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

Many Sources, Many Tasks

- optical
- radar
- quality
- elevation
- climate

classification regression change detection denoising

...and more!

One Learned Embedding Served as Data for All Tasks

solution

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

Geometry, Time, Outputs

Geography: conterminous US + Indonesia + Malaysia

Time: years 2018–2022

Dimension: 64 channels / bands

Resolution: 10m² annualized

Benchmarked 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

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

Inference: Summarizing Sources into Embedding Fields

space ~1 km² area of 128x128 pixels at 10m resolution

1 EF idx = time 1 year of 100+ observations across sources

sources: Sentinel-2, Sentinel-1, Landsat-8/9

Visualization by PCA Reveals Sensitivity to Input Details

Benchmarking Few-Shot Accuracy by kNN + Linear Models

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

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 important

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)R1 30m Pan 15m & FMASK

targets for learning:

- GLO-30 elevation
- GEDI lidar
- ERA5-Land climate

References

- MOSAIKS: Rolf et al. Nature Comms'21
- Earth Obs.: LandSat: Wulder et al. RS of Env.'22
- HHNet: Wang et al. PAMI'20
- MOSAIKS: Pasquarella et al. CVPRW'23
- Eco Regions: RESOLVE
- Dinerstein et al. BioScience '17

GDM*: Equal Contribution