Weakly Supervised Learning for Land Cover Classification from Earth Observation

Conrad Albrecht
Earth Observation Center @ DLR
joint work with the "EvoLand" Horizon Europe Consortium & HelmholtzAI

Jun 12, 2024 @ European Commission's Joint Research Center

DLR

Evoland
LAND MONITORING EVOLUTION

HELMHOLTZAI

Agenda

HELMHOLTZ AI

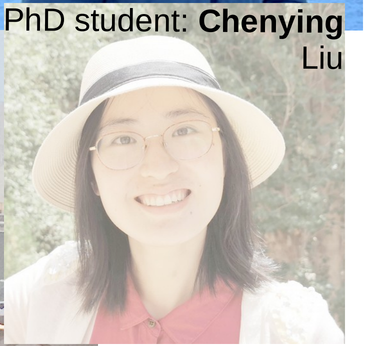
FYI: *hover & click paper references for DOI link*

I Introduction to EvoLand

HelmholtzAI Young Investigator Group
DM4EO@DLR


PI: **Conrad** Albrecht

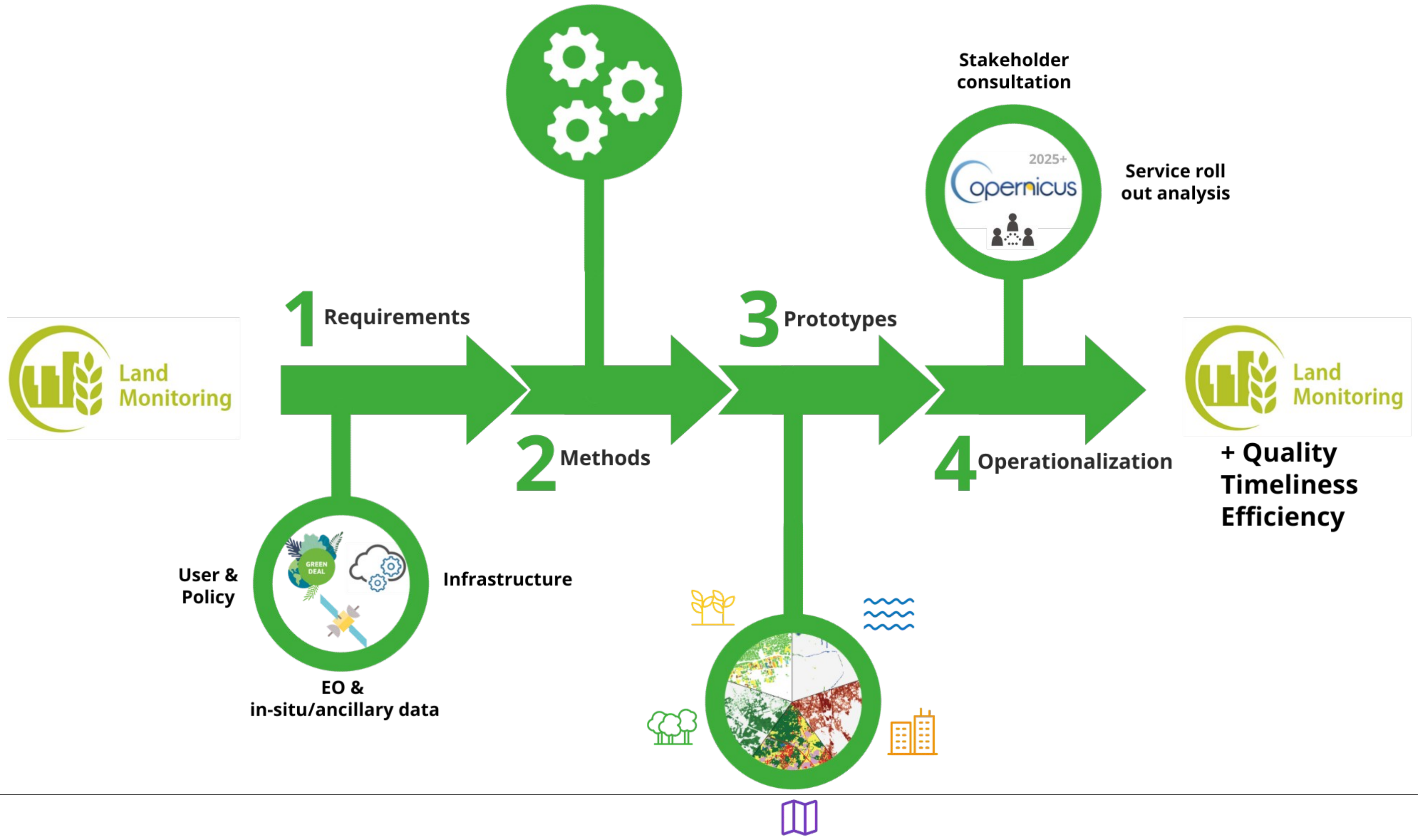

PhD student: **Nassim** Ait Ali Braham


PhD student: **Chenying** Liu


PhD student: **Yi** Wang

Coordinated by vito Remote Sensing

I Introduction to EvoLand



I Introduction to Weakly/Self-Supervised – taxonomy & history

our focus

current research

sources of training data for deep learning models

supervised

domain experts manually annotate data
– labor intensive

semi-supervised

exploit knowlege on structure of data
– models highly depending on data

weakly supervised

efficiently gather low-quality labels
+ exploit automated annotation

- *inexact* – coarse-grained labels
- *inaccurate* – noisy labels
- *incomplete* – missing labels

self-supervised

unsupervised generation of task-agnostic data representations
+ task training on little manual labels

- *generative* – data generation
- *predictive* – data puzzles
- *contrastive* – data correlation

Vision Transformers

Foundation Models

deep learning methodology timeline

>2000 ~2009 ~2017 ~2019 ~2023

I Introduction: Self-Supervised Learning (SSL) in a nutshell

Stage I: *self-supervised pre-training without labels*

methodologies:

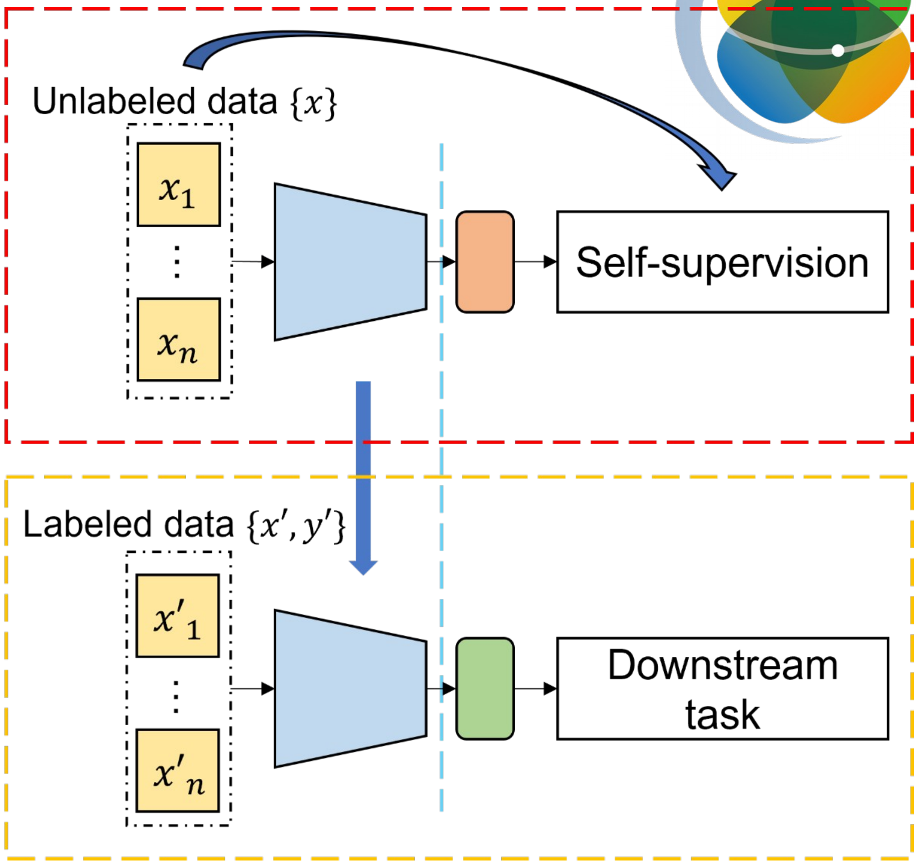
- *Generative* – data (re)construction
- *Predictive* – solve puzzles on data
- Contrastive – exploite data relationships

→ data compression/efficient feature representation

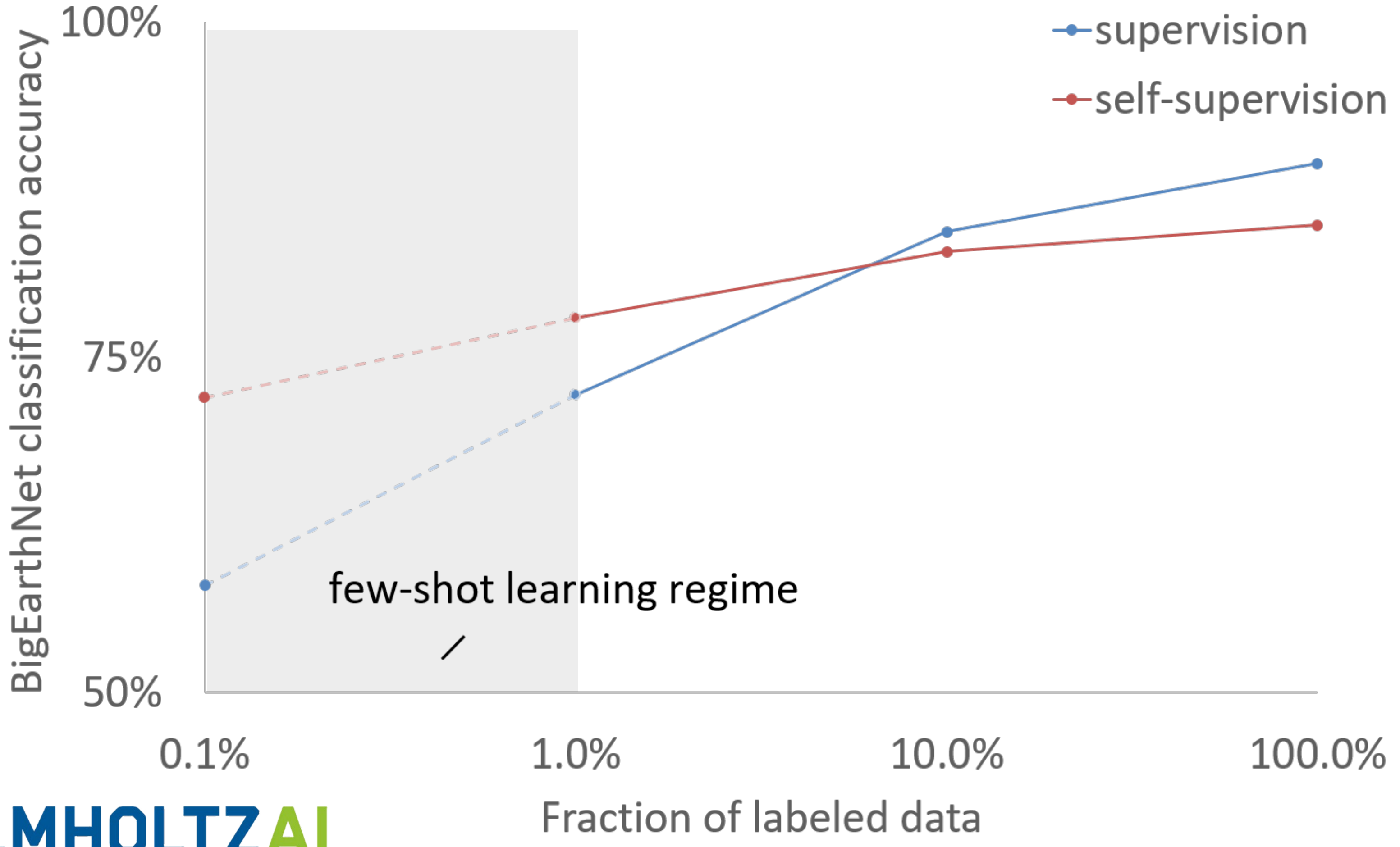
Stage II: *supervised downstream task training with labels*

potential downstream tasks:

- data classification
- object detection
- semantic image segmentation
- data retrieval from feature representations
- etc.



I Introduction: SSL4EO-based deep learning model benefits



II SSL4EO-S12: a Sentinel-1/2 EO benchmark dataset for SSL

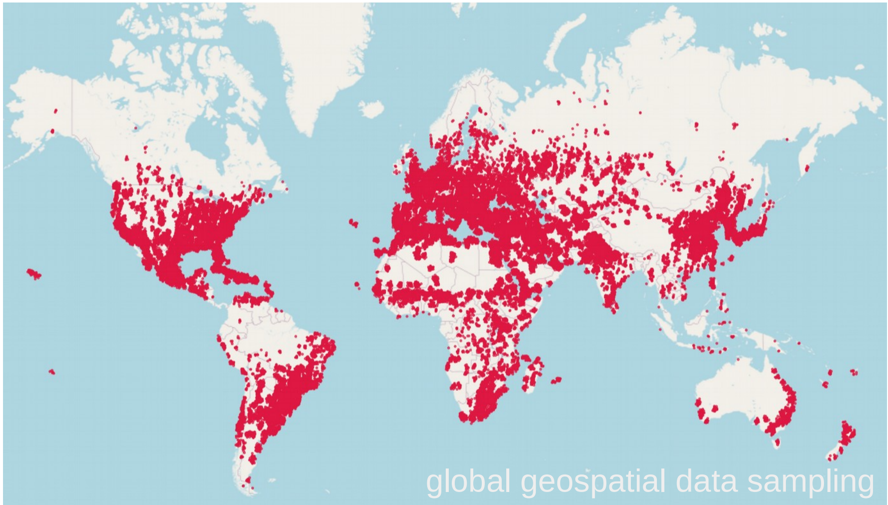
Large scale (~2TB of data)
250k of non-overlapping, 264x264 pixels images

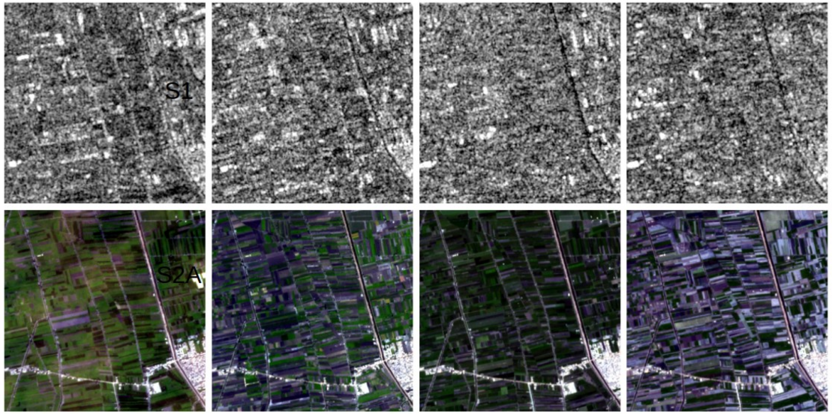
Global scale
top 10k populated cities (# inhabitants >65k)

Diverse geolocations
uniform sampling of cities
Gaussian sampling from city center

Multi-modal
sourced from Sentinel-1 and -2 data products

Multi-temporal
4 season sampling per image patch for year 2021





Sentinel-1 (SAR)
@ 10 meters
→ VH polarization channel
→ VV polarization channel

Sentinel-2 (optical)
@ 10, 20 & 60 meters
→ 13 spectral bands

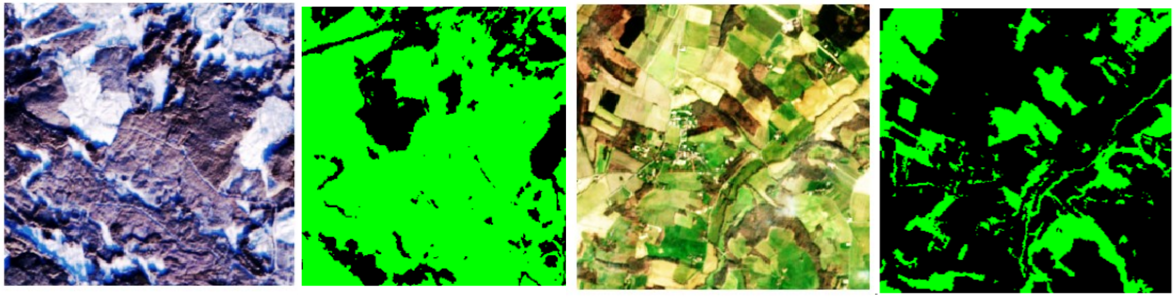
II SSL4EO-EU-Forest: model fine-tuning for European forests

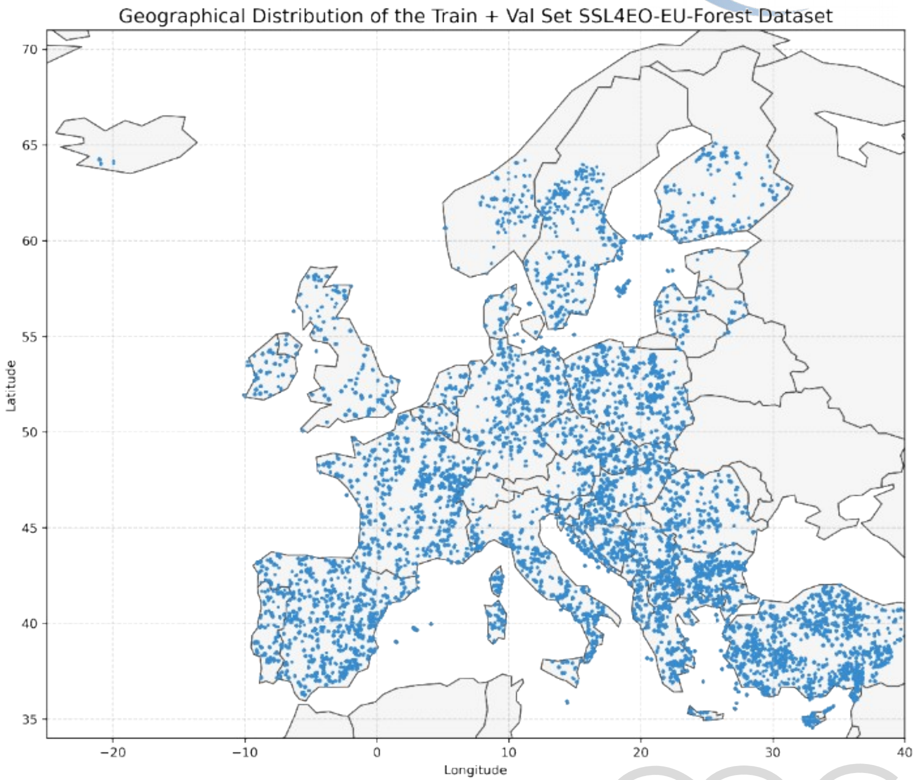
Data Scale (~0.1TB)
16k of non-overlapping, 264x264 pixels patches
of Sentinel-2 images

Forest Centered
sampled over European forests
(reference: 2018 HRL, domain expert picked by GAF AG)

Multi-temporal
4 seasons sampled per image patch for year 2018

Labelled
pixels co-registered with 2018 HRL forest product
(binary mask: tree vs. no-tree)

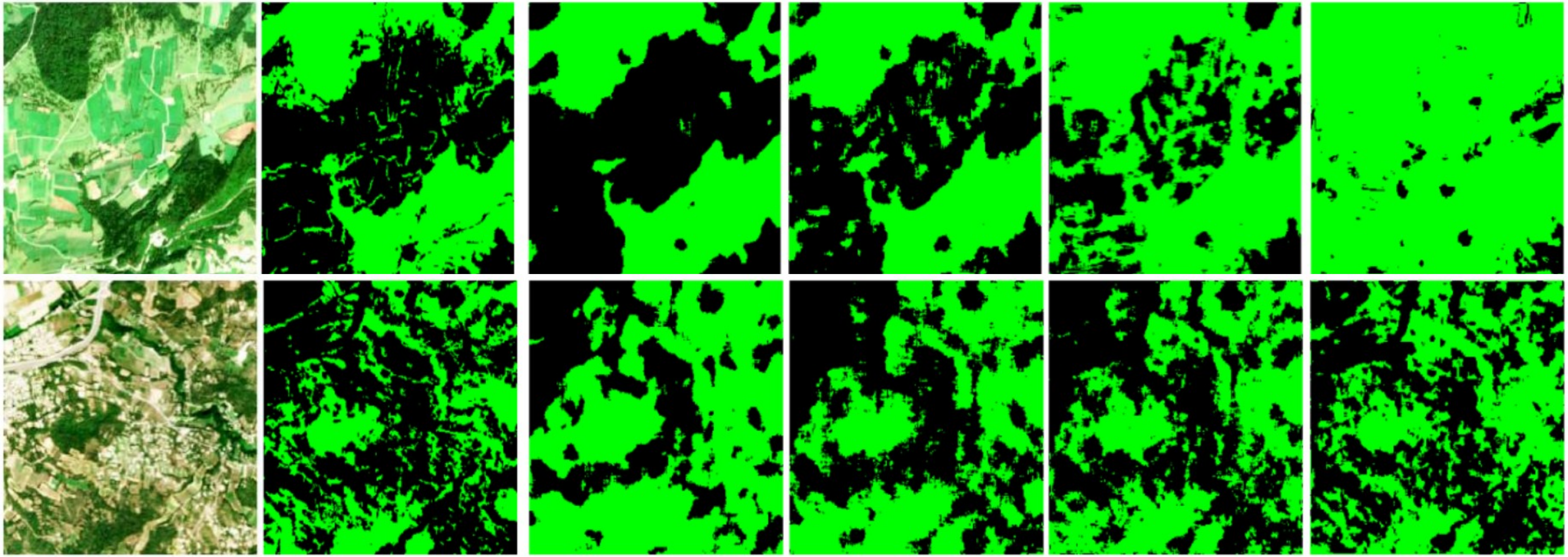




II SSL4EO-EU-Forest: Forest Semantic Segmentation (single timestamp)

ABLATION STUDY

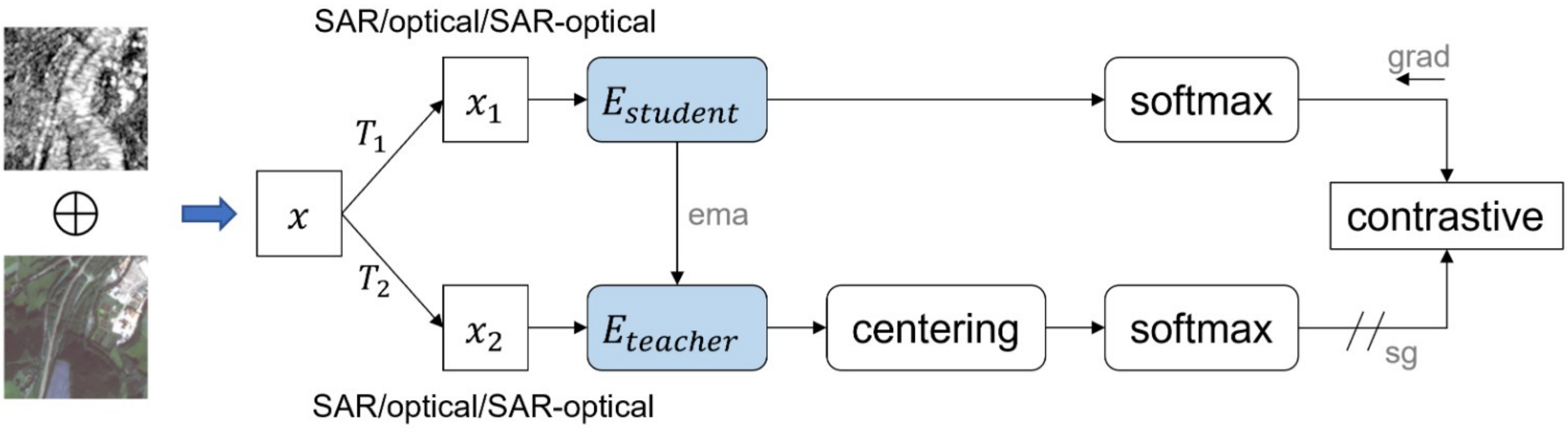
From left to right: Sentinel-2, HRL forest mask (ground truth), FCN-ResNet-18 prediction All layers, 3, 2, and 1 layer(s).



II SSL4EO-EU-Forest: Forest Semantic Segmentation (single timestamp)

Segmentation Protocol	Encoder	Pre-training Weights	Overall Accuracy	Mean IoU
UNet	ResNet-18	Random	84.05	72.49
		MoCo	**85.19**	**74.19**
		DINO	85.14	74.13
	ResNet-50	Random	84.25	72.79
		MoCo	**85.27**	**74.33**
		DINO	84.84	73.68
DeepLabV3+	ResNet-18	Random	84.15	72.63
		MoCo	84.94	73.83
		DINO	**85.15**	**74.14**
	ResNet-50	Random	84.17	72.67
		MoCo	**85.31**	**74.39**
		DINO	84.76	73.55
UpConv	ViT-S	Random	83.47	71.63
		MoCo	85.08	74.03
		DINO	**85.26**	**74.31**

II SSL4EO: DINO-MM – model boost performance by SAR-optical data fusion

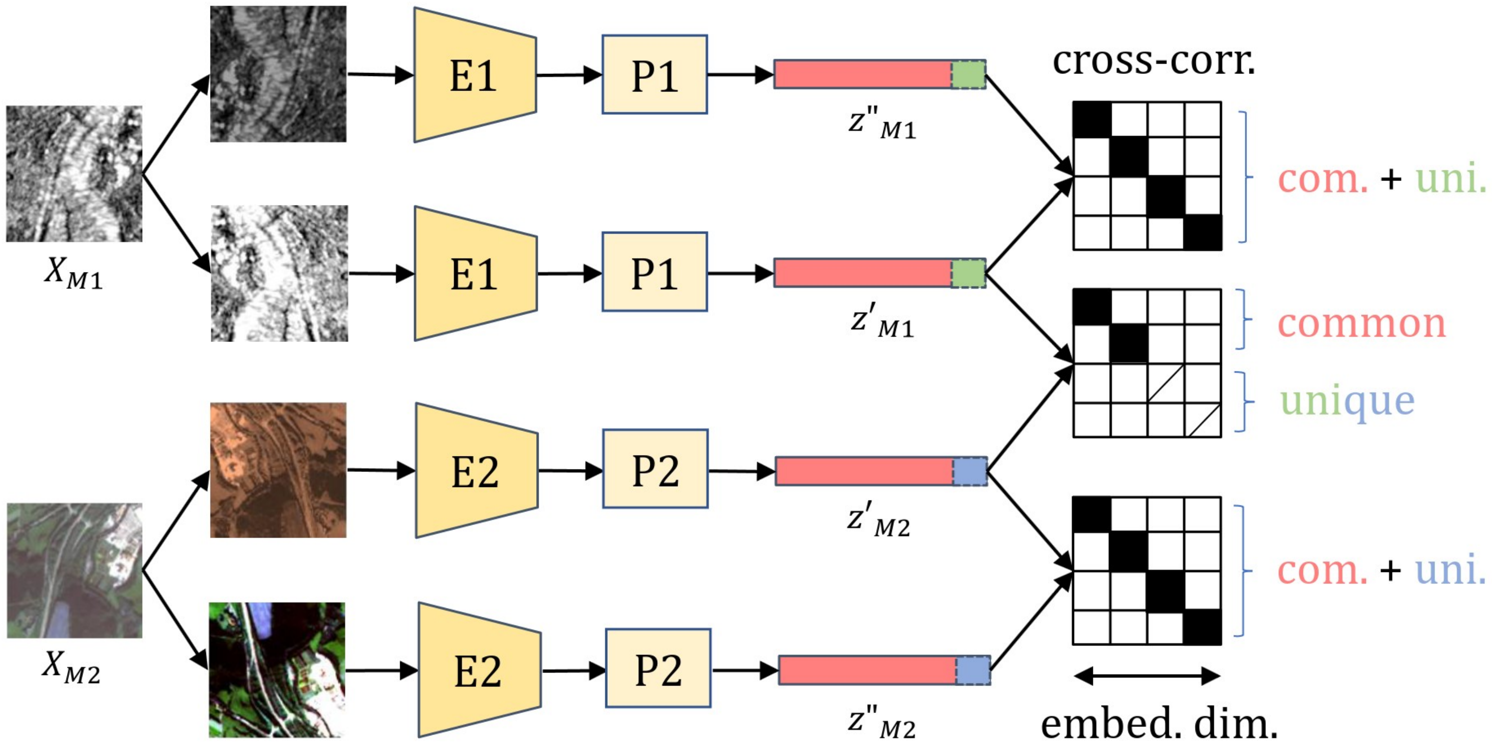


fraction of labels available

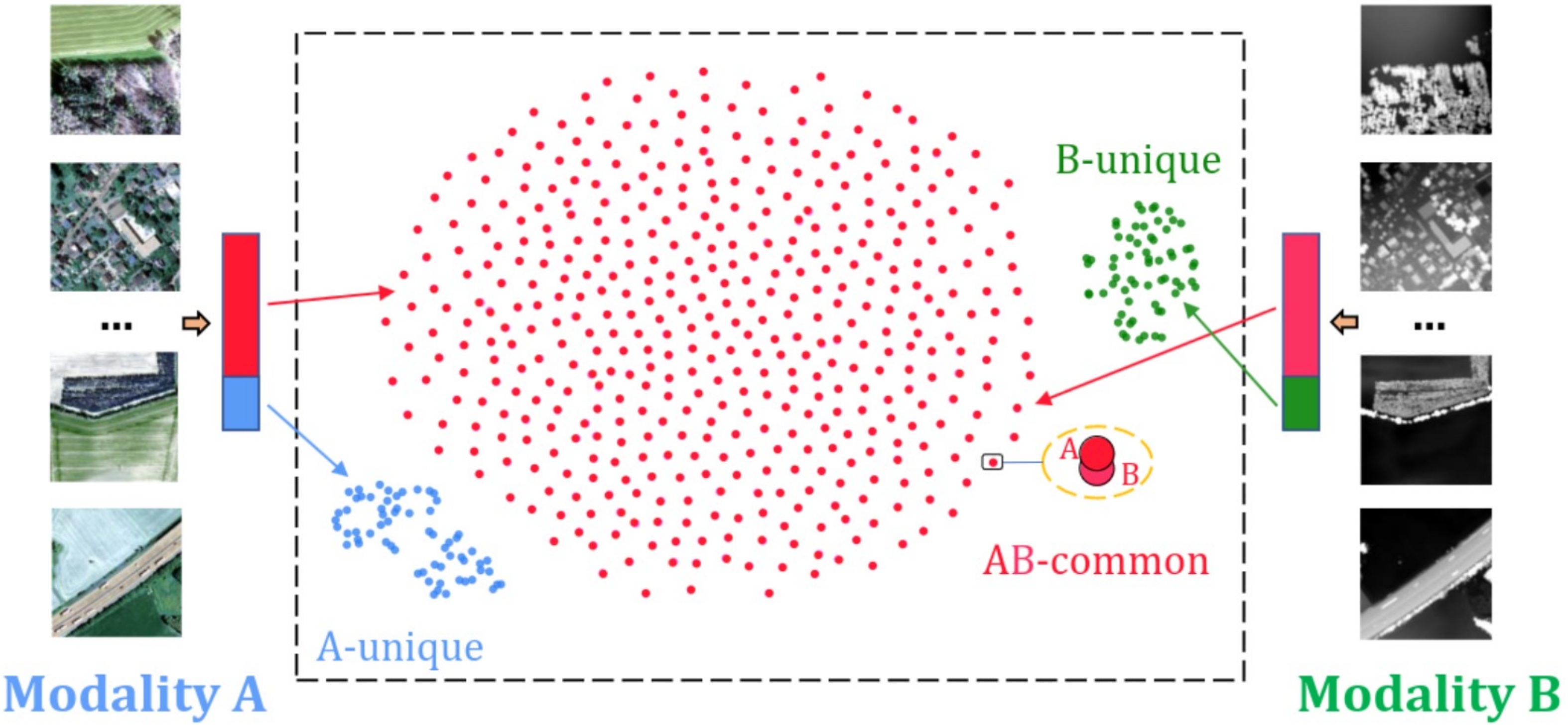
	100%			1%		
	S1	S2	S1+S2	S1	S2	S1+S2
Random	54.6	62.0	64.5	52.7	59.0	62.4
DINO-S1/2	76.2	86.0	–	68.7	82.0	–
DINO-MM	**79.5**	**87.1**	87.1	**75.3**	**82.9**	**82.8**
Supervised	77.1	86.7	**88.6**	63.7	73.6	75.0

Random: random weight initialization (no training)
Supervised: (ordinary) supervised learning
DINO-S1/2: SSL with either Sentinel-1 or -2
DINO-MM: SSL on both Sentinel modalities

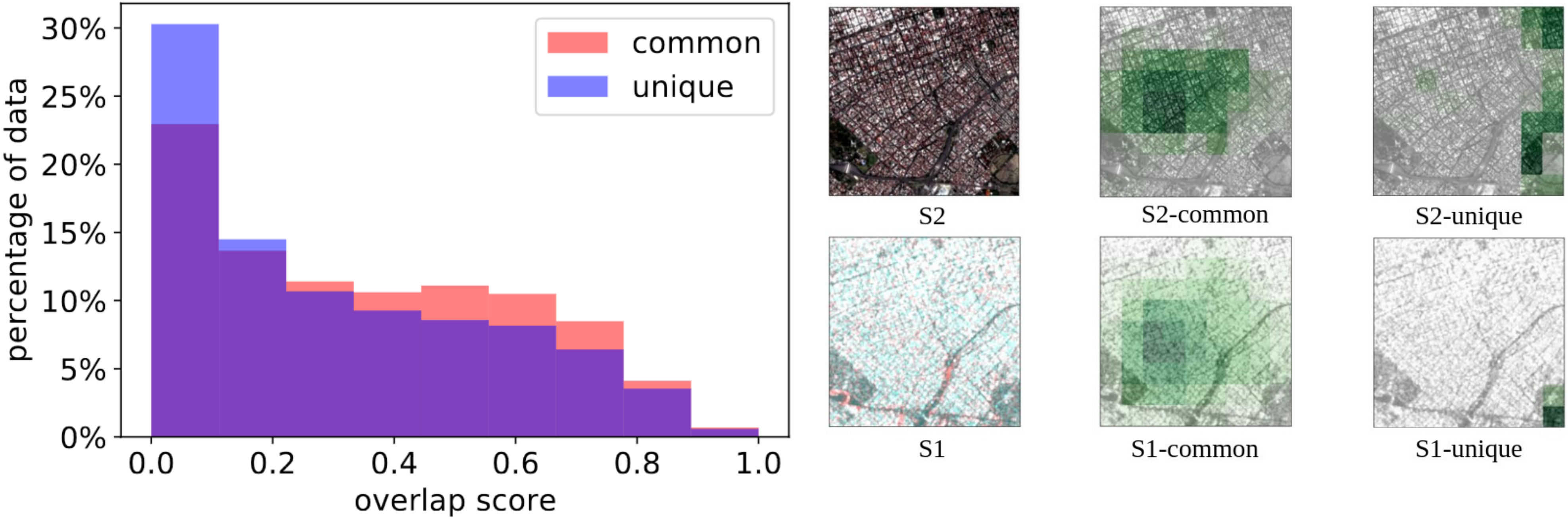
III DeCUR: embedding dimensions in multi-modal remote sensing



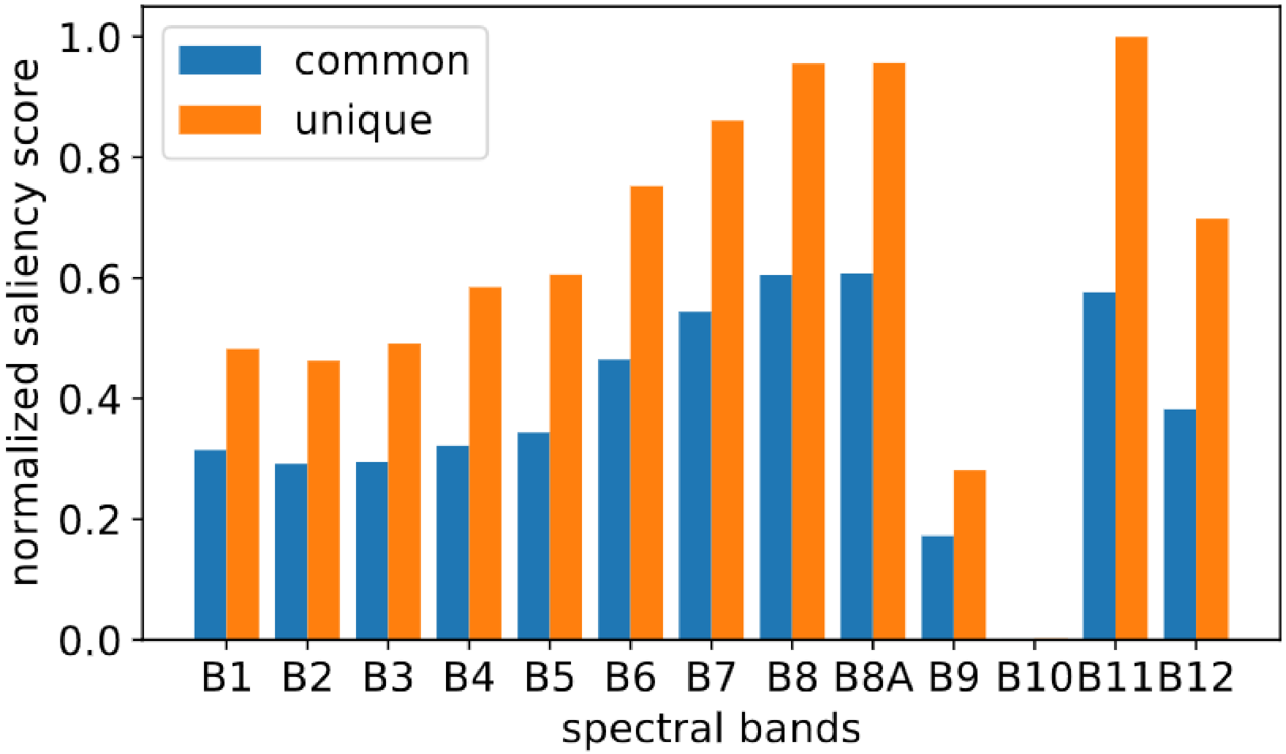
III DeCUR: embedding dimensions in multi-modal remote sensing



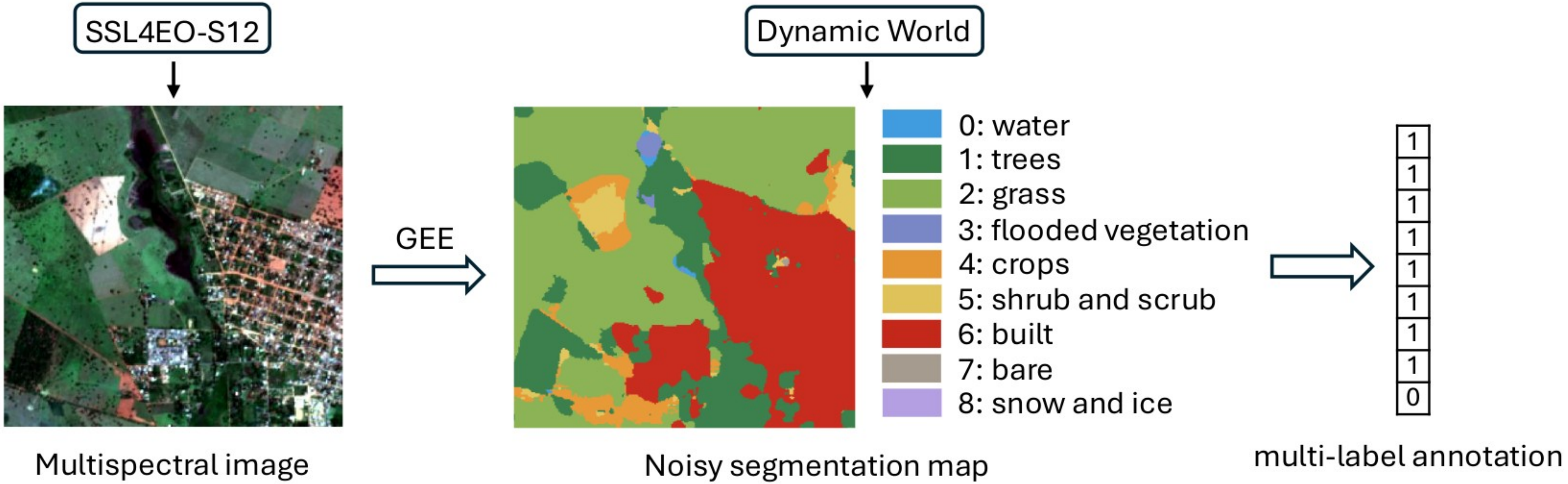
III DeCUR: embedding dimensions in multi-modal remote sensing



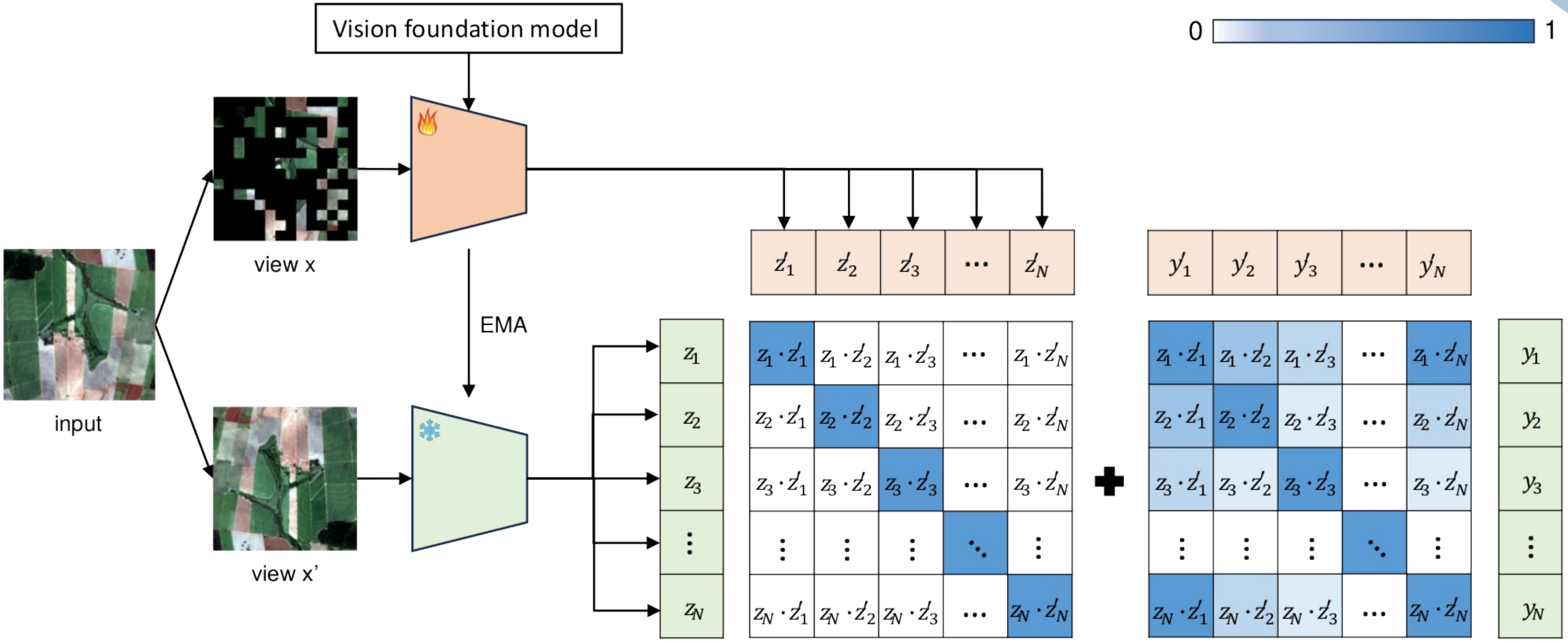
III DeCUR: embedding dimensions in multi-modal remote sensing



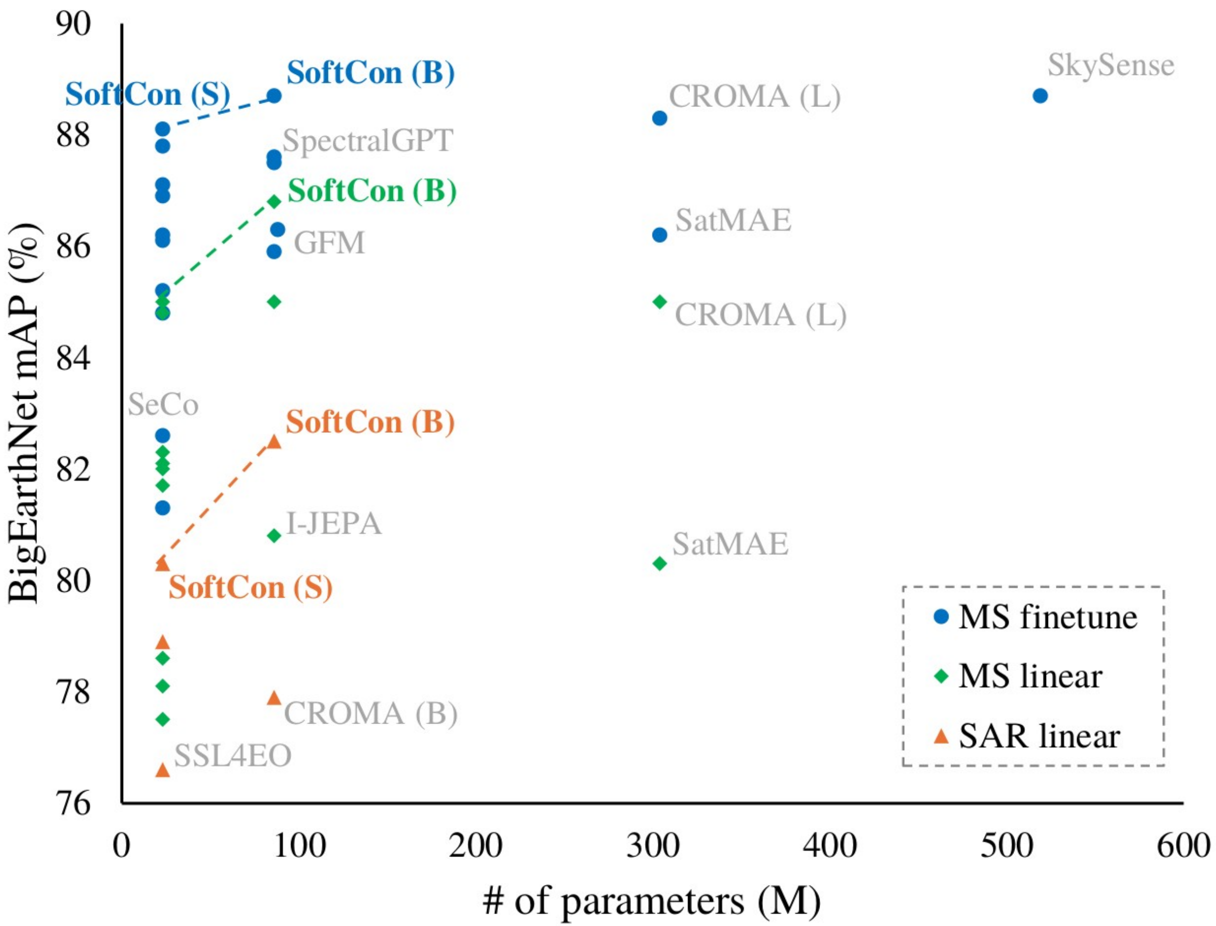
IV SoftCon: exploit land cover labels for foundation model training



IV SoftCon: exploit land cover labels for foundation model training

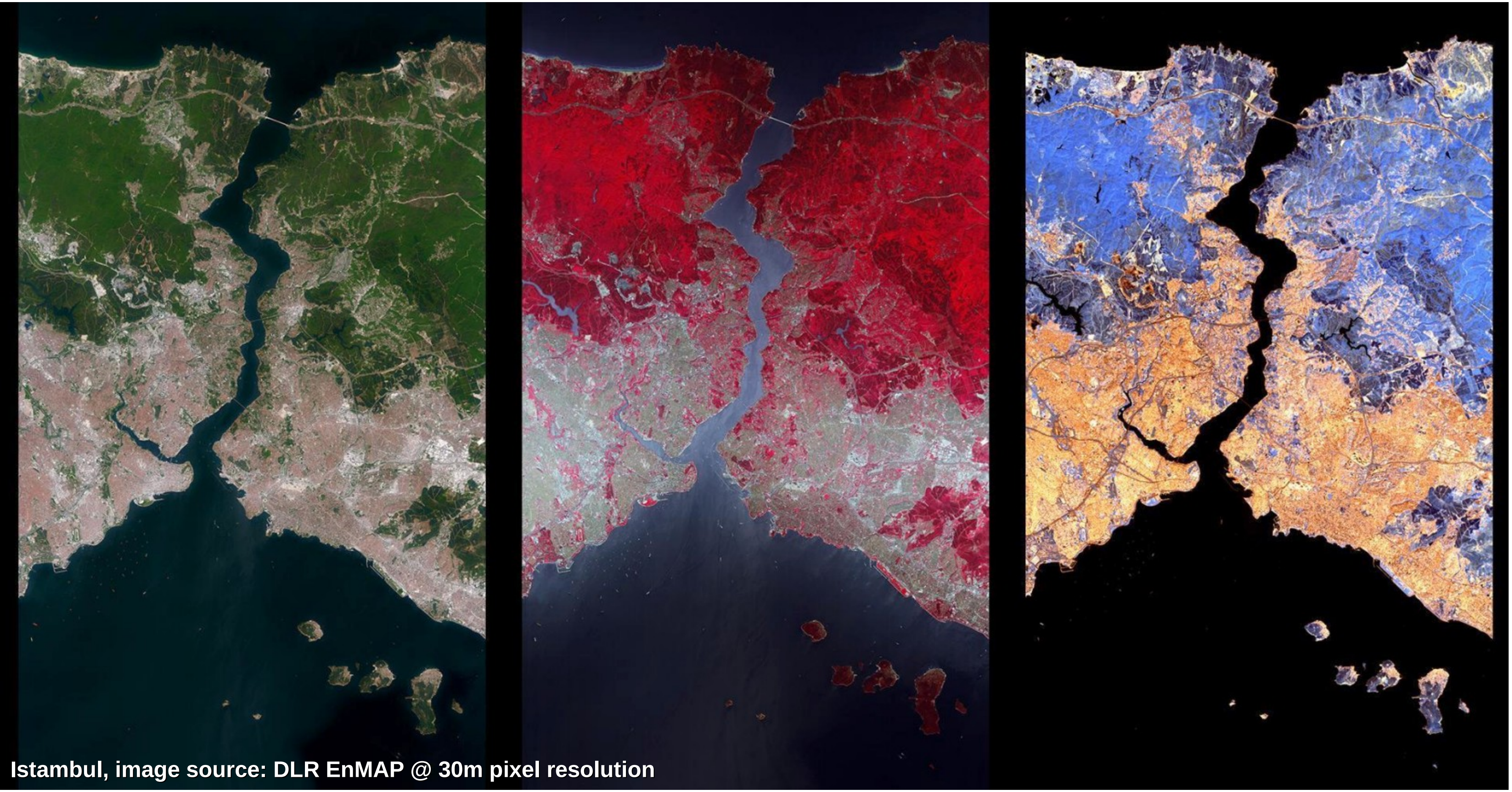


IV SoftCon: exploit land cover labels for foundation model training



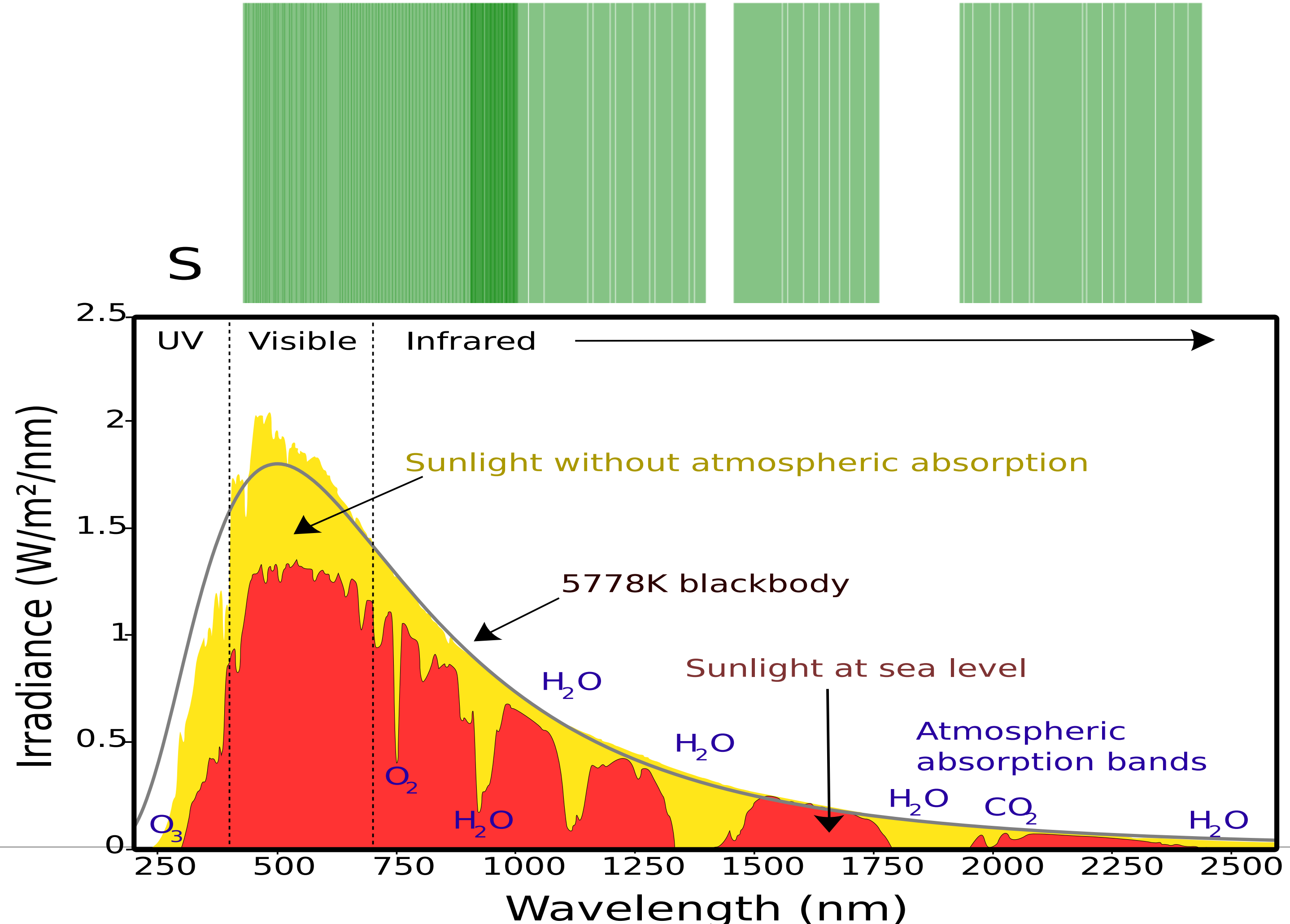
V SpectralEarth: SSL for Hyperspectral Remote Sensing

V SpectralEarth: EnMAP operated by the German Aerospace Center



Istanbul, image source: DLR EnMAP @ 30m pixel resolution

V SpectralEarth: EnMAP spectral channels



DLR EnMAP:
224 hyperspectral bands
of bandwidth ~6nm

top-of-atmosphere solar spectrum

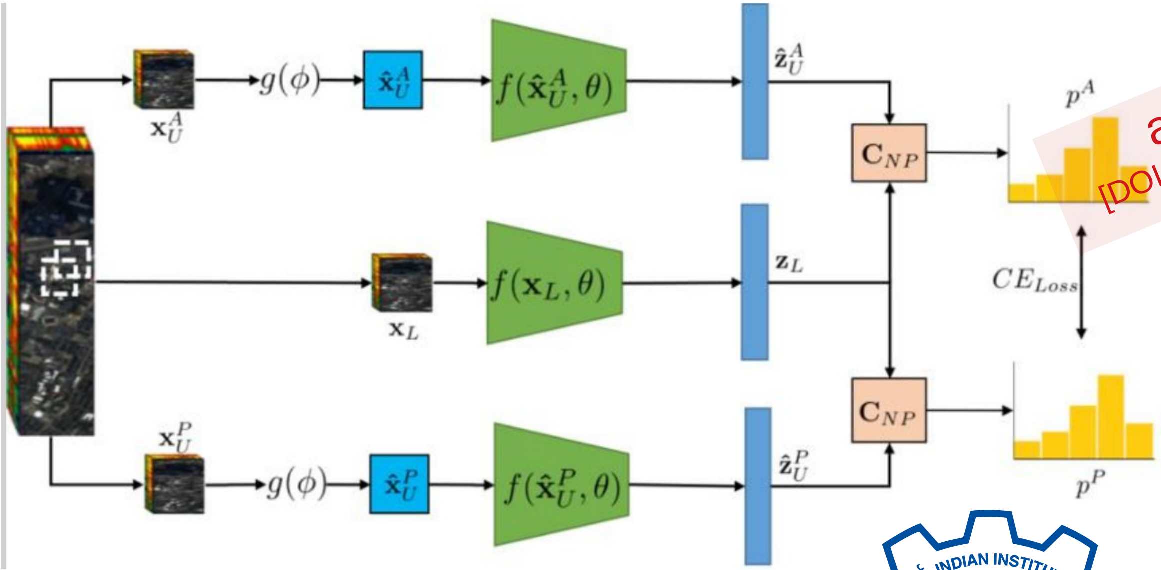
solar spectrum at Earth surface
as impacted by atmospheric absorption

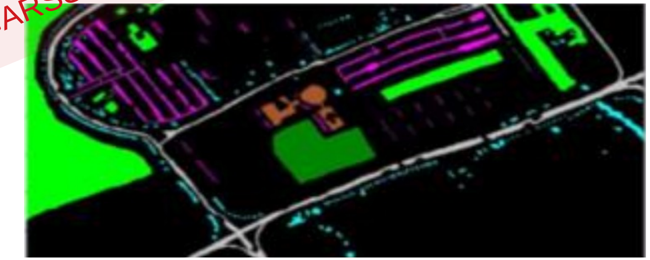
solar spectra image source:
https://en.wikipedia.org/wiki/Sunlight

some (Land Cover Classification) works beyond *EvoLand*

V SpectralEarth: Land Cover assisted SSL for Hyperspectral Earth Observation

an IGARSS Best Paper
[DOI:10.1109/IGARSS52108.2023.10282971]



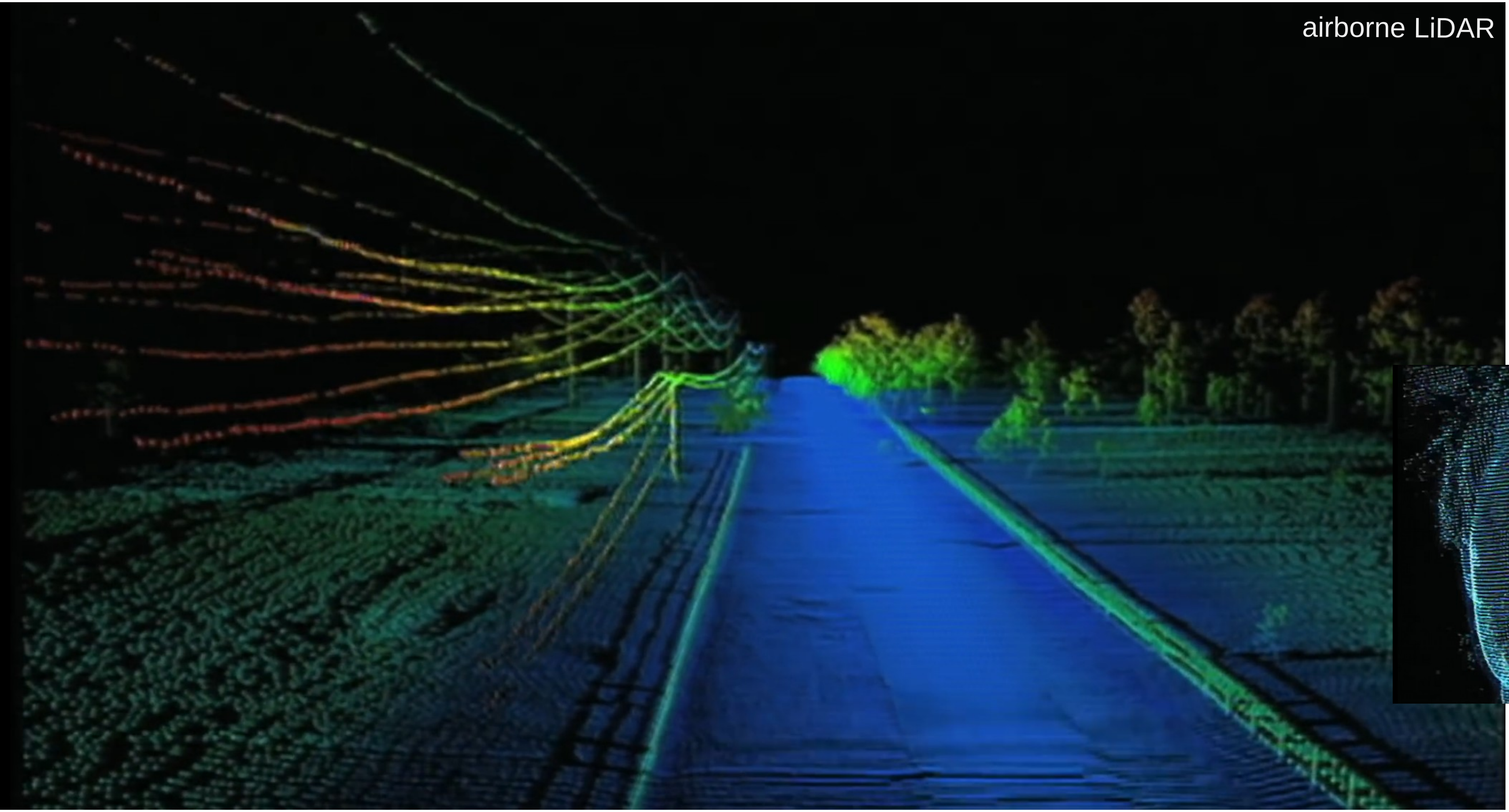


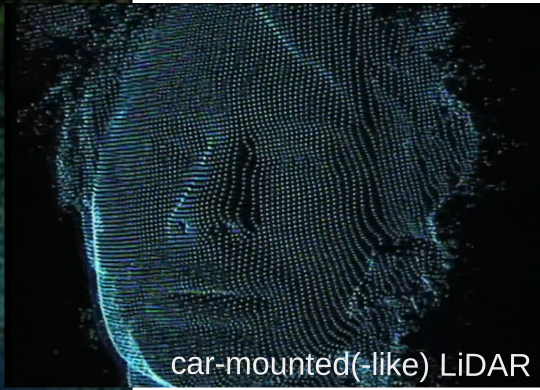
Encoder	Augmentations	Classifier	Accuracy
CNN	No	Supervised	79.10%
Barlow Twins [2]	Yes	Fine Tune	77.14%
None	No	Linear	64.21%
PAWS trained	Yes	Linear	80.92%
PAWS trained	Yes	Fine Tune	**83.12%**
None	No	SNN	32.63%
Untrained	No	SNN	34.28%
PAWS trained	Yes	SNN	**75.31%**

PhD student: **Shivam Pande**

HELMHOLTZ AI

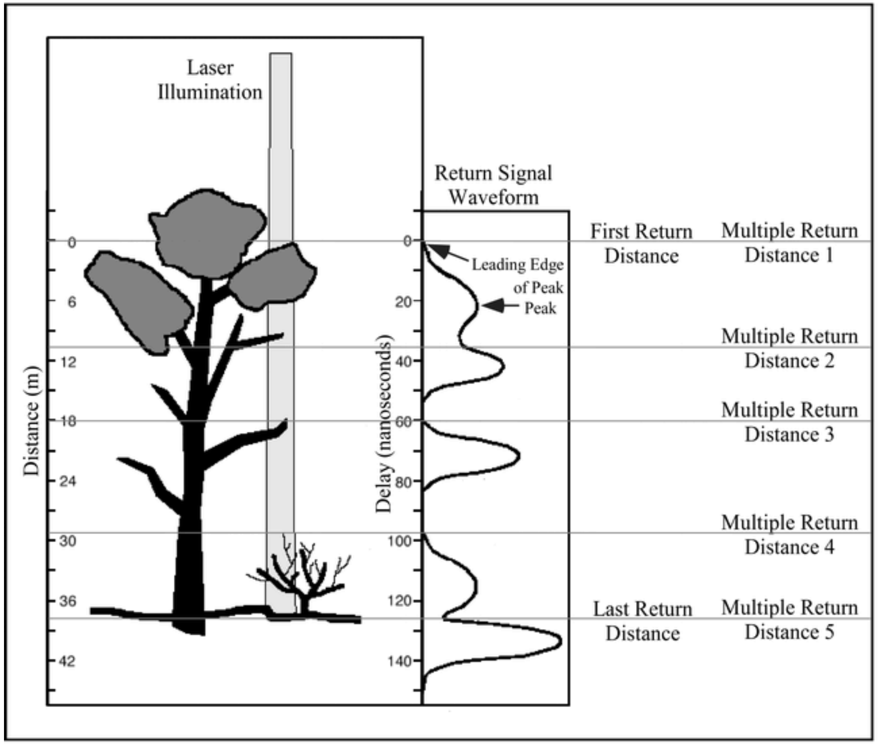
VI LiDAR in popular music / the arts





track & screenshots: copyright by Radiohead – House of Cards (source: YouTube)

VI a remote sensing modality worth the effort?

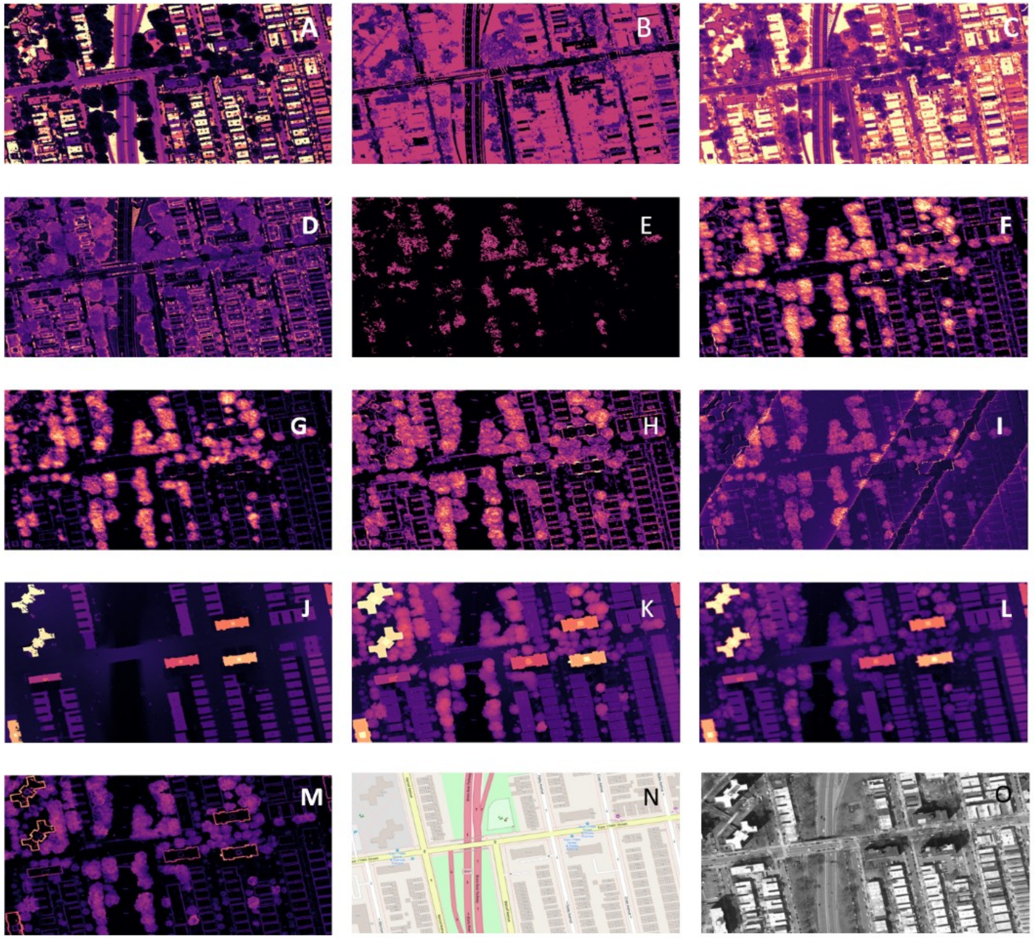


LiDAR cartoon: Lefsky et al. (2002), *Elon on LiDAR screenshots*: YouTube clip Elon Musk says losers use LiDAR. [Explanation video]

VI LiDAR data harmonization by raster statistics

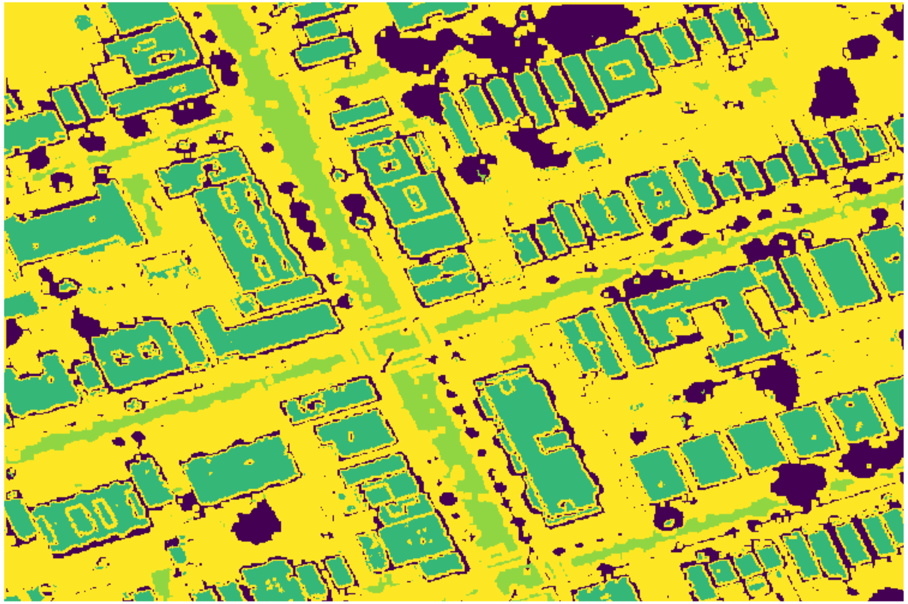
POINT CLOUD STATISTICS

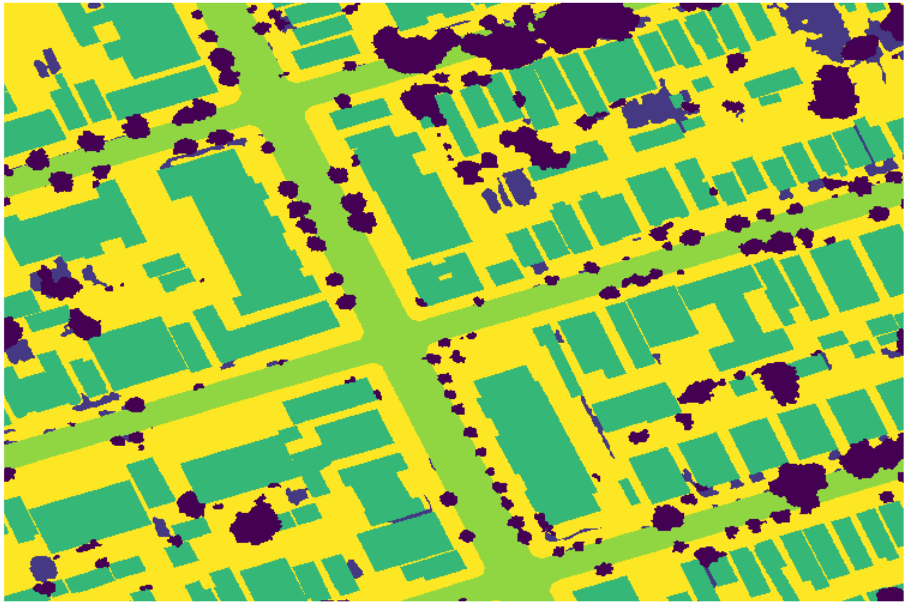
attribute	statistics	Fig. index
reflectance r	minimum r_-	A
	maximum r_+	B
	mean $\bar{r}$	C
	standard deviation r_Δ	D
count c	minimum c_-	E
	maximum c_+	F
	mean $\bar{c}$	G
	standard deviation c_Δ	H
	sum $\sum$	I
elevation e	minimum e_-	J
	maximum e_+	K
	mean $\bar{e}$	L
	standard deviation e_Δ	M



HELMHOLTZ AI

VI AutoGeoLabel: generation of rule-based annotations

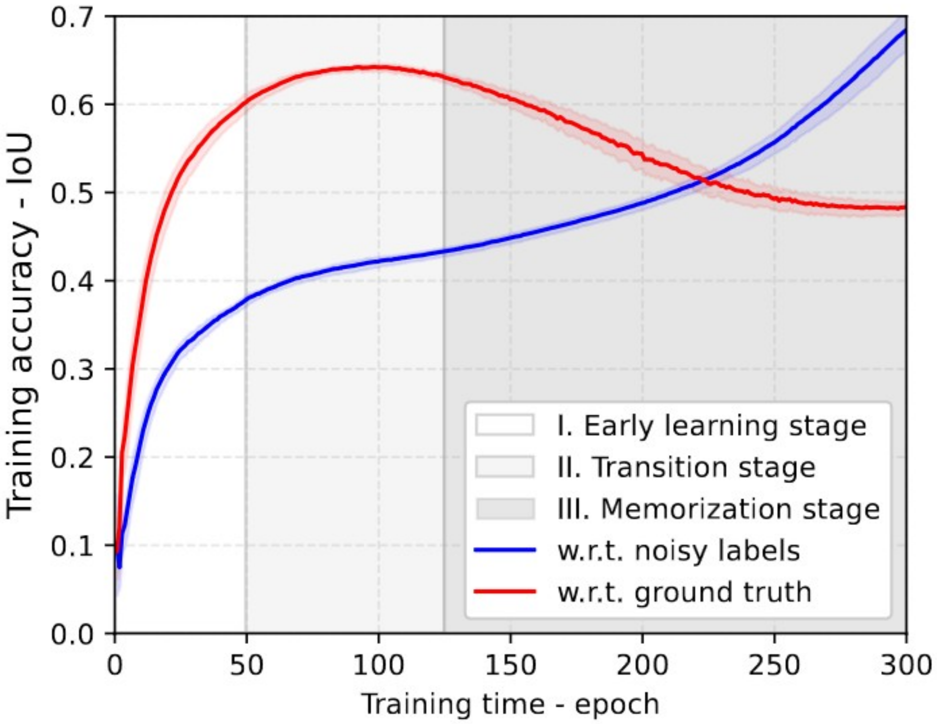


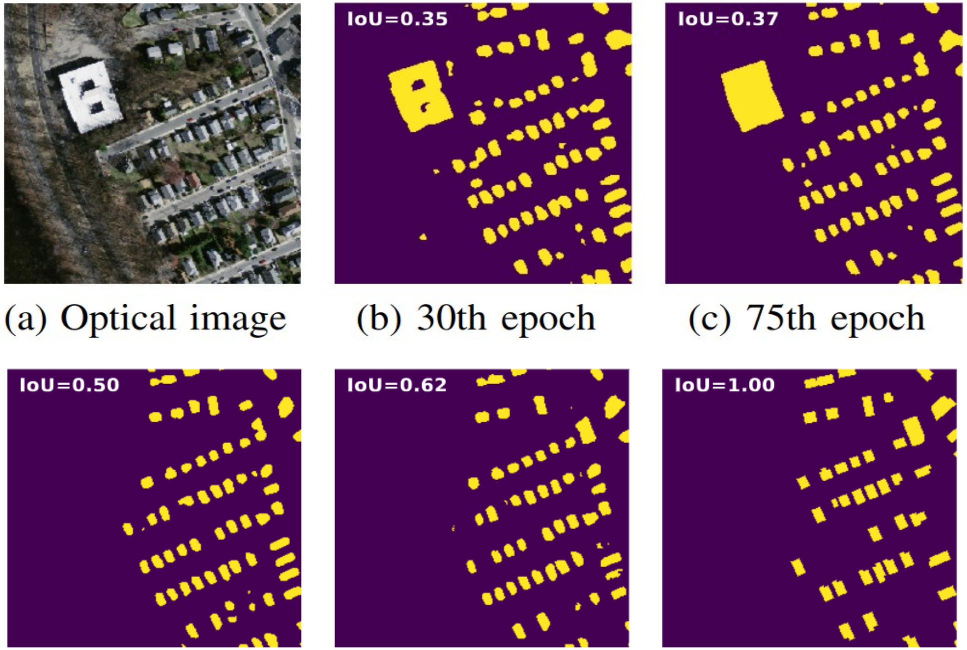


LABELING RULES FROM LIDAR STATISTICS

class	pseudo (R,G,B)	binary classification rule
buildings	(e_-, e_Δ, e_+)	$e_- > \langle e_- \rangle \wedge e_\Delta < \langle e_\Delta \rangle \wedge e_+ > \langle e_+ \rangle$
vegetation	$(c_+, e_\Delta, c_\Delta)$	$c_+ > \langle c_+ \rangle \wedge e_\Delta > \langle e_\Delta \rangle \wedge c_\Delta > \langle c_\Delta \rangle$
roads	$(r_-, \overline{r}, e_-)$	$r_- > .1 r_{\max} \wedge \overline{r} < .6 r_{\max} \wedge e_- < .1 e_{\max}$

VI AIO2: Adaptive Online Object-wise correction

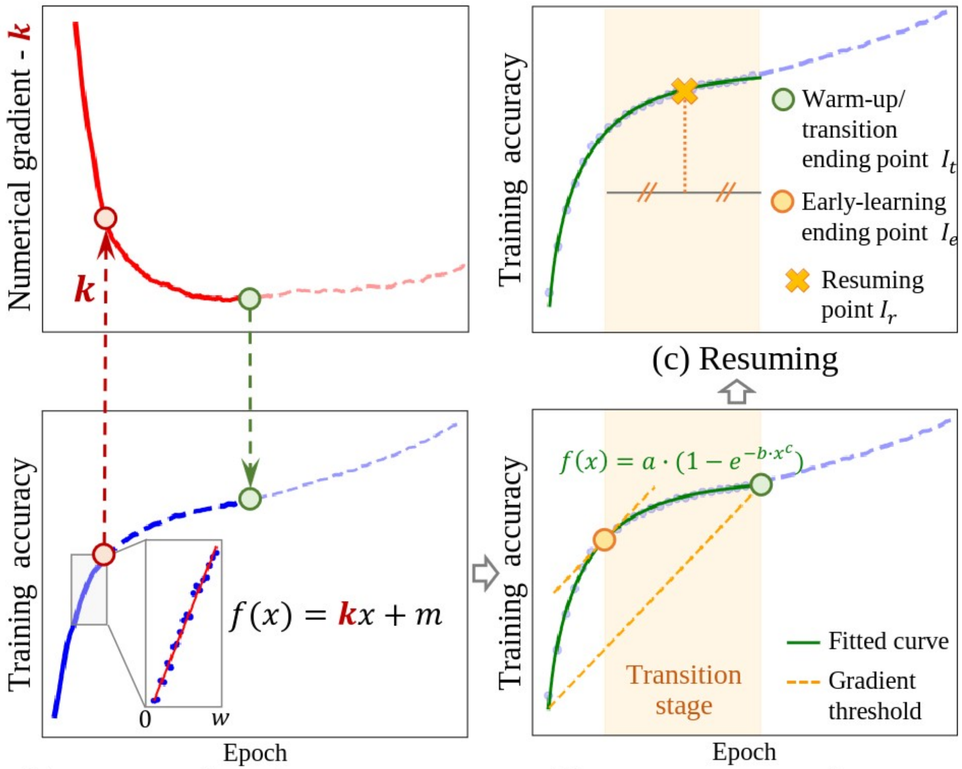




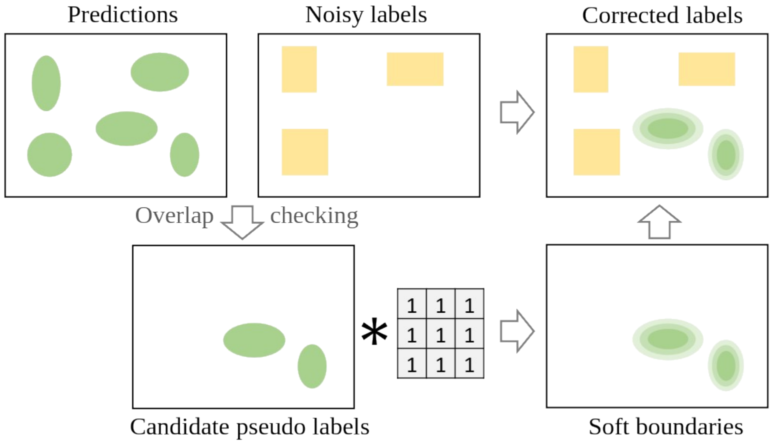
(a) Optical image (b) 30th epoch (c) 75th epoch

(d) 150th epoch (e) 300th epoch (f) Noisy labels

VI AIO2: Adaptive Online Object-wise correction

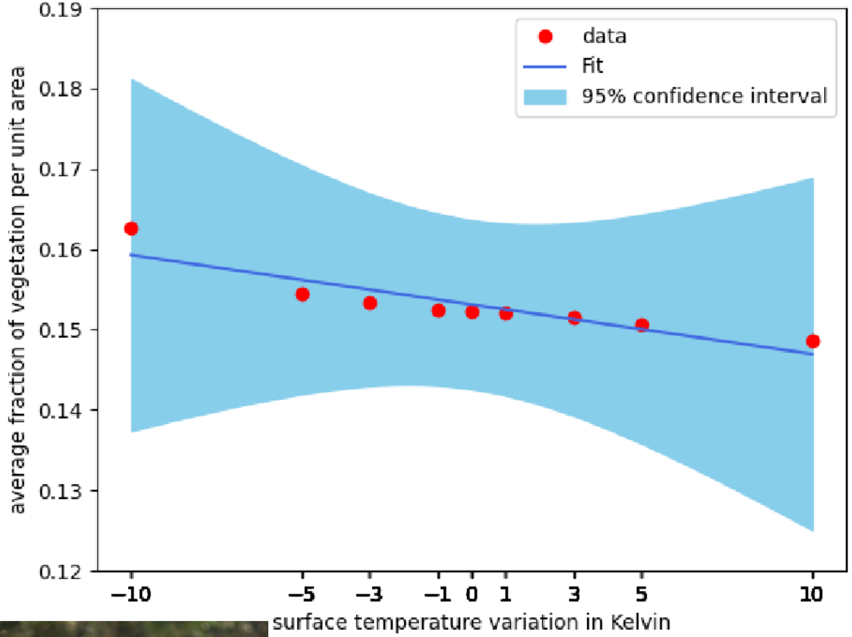


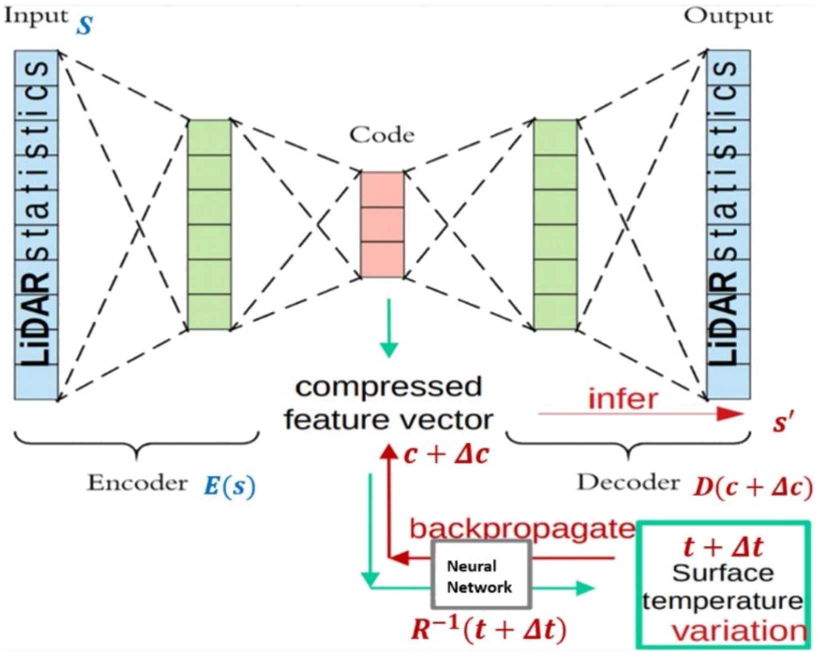
(a) Numerical gradient estimation (b) Transition stage detection



IoU (%)		Final		
α_0		0.3	0.5	0.7
Baseline	Train w/ GT	72.47		
	Train w/ noisy labels	58.06±2.13	40.43±2.17	22.84±0.48
Regularization	Consistency constraint	59.94±1.69	39.59±0.66	22.90±0.86
	Bootstrapping	66.63±0.19	63.23±1.01	56.11±0.58
	Noisy label regularization	69.71±0.58	66.88±0.51	60.09±1.14
Correction	Pixel-wise	68.45±0.60	63.71±0.42	56.63±1.18
	Adaptive pixel-wise	69.06±0.31	66.29±0.35	60.25±1.21
	AIO2 (proposed)	**71.56±0.29**	**70.47±0.17**	**65.45±0.27**

VI DeepLCZChange: Estimating the Climate Impact on Cities





$$s' = D(c+\Delta c) = D\left(E(s) + R^{-1}\left[R(E(s)) + \Delta t\right]\right)$$
$$= s'(s, \Delta t) \quad .$$

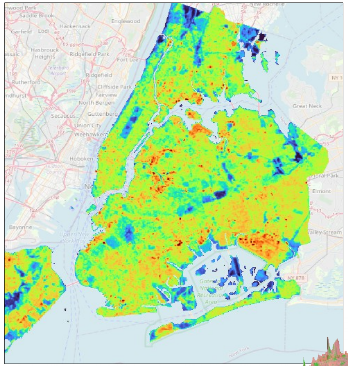
$$\vec{g} = \nabla R \in \mathbb{R}^n$$

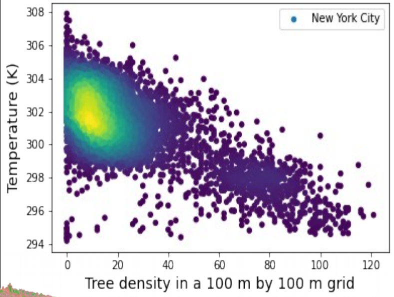
$$\Delta\vec{c} = \zeta\vec{g}/g \qquad \Delta t = \Delta\vec{c}\cdot\vec{g} = \zeta\vec{g}\cdot\vec{g}/g = \zeta g$$

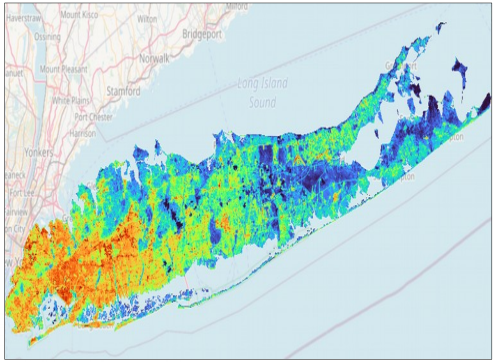
$$\Delta\vec{c}(R, \Delta t) = \frac{\Delta t}{g^2}\vec{g}$$

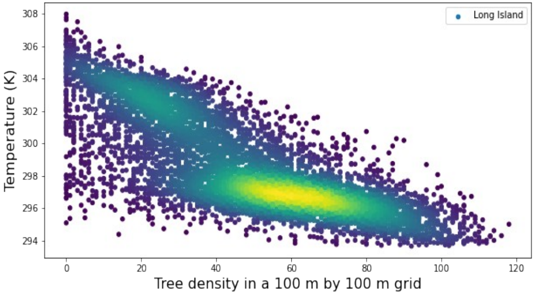
TUM master student: **Wenlu** Sun

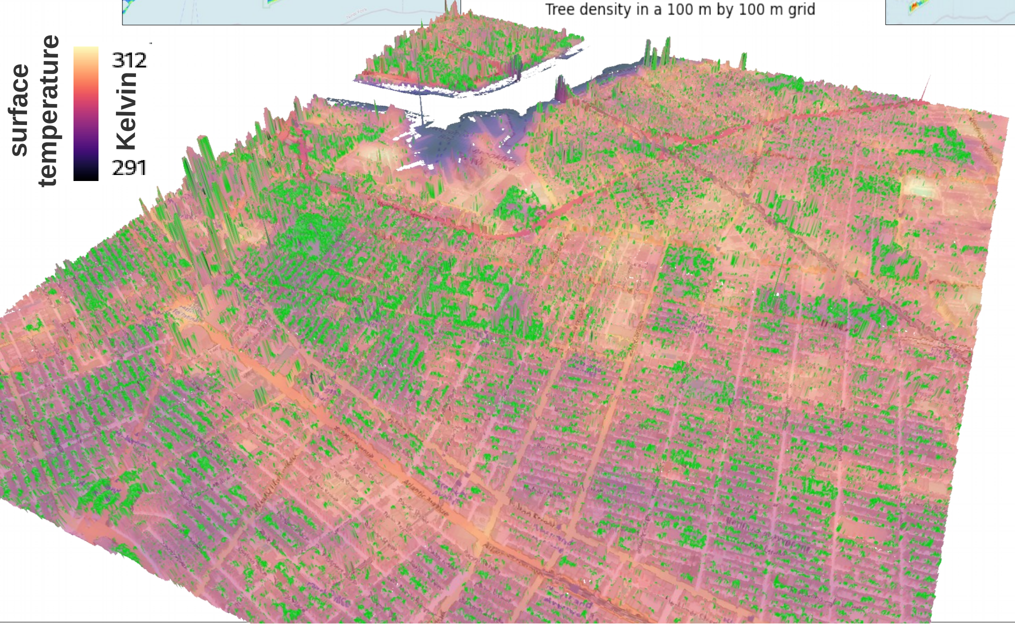
VI LCZ: Inspecting the Correlation of LCZs and Heat Islands

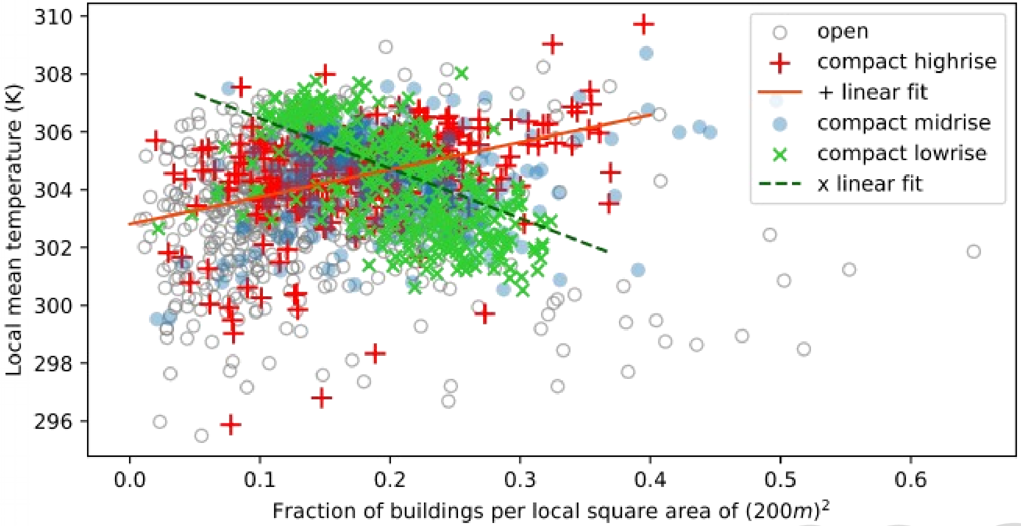












VI AutoLCZ: Local Climate Zones autogenerated from LiDAR

No. LCZ	Built type	Geometric and surface cover parameters							Thermal, radiative, and metabolic parameters		
		SVF[a]	AR[b]	BSF[c]	ISF[d]	PSF[e]	HRE[f]	TRC[g]	SAD[h]	SAL[i]	AHO[j]
1	Compact high-rise	0.2-0.4	>2	40-60	40-60	<10	>25	8	1,500-1,800	0.10-0.20	50-300
2	Compact midrise	0.3-0.6	0.75-2	40-70	30-50	<20	10-25	6-7	1,500-2,200	0.10-0.20	<75
3	Compact low-rise	0.2-0.6	0.75-1.5	40-70	20-50	<30	3-10	6	1,200-1,800	0.10-0.20	<75
4	Open high-rise	0.5-0.7	0.75-1.25	20-40	30-40	30-40	>25	7-8	1,400-1,800	0.12-0.25	<50
5	Open midrise	0.5-0.8	0.3-0.75	20-40	30-50	20-40	10-25	5-6	1,400-2,000	0.12-0.25	<25
6	Open low-rise	0.6-0.9	0.3-0.75	20-40	20-50	30-60	3-10	5-6	1,200-1,800	0.12-0.25	<25
7	Lightweight low-rise	0.2-0.5	1-2	60-90	<20	<30	2-4	4-5	800-1,500	0.15-0.35	<35
8	Large low-rise	>0.7	0.1-0.3	30-50	40-50	<20	3-10	5	1,200-1,800	0.15-0.25	<50
9	Sparsely built	>0.8	0.1-0.25	10-20	<20	60-80	3-10	5-6	1,000-1,800	0.12-0.25	<10
10	Heavy industry	0.6-0.9	0.2-0.5	20-30	20-40	40-50	5-15	5-6	1,000-2,500	0.12-0.20	>300

[a] Sky View Factor (SVF): ratio of the amount of sky hemisphere visible from ground level to that of an unobstructed hemisphere; [b] Aspect ratio (AR): mean height-to-width ratio of street canyons (LCZs 1-7), and building spacing (LCZs 8-10); [c] Building Surface Fraction (BSF): ratio of building area to total area; [d] Impervious Surface Fraction (ISF): ratio of impervious area to total area; [e] Pervious Surface Fraction (PSF): ratio of pervious area to total area; [f] Height of Roughness Elements (HRE): geometric average of building heights; [g] Terrain Roughness Class (TRC): classification of effective terrain roughness for city and country landscapes; [h] Surface ADmittance (SAD); [i] Surface ALbedo (SAL); [j] Anthropogenic Heat Output (AHO)

VI AutoLCZ: Local Climate Zones autogenerated from LiDAR

$\hat{z}$ = LCZ	x_1 = BSF	x_2 = ISF	x_3 = PSF	x_4 = HRE
1	0.4-0.6 **[0.30, 0.65]**	0.4-0.6 **[0.25, 0.55]**	<0.1 **[0.00, 0.26]**	>25 **[18.98, 69.14]**
2	0.4-0.7 **[0.18, 0.50]**	0.3-0.5 **[0.31, 0.62]**	<0.2 **[0.04, 0.32]**	10-25 **[4.81, 24.25]**
3	0.4-0.7 **[0.22, 0.42]**	0.2-0.5 **[0.35, 0.58]**	<0.3 **[0.09, 0.32]**	3-10 **[4.20, 17.72]**
4	0.2-0.4 **[0.05, 0.34]**	0.3-0.4 **[0.21, 0.58]**	0.3-0.4 **[0.17, 0.61]**	>25 **[2.20, 29.78]**
5	0.2-0.4 **[0.11, 0.37]**	0.3-0.5 **[0.22, 0.53]**	0.2-0.4 **[0.21, 0.54]**	10-25 **[6.44, 25.09]**
6	0.2-0.4 **[0.04, 0.28]**	0.2-0.5 **[0.20, 0.56]**	0.3-0.6 **[0.23, 0.68]**	3-10 **[0.09, 18.11]**
8	0.3-0.5 **[0.04, 0.59]**	0.4-0.5 **[0.31, 0.81]**	<0.2 **[0.00, 0.27]**	3-10 **[3.25, 12.21]**
10	0.2-0.3 **[0.03, 0.49]**	0.2-0.4 **[0.32, 0.81]**	0.4-0.5 **[0.00, 0.30]**	5-15 **[2.59, 14.16]**

z = LCZ	Using GT labels		Using noisy labels	
	Given thresholds	**Estimated thresholds**	**Given thresholds**	**Estimated thresholds**
1	43.02	81.12 (+38.10)	56.73	69.13 (+12.40)
2	48.58	54.78 (+ 6.20)	69.11	61.30 (- 7.81)
3	30.56	44.22 (+13.66)	11.07	30.56 (+19.49)
4	8.54	56.66 (+48.12)	7.35	59.36 (+52.01)
5	34.05	50.48 (+16.43)	39.52	40.00 (+ 0.48)
6	60.81	72.42 (+11.61)	21.18	35.36 (+14.18)
8	26.84	82.89 (+56.05)	20.26	86.05 (+65.79)
10	1.41	82.13 (+80.72)	6.11	83.23 (+77.12)
All (OA)	44.85	59.11 (+14.26)	38.91	48.56 (+ 9.65)

VII Conclusions

I data mining for land cover analysis

a. **vast amounts of remote sensing data** call for state-of-the-art **Big Data processing** paradigms

b. **we explore** the use of self-/**weakly-supervised** (deep) **learning** for novel land cover prototypes **in *EvoLand***

II **SSL4EO & Spectral Earth**: feature generation from SAR-optical sensor fusion

a. **SSL in Earth observation** demands **terabyte-scale** diverse **benchmark** datasets

b. **SSL** efficiently **compresses multi-modal data** for **boost** in model **performance**

c. **hyperspectral satellite** datasets enter the **era of Big Data**

III **Auto{GeoLabel,LCZ} & AIO2**: automated segmentation maps multi-modal

a. **LiDAR** provides a valuable **sensor modality** to **automatically generate** semantic segmentation **labels**

b. **deep learning** training **with inaccurate annotation** requires **adaptation of training** procedures

c. **human inspection** of local climate zones may get **automatized by airborne LiDAR**

Q&A Session

beyond, go visit https://conrad-m-albrecht.github.io