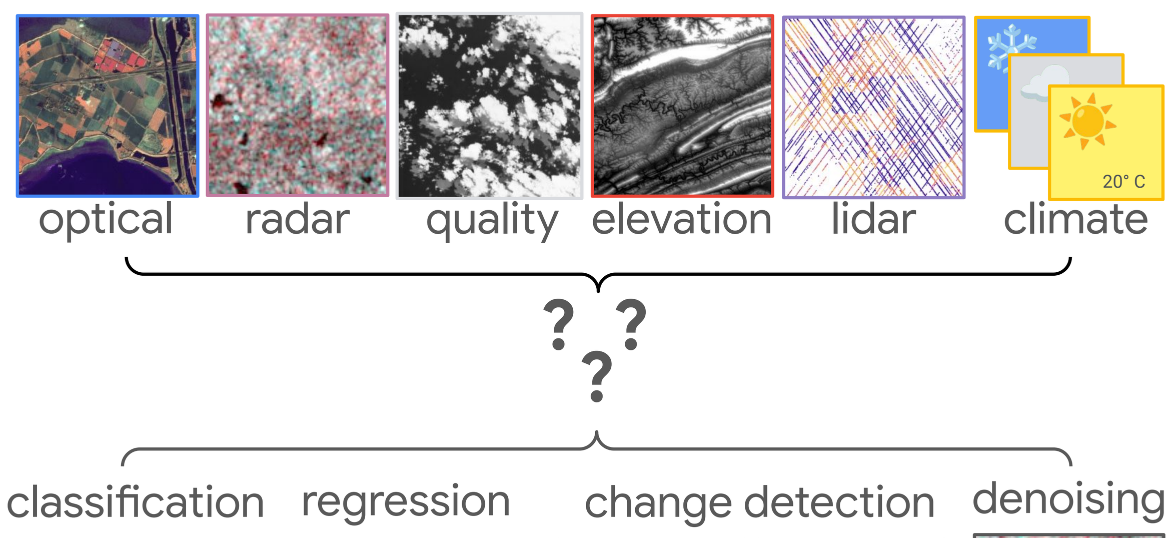
Learned Embedding Fields for Multi-Source/Multi-Temporal Earth Observation Imagery

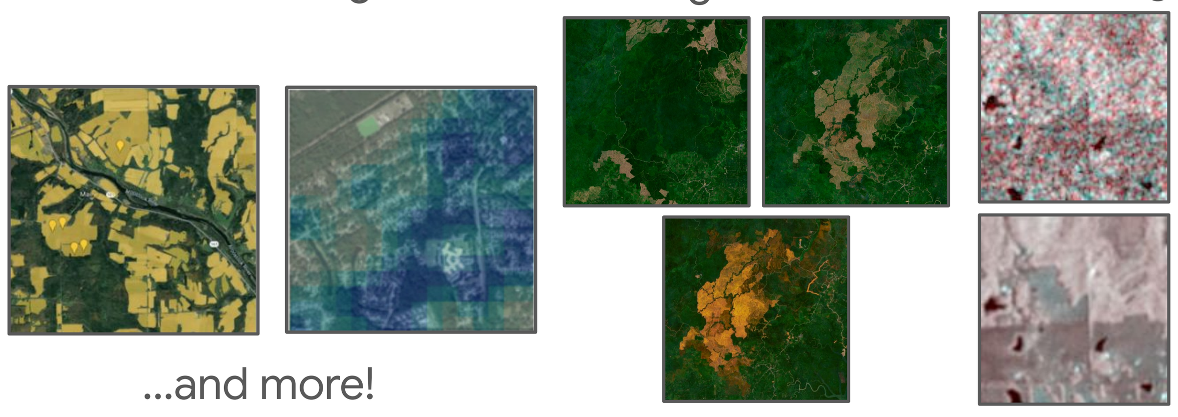
Chris Brown[GDM*], Michal Kazmierski[GDM*], William Rucklidge[G], Valerie Pasquarella[G],
Sophia Alj[GDM], Emily Schechter[G], Sean Askay[G], Alexis Boukouvalas[GDM], Evan Shelhamer[GDM]

[GDM] Google DeepMind [G] Google * Equal Contribution

problem

Many Sources, Many Tasks





...and more!

solution

One Learned Embedding Served as Data for All Tasks

sources → EF → tasks

interface: index by space & time
`EF[lat, lon, year]`

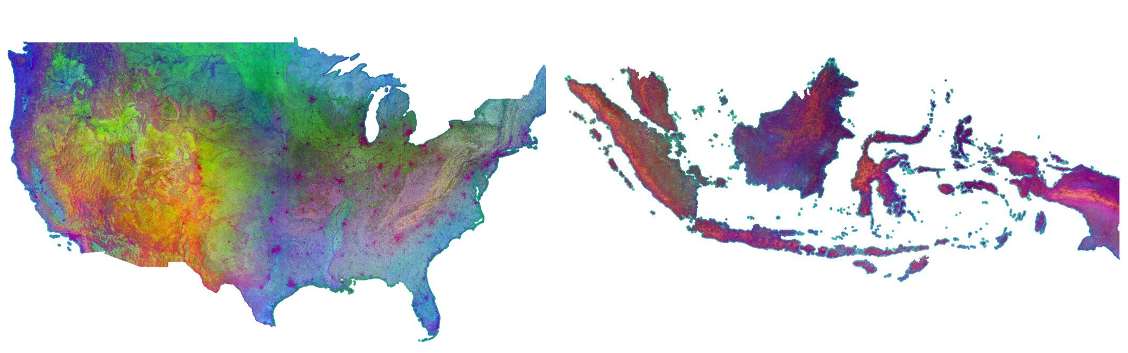
less direct reliance on sources
= less expertise needed

no deep network re-training
= no waiting (interactive even!)

more learning in the embedding
= more maps in fewer labels

scope

Geography, Time, Outputs



Geography conterminous US
+ Indonesia + Malaysia
Time years 2018–2022
Dimension 64 channels / bands
Resolution $10m^2$ annualized

data

Benchmarking Tasks + Labels

Land Use LCMAP CONUS 2018
Forest, Devel., Range., Agri., Wet., Other
250-shot episodes x 100 on ~8k test set

Trees iNaturalist CONUS 2020-2022
28 genera in union of top-5 for each state
150-shot episodes x 10 on ~11k test set

Eco. Regions RESOLVE IDN-MYS '17
38 zones like New Guinea mangrove
100-shot episodes x 125 on ~14k test set

method

Learning: Self-Supervised Auto-Encoding + Trans-Coding

input sources -> shared encoder -> latent -> decoders -> input & target sources

encoder
- hybrid convolution/attention arch
- hierarchical over space and time
- preserves spatial dims like HRNet

latent = EF

decoders
- MLP taking latent + datetime
- per-source parameters
- explicit space / implicit time

■ = input source for auto-encoding ☐ = target source for trans-coding

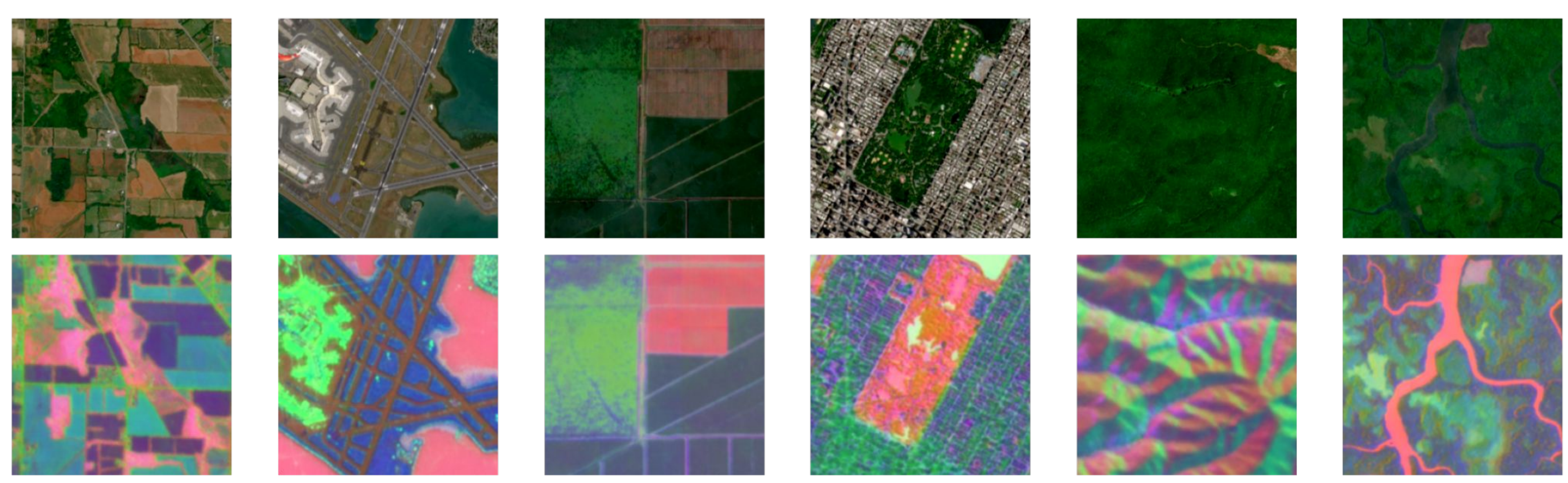
method

Inference: Summarizing Sources into Embedding Fields

1 EF idx = **space** ~1 km^2 area of 128x128 pixels at 10m resolution
time 1 year of 100+ observations across sources
sources Sentinel-2, Sentinel-1, Landsat-8/9

result

Visualization by PCA Reveals Sensitivity to Input Details



result

Benchmarking Few-Shot Accuracy by kNN + Linear Models

Method	Dim.	Sources	Land Use	Trees	Eco. Regions
Data	5	S2	51.8 / 66.9	8.7 / 10.8	10.1 / 14.0
Data	17	S2,S1,LS	60.0 / 1.8	9.8 / 3.6	14.3 / 2.6
MOSAIKS	8192	S2(RGB)	49.7 / 22.8	7.0 / 3.6	8.4 / 4.0
MOSAIKS	8192	S2	60.0 / 55.2	9.6 / 3.8	11.9 / 3.5
MOSAIKS	64	S2	59.7 / 65.8	9.3 / 13.9	11.9 / 16.5
MOSAIKS	66	S2,S1,LS	61.8 / 1.8	9.7 / 3.6	14.2 / 2.6
MOSAIKS	192	S2,S1,LS	63.1 / 1.8	10.0 / 3.6	14.9 / 2.6
EF (Random)	64	S2,S1,LS	39.2 / 59.1	7.5 / 11.2	24.0 / 38.8
EF (Ours)	64	S2,S1,LS	**76.9 / 82.6**	**18.3 / 22.9**	**44.8 / 50.7**

Tasks Multi-class annualized recognition in US, Indonesia, and Malaysia
Metric: Accuracy % **Classifiers**: Nearest Neighbor (k=3) **/** Linear Model
Summary: more sources = better, learning helps (EF Random vs. Ours), time is is important

data

Data Sources: Inputs + Targets

inputs for inference and learning:

Sentinel-2 L1C: RGB, NIR 10m SWIR1 20m & Cloud Score+
Sentinel-1 L1GRD: VV, VH, HH, HV, Angle 10m
Landsat-8/9 T1: RGB, {N,SW,T}IR1 30m Pan 15m & FMASK

targets for learning:

GLO-30 elevation
GEDI lidar
ERA5-Land climate

References

VAE Kingma & Welling. ICLR'13
Land Use / LCMAP Pengra et al. usgs.gov'21
MOSAIKS Rolf et al. Nature Comms'21
Earth Obs.: LandSat Wulder et al. RS of Env.'22
HRNet Wang et al. PAMI'20
Trees / iNaturalist iNaturalist.org & gbif.org'23
Analysis-ready Data Dwyer et al. Remote Sensing'18
Earth Obs.: Sentinel Berger et al. RS of Env.'12
Cloud Score+ Pasquarella et al. CVPRW'23
Eco Regions / RESOLVE Dinerstein et al. BioScience'17