Earth Embeddings: Learning Mental Maps in Neural Nets

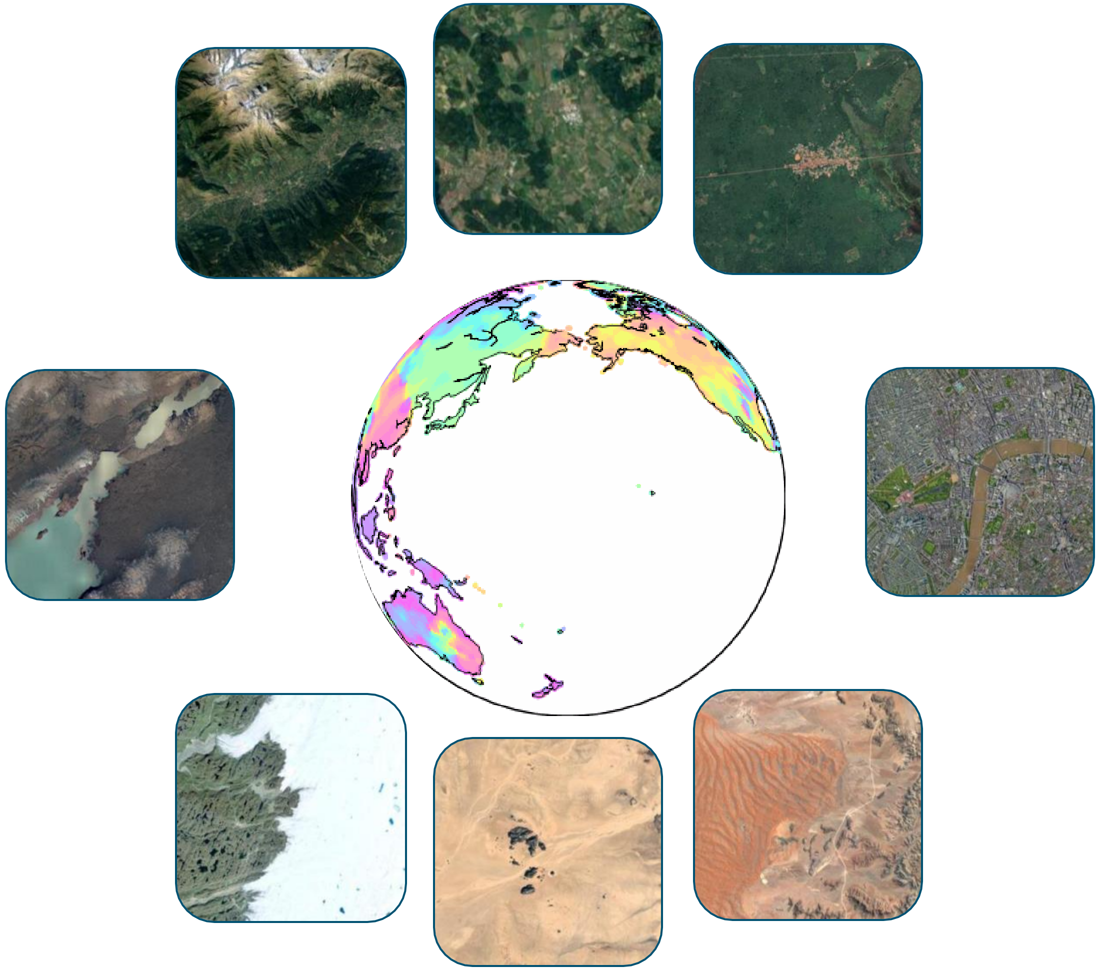
Continents Webinar
The University of Edinburgh

November 12th 2025

Marc Rußwurm

Assistant Professor

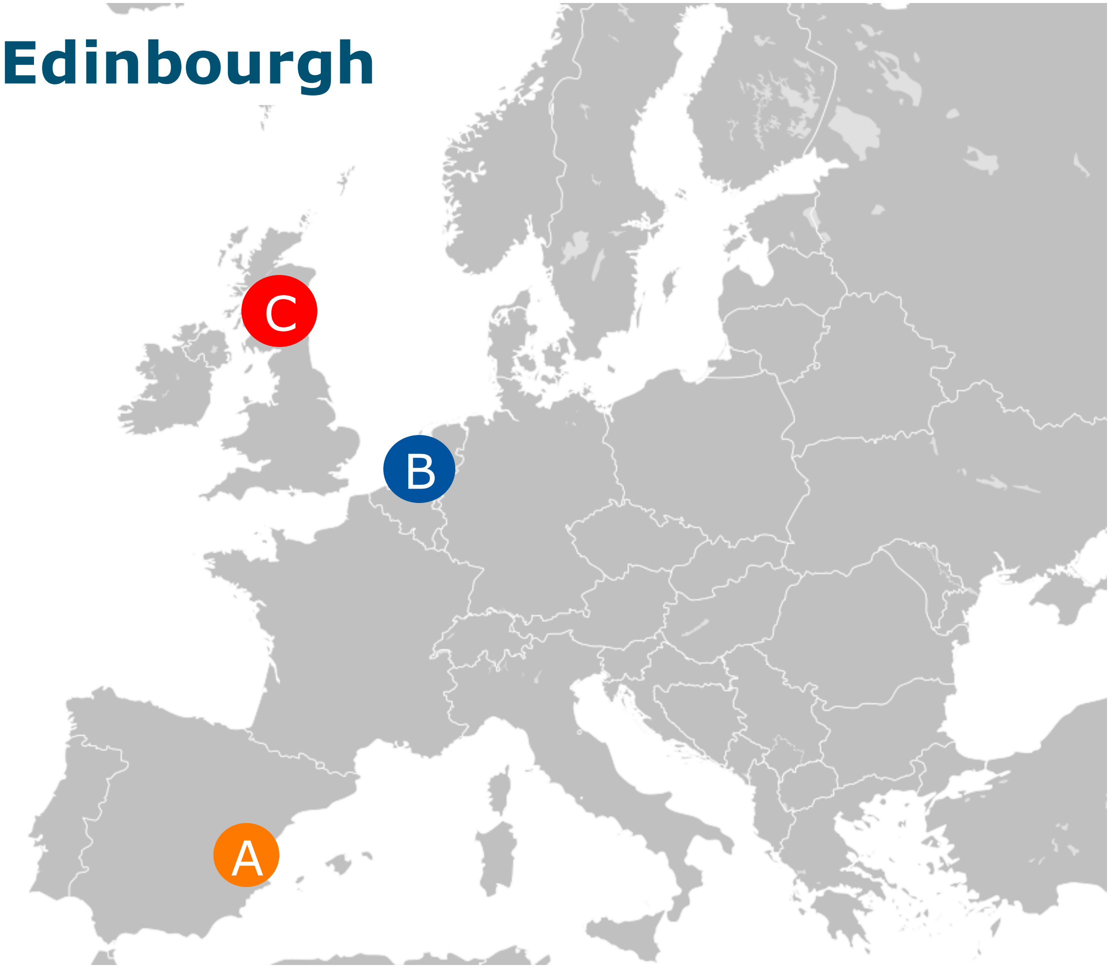
Wageningen University



WAGENINGEN
UNIVERSITY & RESEARCH

Where is this image from? A, B, C?

It’s C, Edinbourgh



Geolocalization involves **Vision** and **Mapping**

Vision

- irregular city blocks,
- Massive stone houses
- Castle on a hill
- Green parks

Geo localization

- Medieval city/Castle 12th–16th centuries
- 18-century street layout (City blocks)
- Mid-large sized city

Location

(Mental) Mapping or Context

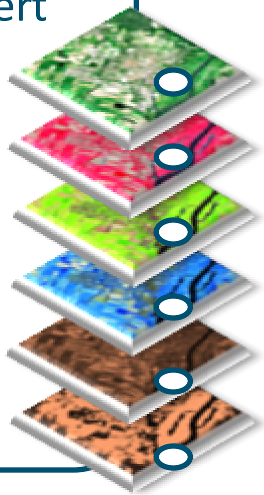
Geospatial Embeddings: Encoding Geospatial Information

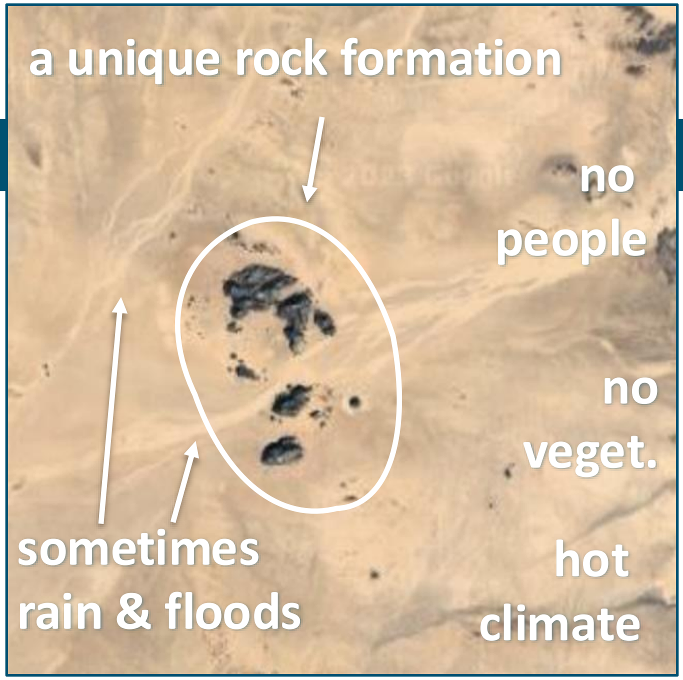
How to describe this place?

with all physical data $\boldsymbol{d}[\lambda, \phi]$ we can **measure**

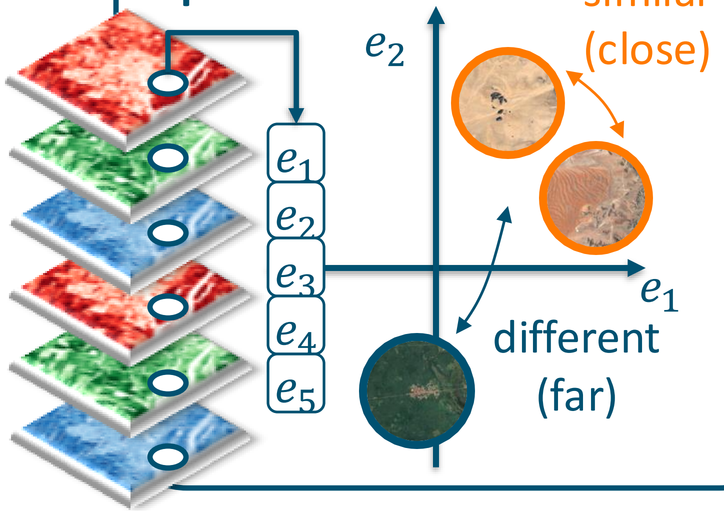
all geodata layers

- d_1 text: “hot, desert”
- d_2 RGB image
- d_3 high temp,
- d_4 elev. over sea
- d_5 low pop. dens
- d_6 eco system
- d_7 low predip.

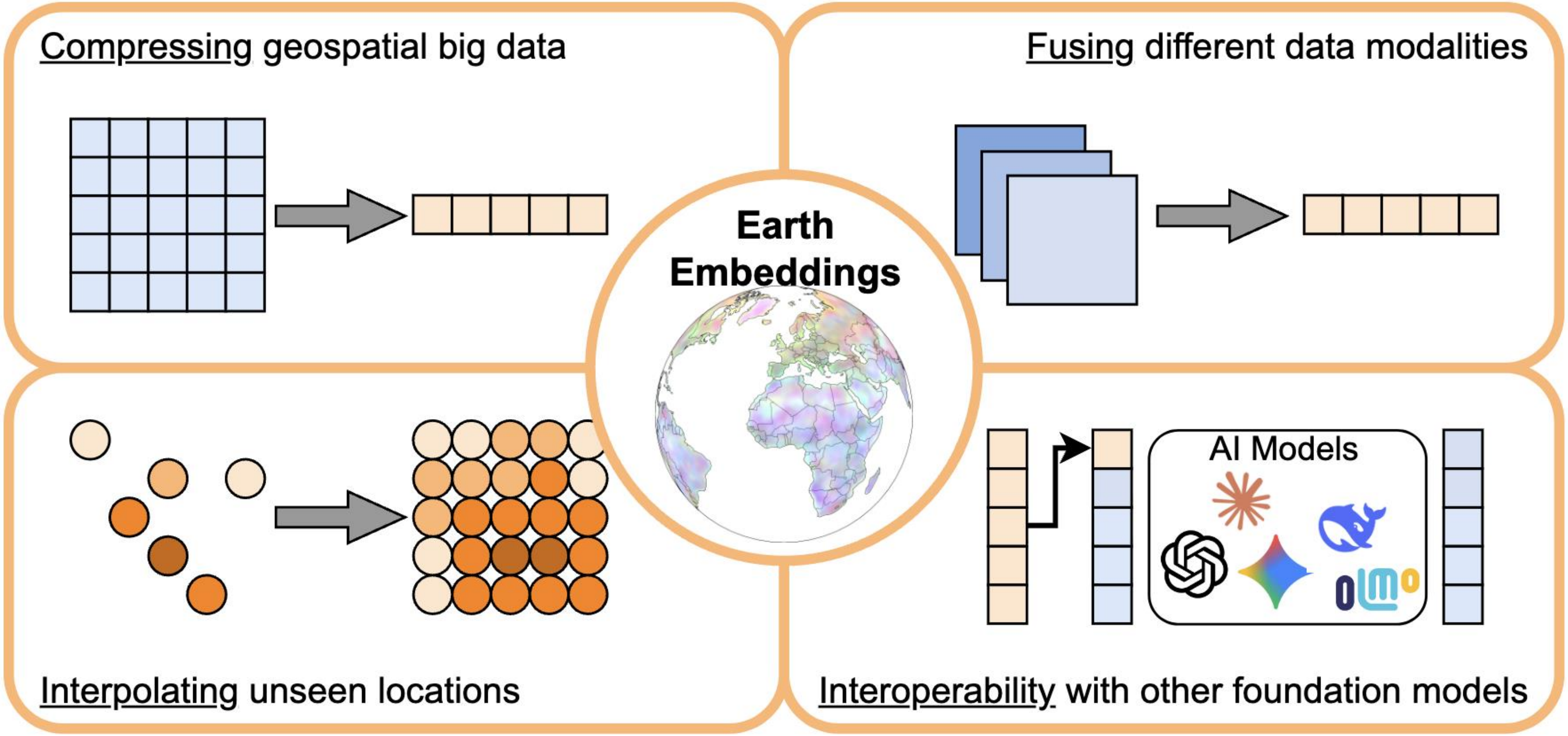




by what makes it **unique** in a metric **embedding** $\boldsymbol{e}[\lambda, \phi]$ space

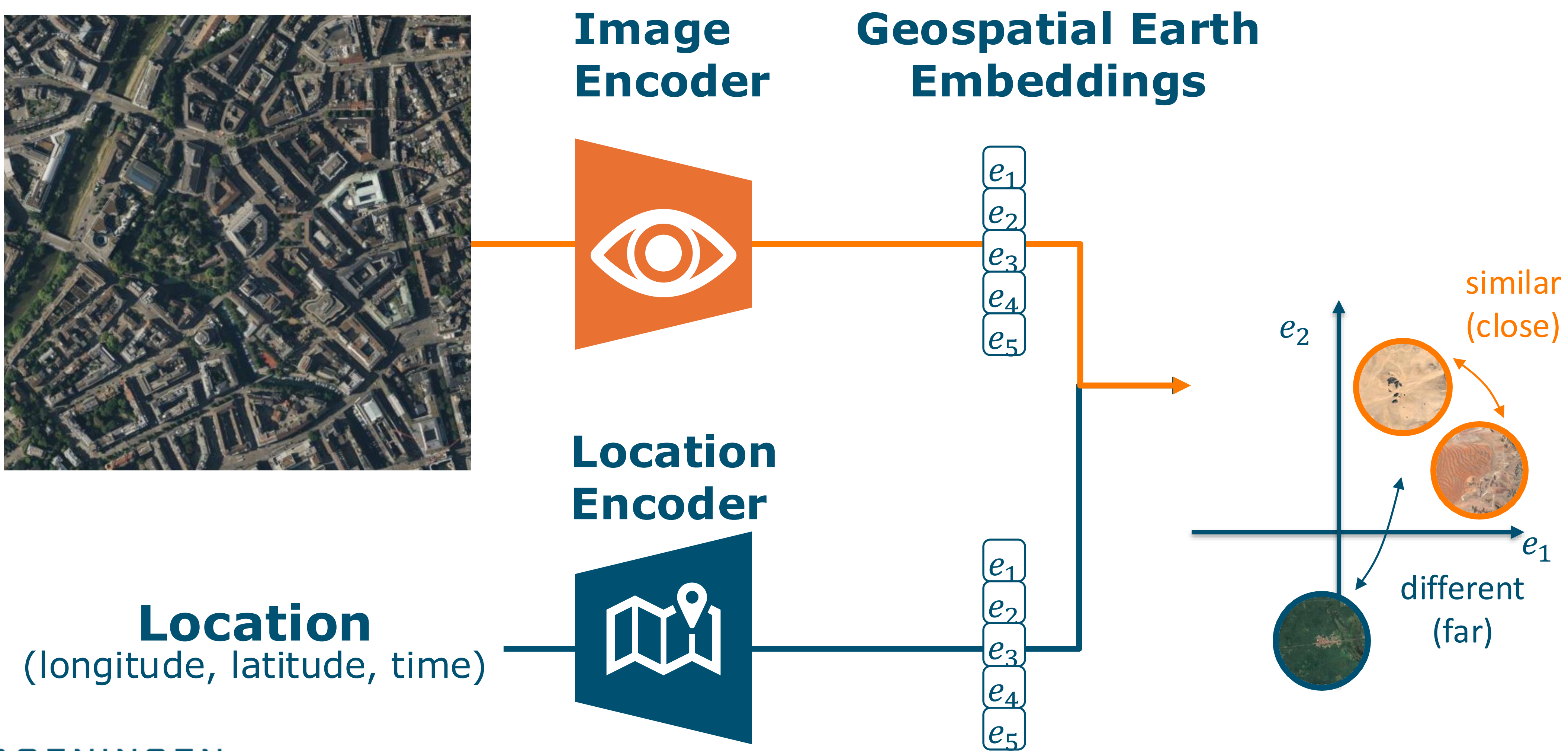


The Functions of Embeddings

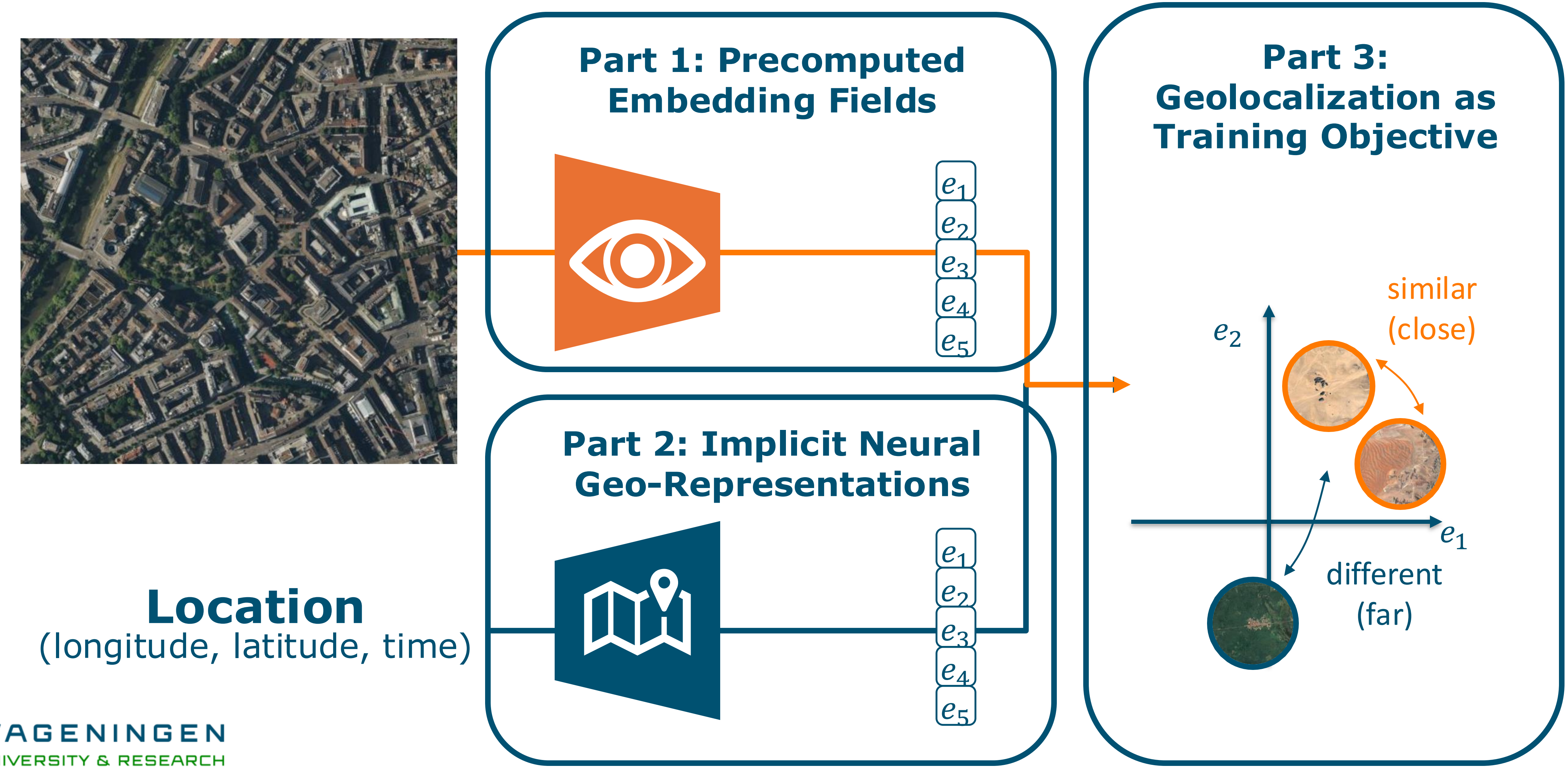


Klemmer K., Rolf. E., Rußwurm M. et al., Earth Embeddings:
Towards AI-centric Representations of our Planet. Perspective in preparation

Geolocalization as Retrieval in an Embedding Space



Outline of this Talk: Learning Mental Maps in Neural Nets



WAGENINGEN
UNIVERSITY & RESEARCH

Outline of this Talk

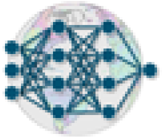
Location
(longitude, latitude, time)

Part 1: Precomputed Embedding Fields

Part 1: Explicit Feature Extraction and Embedding Databases

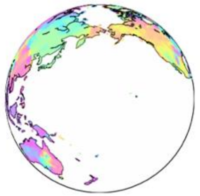
Part 2: Implicit neural Geo-Representations

Part 2: Implicit neural Geo-Representations



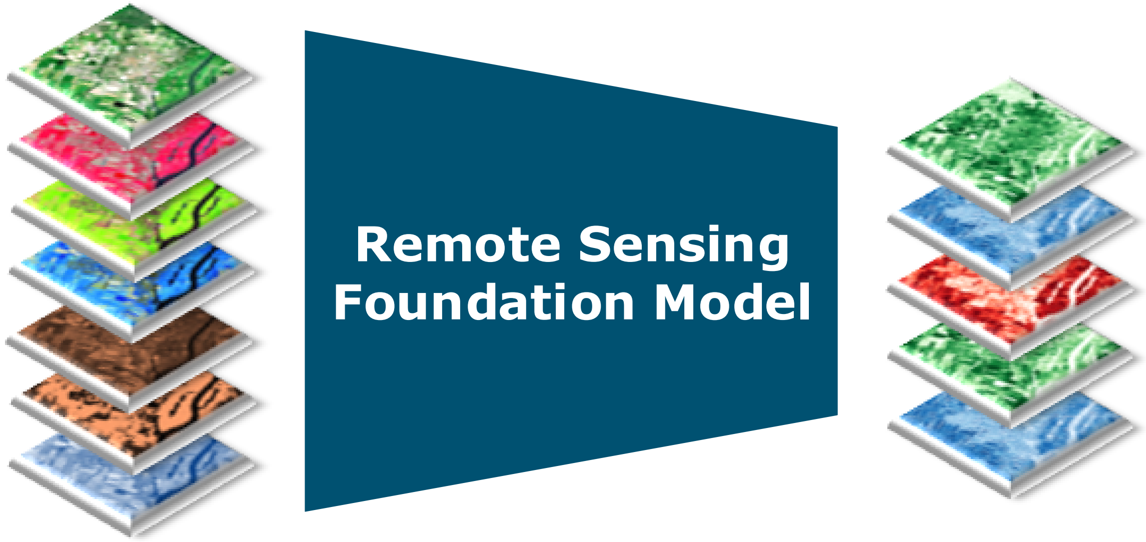
Part 3: Geolocalization as Training Objective

Part 3: Implicit Embedding Models through Geolocalization



WAGENINGEN UNIVERSITY & RESEARCH

Part 1: Explicit Feature Extraction and Embedding Databases



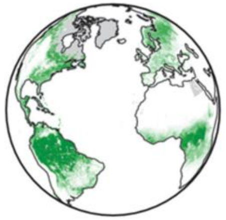
Embedding Databases as pre-computed Features

Embedding Databases as pre-computed Features

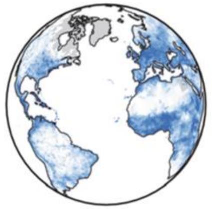
Geospatial database (e.g., raster)

Embedding database (e.g., raster)

Linear Regression or KNN



Downstream Task 1



Downstream Task 2

loading of remote sensing data

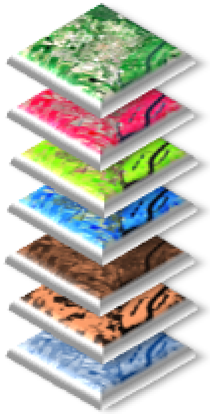
$\boldsymbol{d}[\lambda, \phi]$

storage of Compressed embeddings

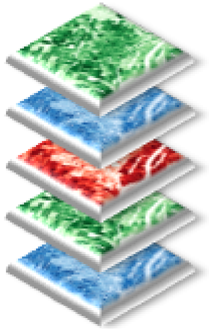
$\boldsymbol{e}[\lambda, \phi]$

Downstream Domain Expertise

Deep Learning Expertise



Remote Sensing Foundation Model



Advantages:

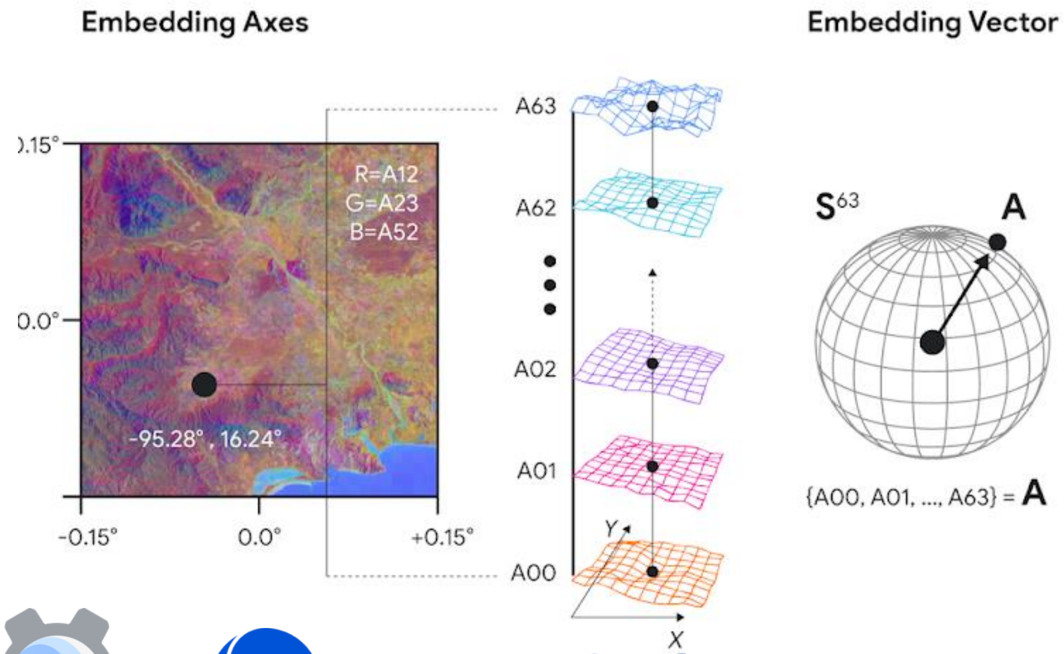
- No deep model inference – just database lookups
- Very high compression rates

WAGENINGEN UNIVERSITY & RESEARCH

Inspired by Konrad Schindler's talk on Foundation Models @LPS 2025

Example Embedding Databases

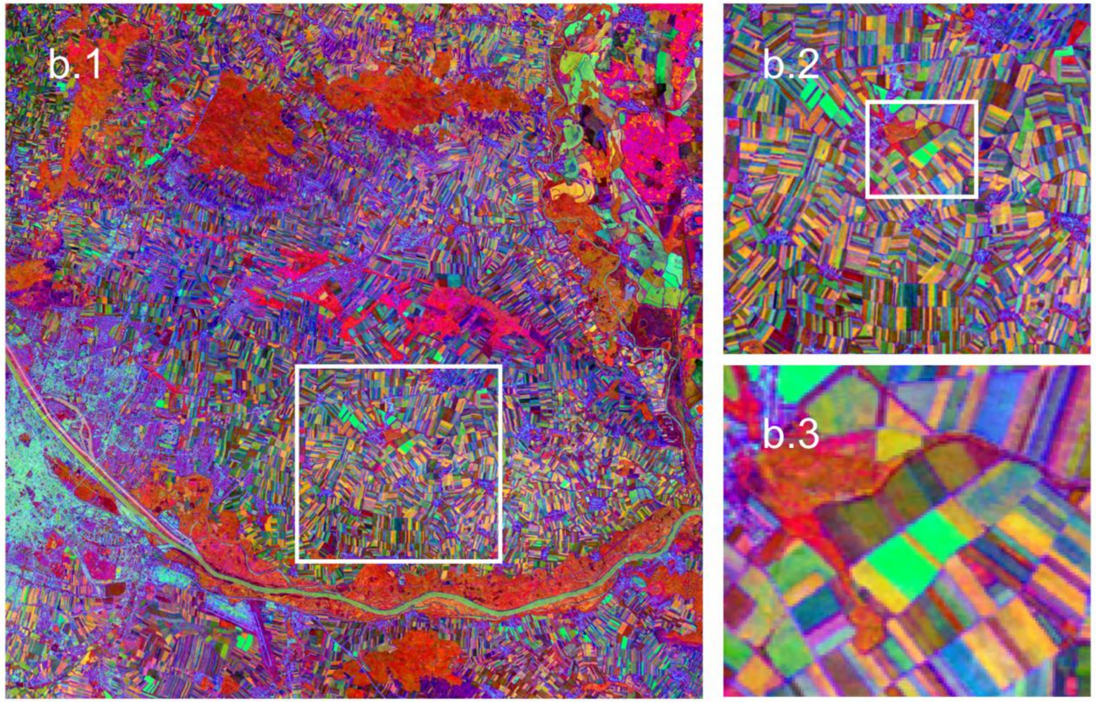
AlphaEarth Embeddings



DeepMind

Brown, C. F., et al., 2025. **AlphaEarth Foundations: An embedding field model for accurate and efficient global mapping from sparse label data**. *arXiv preprint arXiv:2507.22291*.

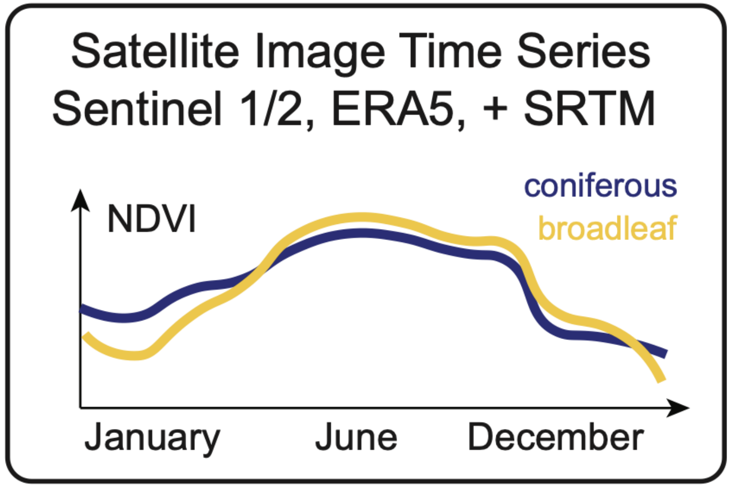
Tessera Embeddings



Feng, Zhengpeng, Clement Atzberger, Sadiq Jaffer, Jovana Knezevic, Silja Sormunen, Robin Young, Madeline C. Lisaius et al. "**TESSERA: Precomputed FAIR Global Pixel Embeddings for Earth Representation and Analysis**." arXiv preprint arXiv:2506.20380 (2025).

Example Application: Tree Species Identification

Monthly
Sentinel-1 & 2 + ERA5
time series

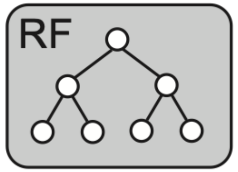


Embedding Approach

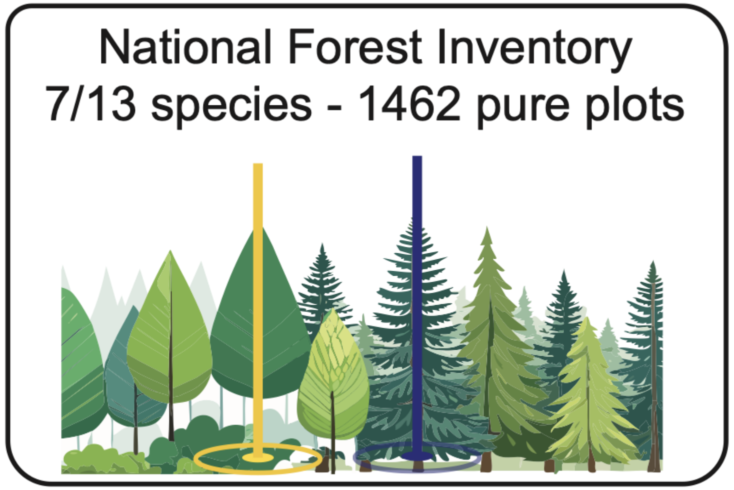
Classic Approach Francini et al., 2024

Harmonic Features

$$P_i = \beta_0 + \beta_1 t + \beta_2 \cos(\cdot) + \ldots$$



Tree Species Classification



Ishikawa, T., Bonannella, C., Lerink, B. J., & Rußwurm, M. (2025). Deep Pre-trained Time Series Features for Tree Species Classification in the Dutch Forest Inventory. arXiv preprint arXiv:2508.18829.

Takayuki Ishikawa

Results: Tree Species Identification

* These numbers are work-in-progress and may still change slightly

datasets	#Cl	Random Forest (Francini et al., 2024)	AlphaEarth (frozen)	TESSERA (frozen)	PRESTO (frozen)	PRESTO * (finetuned)
NFI (COMB)	13	0.63 ± 0.01	0.63 ± 0.01	**0.66 ± 0.01**	0.48 ± 0.01	**0.67** ± 0.02
Francini (SIBA)	7	0.84 ± 0.00	**0.90 ± 0.01**	0.76 ± 0.01	0.38 ± 0.01	**0.95** ± 0.01

Takeaway 1

Frozen embeddings beat hand-crafted harmonic features

Takeaway 2

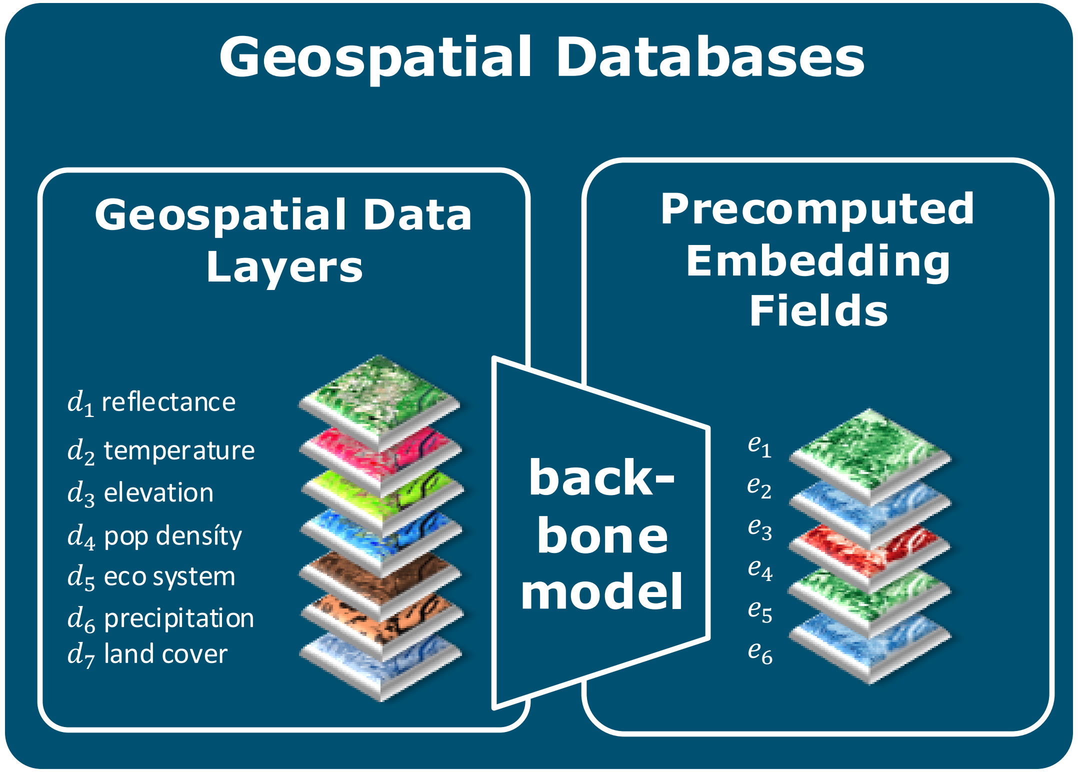
Fine-tuning deep models like Presto (Tseng et al., 2023) beat embeddings

Ishikawa, T., Bonannella, C., Lerink, B. J., & Rußwurm, M. (2025). **Deep Pre-trained Time Series Features for Tree Species Classification in the Dutch Forest Inventory.** arXiv preprint arXiv:2508.18829.

Takayuki Ishikawa

Summary: Precomputed Embedding Databases

Precomputed Embedding fields store **deep features** in a compressed form.



They are

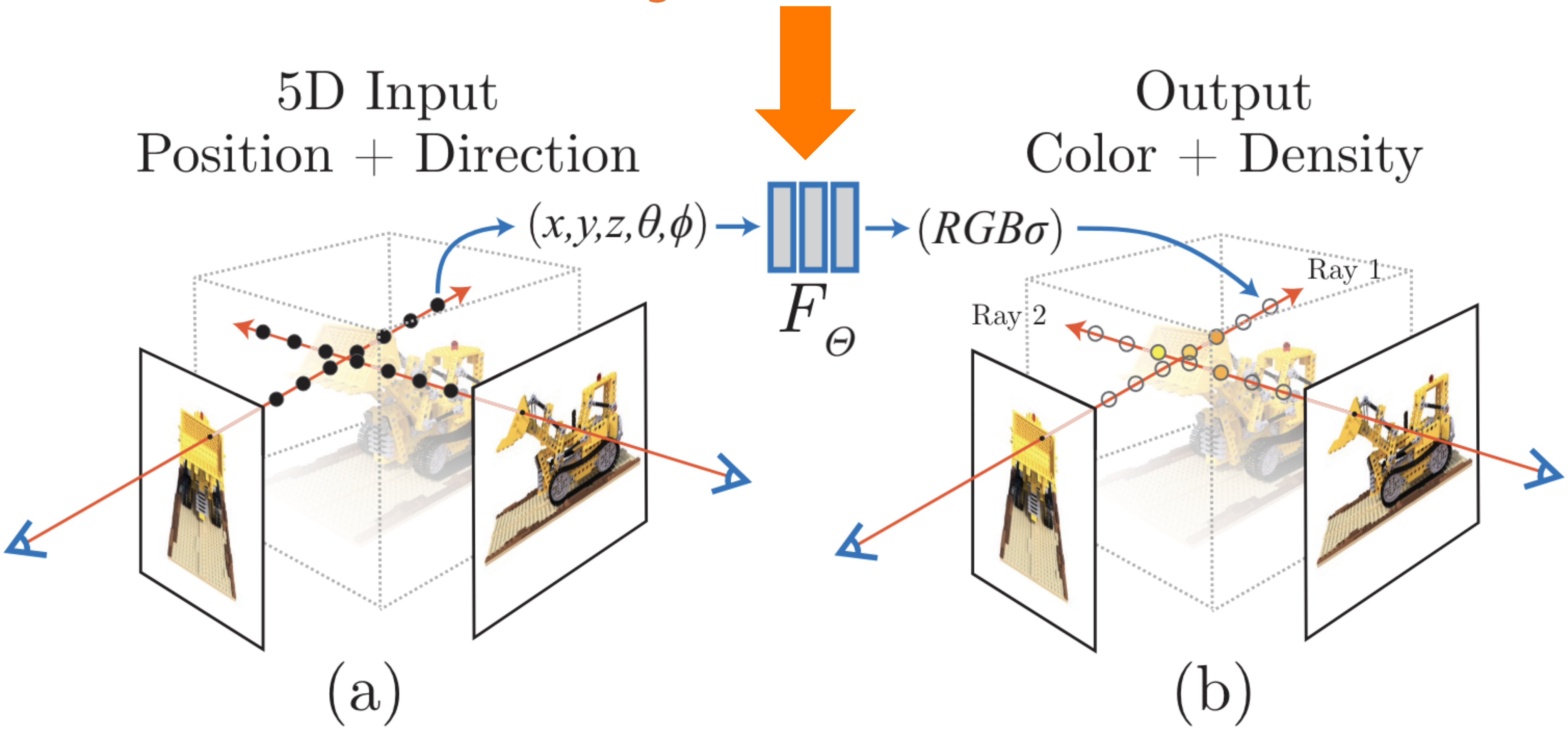
- an output of remote sensing foundation models (Alpha Earth, or Presto)
- very user-friendly: embedding "features" can be downloaded just like remote sensing data.
- But: they store **generic patterns** – **not** explicitly **what describes a location uniquely**

Part 2: Implicit neural Geo-Representations

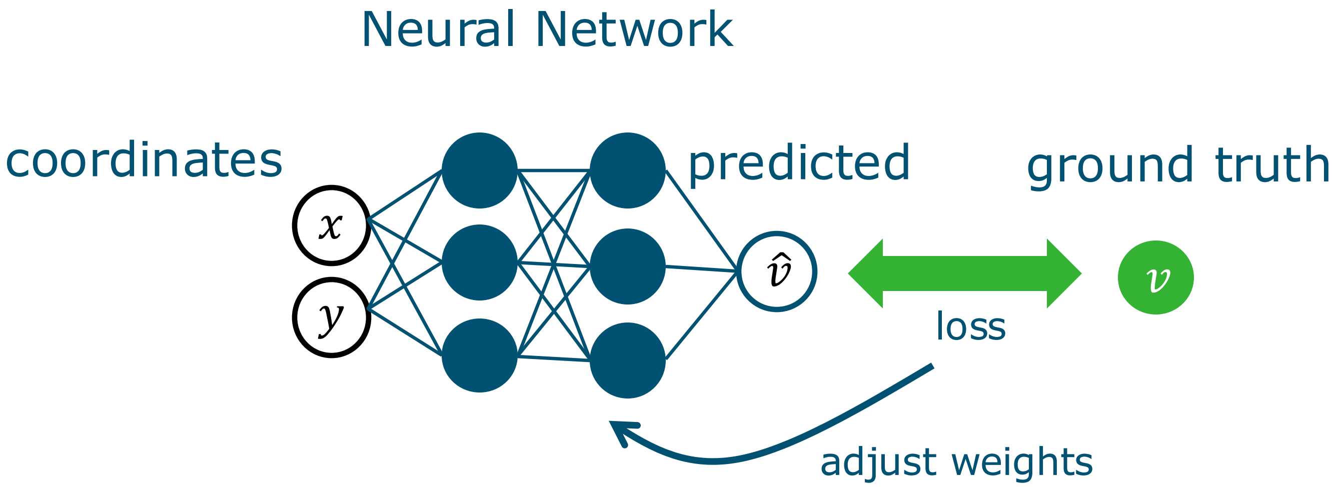
Implicit Neural Representations as active research field

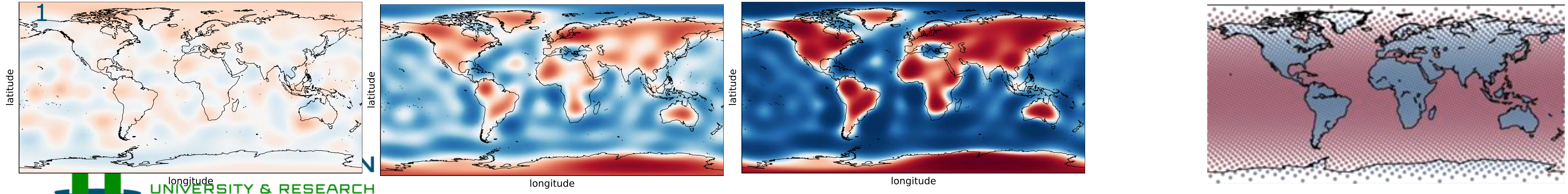
For instance, in Neural Radiance Fields (NeRF) Mildenhall et al., 2020

Neural Network weights encode information on the scene



Supervised Learning encodes Geodata in Neural Nets

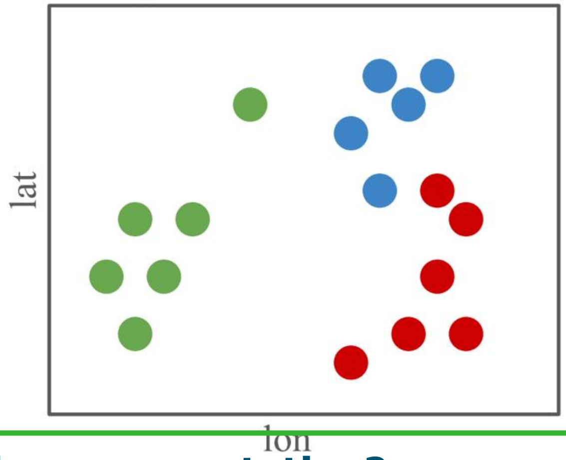




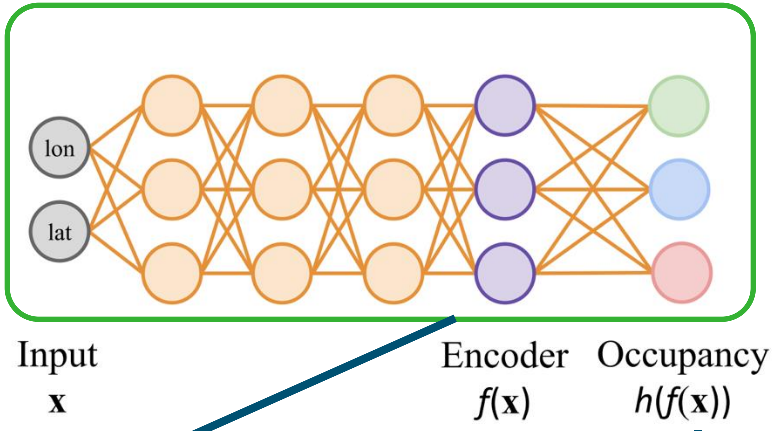
Implicit Neural Representations for Species Mapping

Input Data

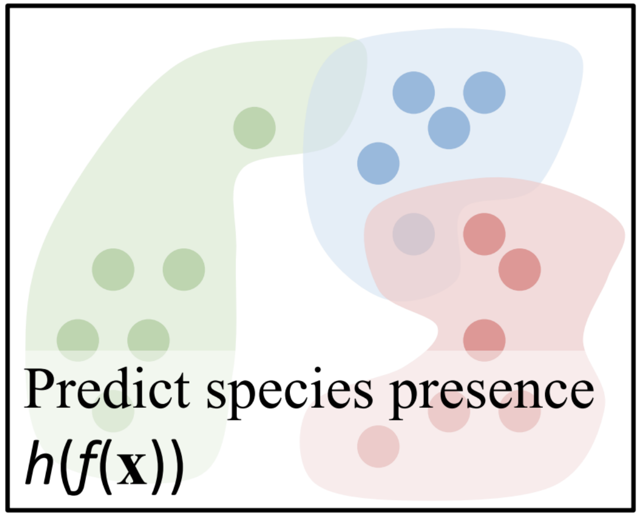
Species Presence



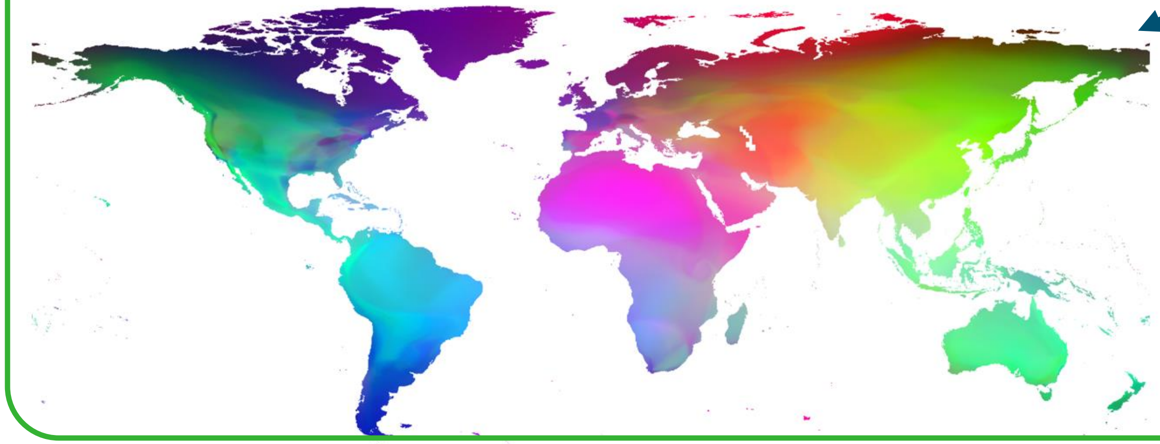
Species Range Model



Species Range

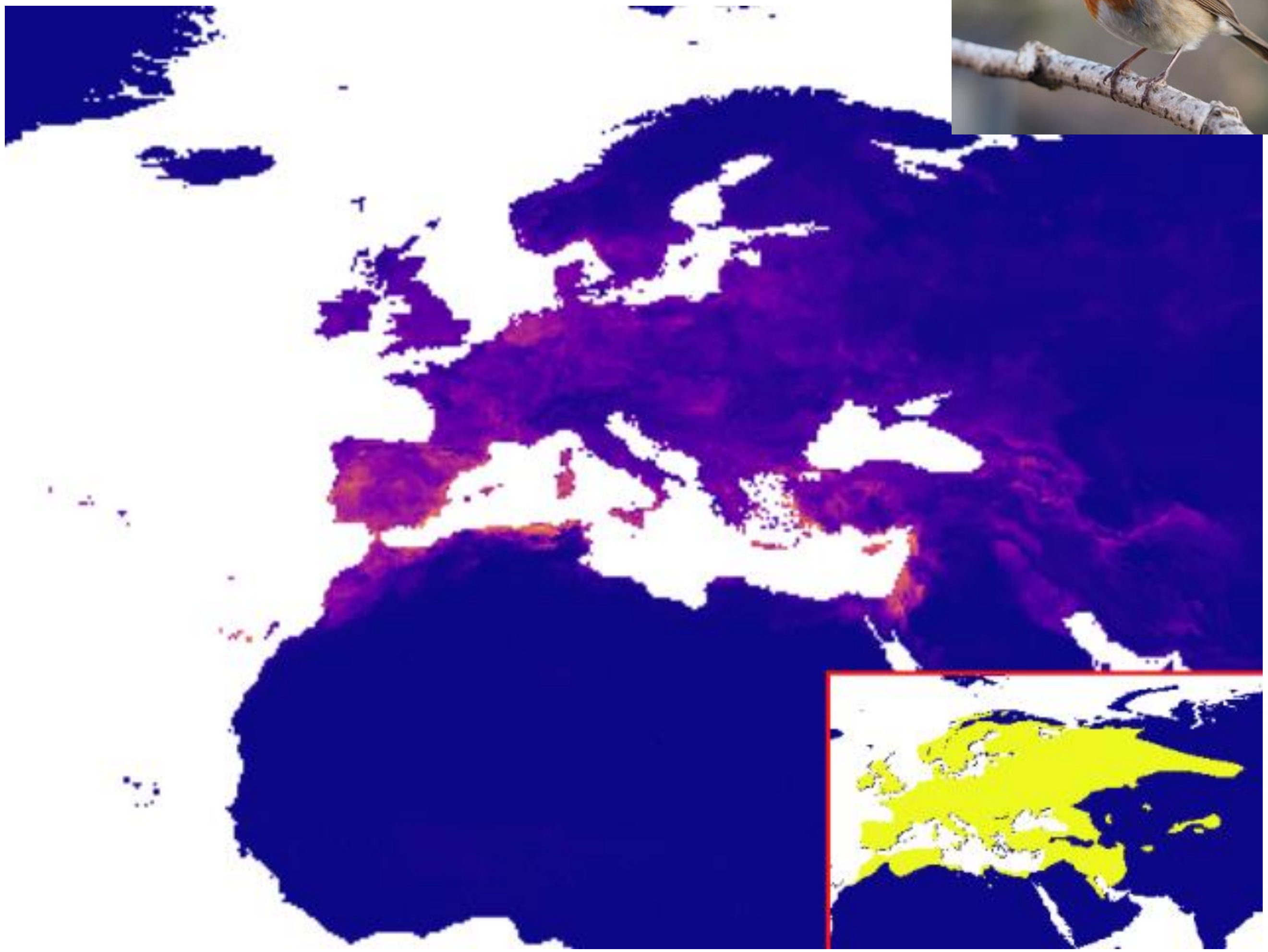


What can we do with the representation?



Cole et al., 2023. Spatial Implicit Neural Representations for Global-Scale Species Mapping. International Conference on Machine Learning (ICML)

FS-SINR: Few-Shot Species Range Mapping



Embeddings provide a geospatial context of “similarity”

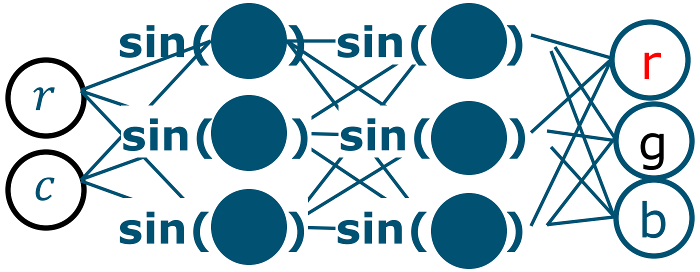
Fewer species observations are **needed** for a species range map.

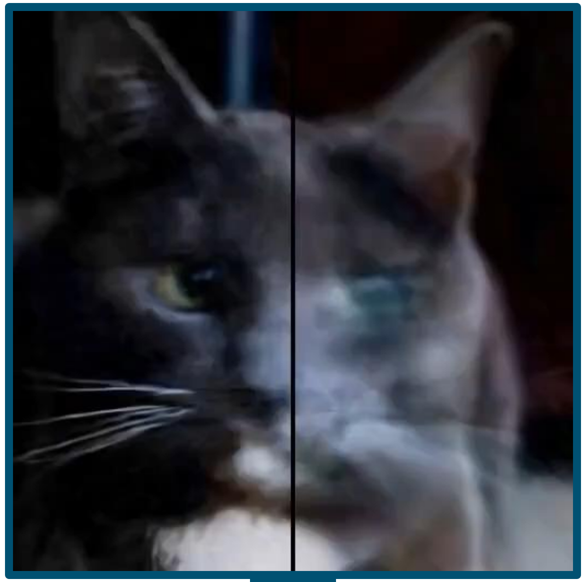
WAGENINGEN UNIVERSITY & RESEARCH

Lange, C., et al. 2025, O. Feedforward Few-shot Species Range Estimation. In *Forty-second International Conference on Machine Learning*.

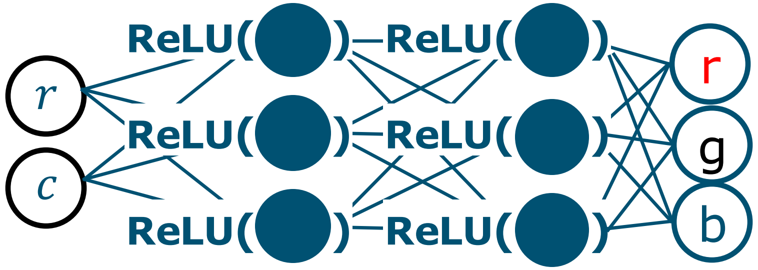
INR Lessions Learned 1: Sine activations work better

Sirens
a MLP with sine activations





ReLU MLP:
Multi-layer Perceptron (MLP)
with ReLU activations



Takeaway: Sine activation functions are better when storing data in neural network weights

Sitzmann, V., Martel, J., Bergman, A., Lindell, D., & Wetzstein, G. (2020). Implicit neural representations with periodic activation functions. Advances in neural information processing systems, 33, 7462-7473.

WAGENINGEN
UNIVERSITY & RESEARCH

Sine Activations Analogue to Fourier Coefficients

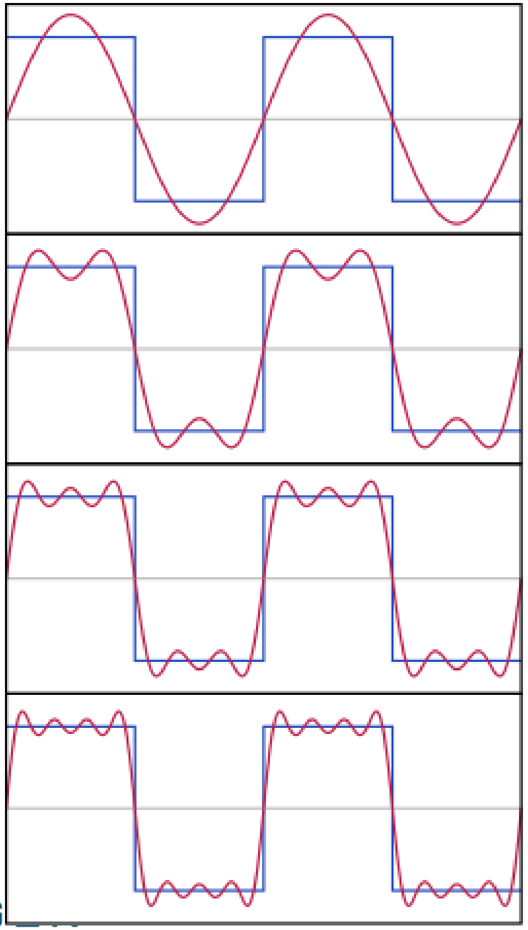
any signal can be approximated with sine and cosine basis functions

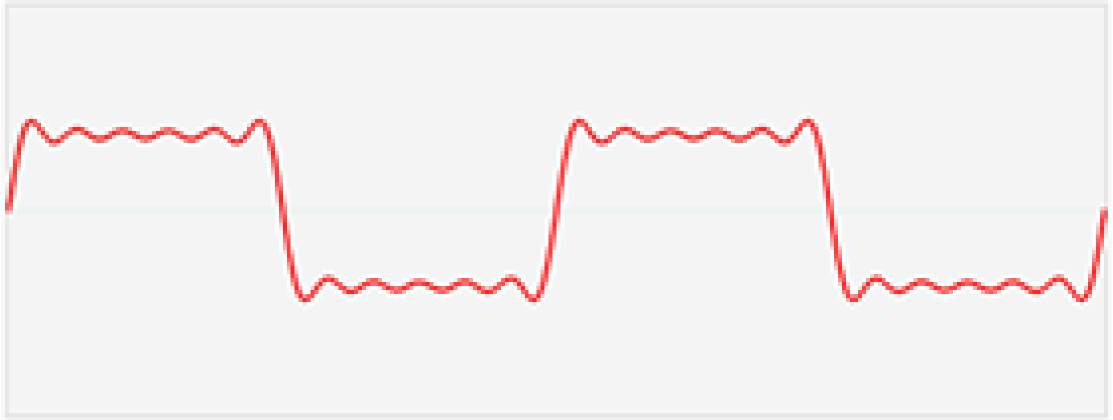
scale 1

scale 2

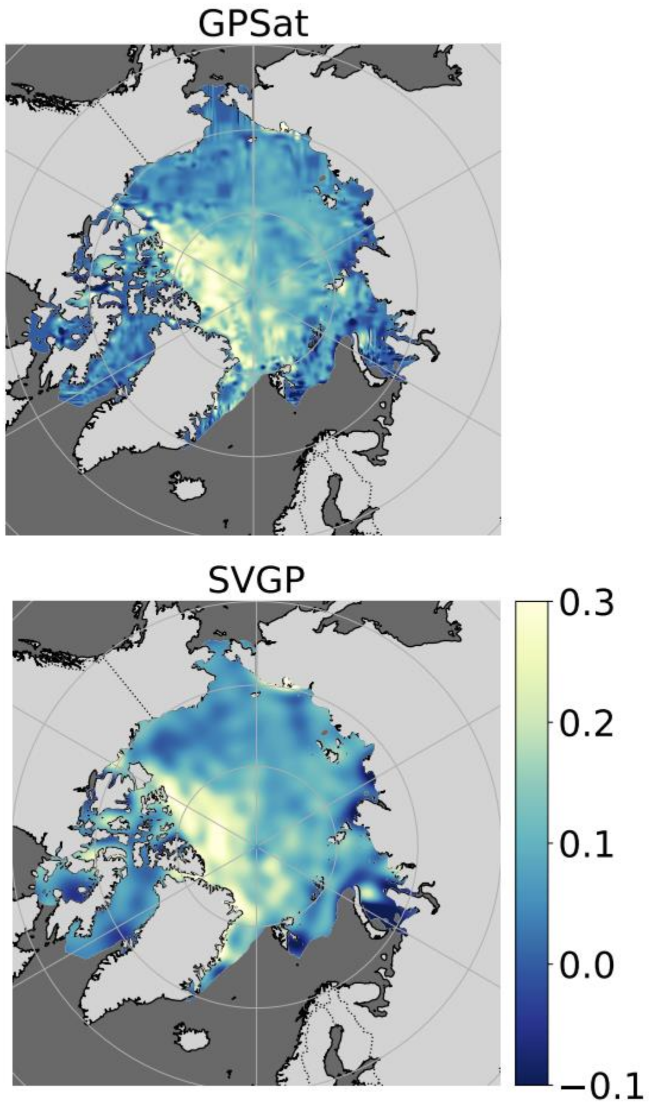
scale 3

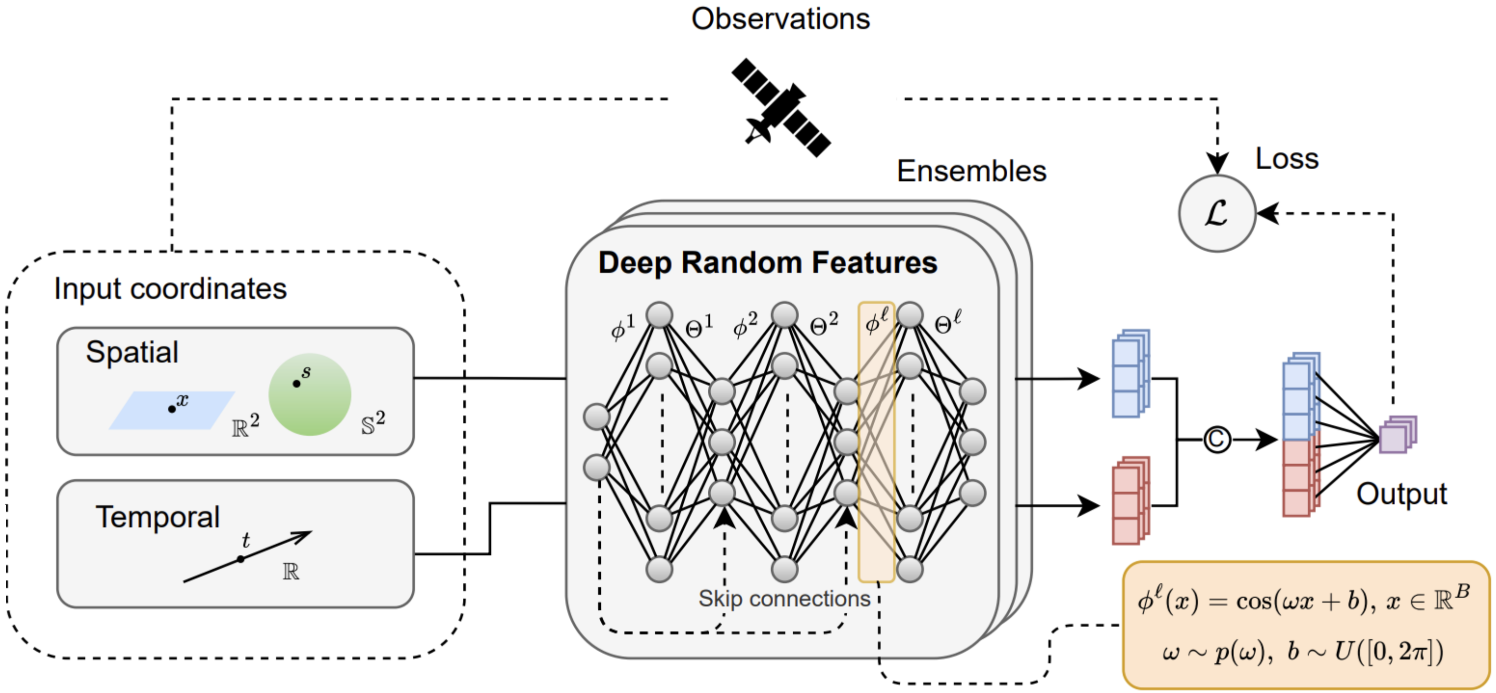
scale 4





Sea Ice Thickness through Deep Random Features





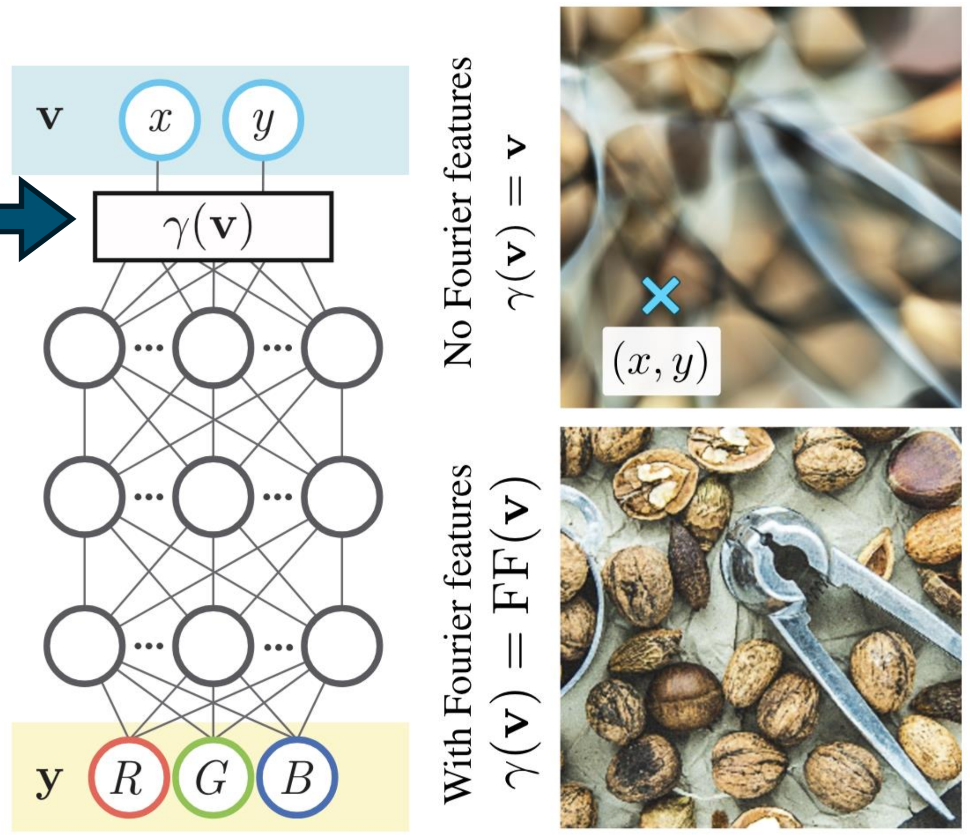
Chen, W., Mahmood, A., Tsamados, M., & Takao, S. (2025). **Deep random features for scalable interpolation of spatiotemporal data.** In *The Thirteenth International Conference on Learning Representations*.

INR Lessions Learned 2: Positional Encoding of Coords

Fourier Features (Tancik et al., 2020)

Positional encoding (no trainable weights)

$$\gamma(v) = (sin(2^0\pi v), cos(2^0\pi v), sin(2^1\pi v), cos(2^1\pi v), \\ ..., sin(2^{N-1}\pi v), cos(2^{N-1}\pi v)),$$



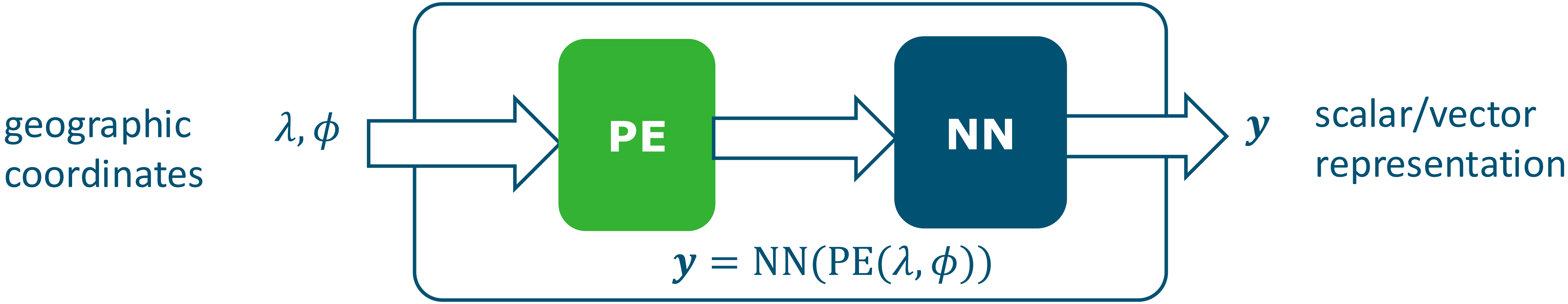
(a) Coordinate-based MLP

(b) Image regression $(x,y) \rightarrow$ RGB

Tancik, M., Srinivasan, P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., et al. & Ng, R. (2020). **Fourier features let networks learn high frequency functions in low dimensional domains.** *Advances in neural information processing systems*, *33*, 7537-7547.

Location Encoder: Positional Encoding & Neural Network

Location Encoder Network



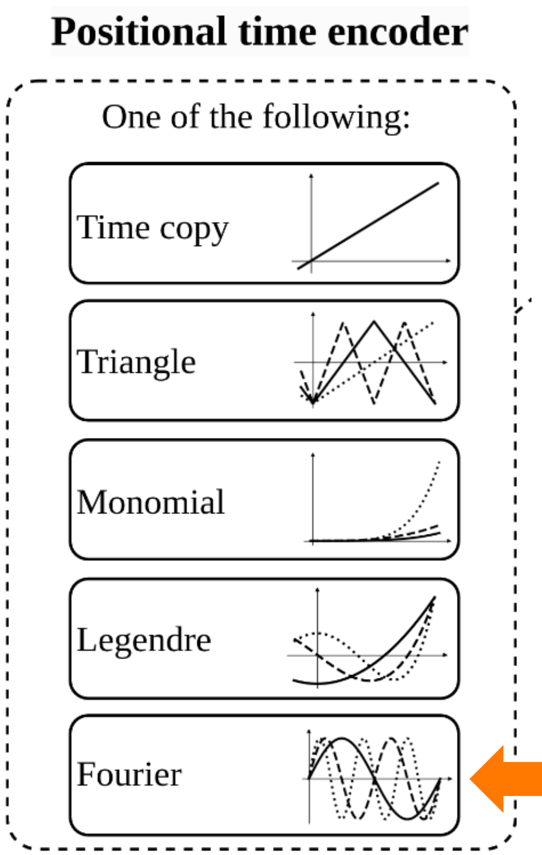
Positional Encoding PE
(non-parametric function)

Neural Network NN
(with trainable weights)

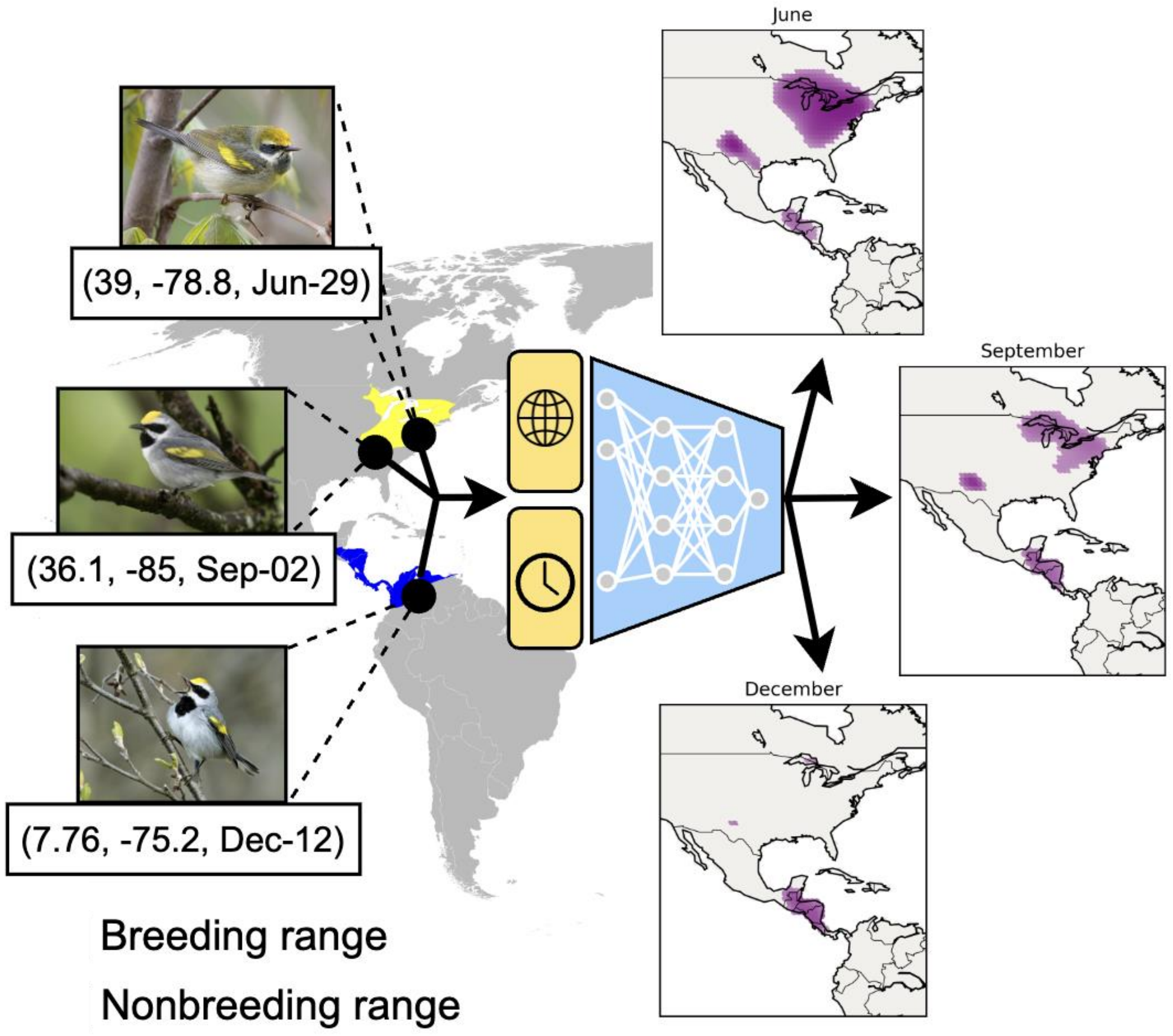
Review of location encoders:
Mai, G., Janowicz, K., Hu, Y., Gao, S., Yan, B., Zhu, R., & Lao, N. (2022). **A review of location encoding for GeoAI**: methods and applications. International Journal of Geographical Information Science, 36(4), 639-673.

WAGENINGEN UNIVERSITY & RESEARCH

Adding Time through Fourier Time Encoding



Fourier Sine Cosine encoding makes Sense



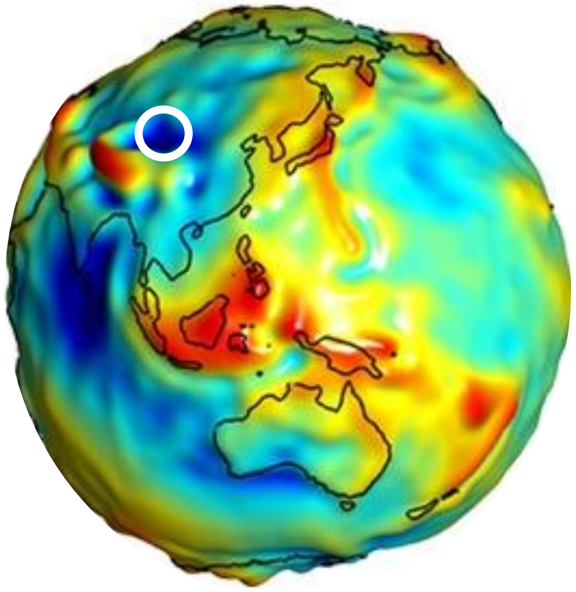
Mickisch, D., Klemmer, K., Rußwurm, M., Rolf, E. , Teng, M., & Rolnick, D. **A Joint Space-Time Encoder for Geographic Time-Series Data.**
Previous version In *ICLR 2025 Workshop on Machine Learning Multiscale Processes*.

Spherical Harmonics as Positional Encoding Function

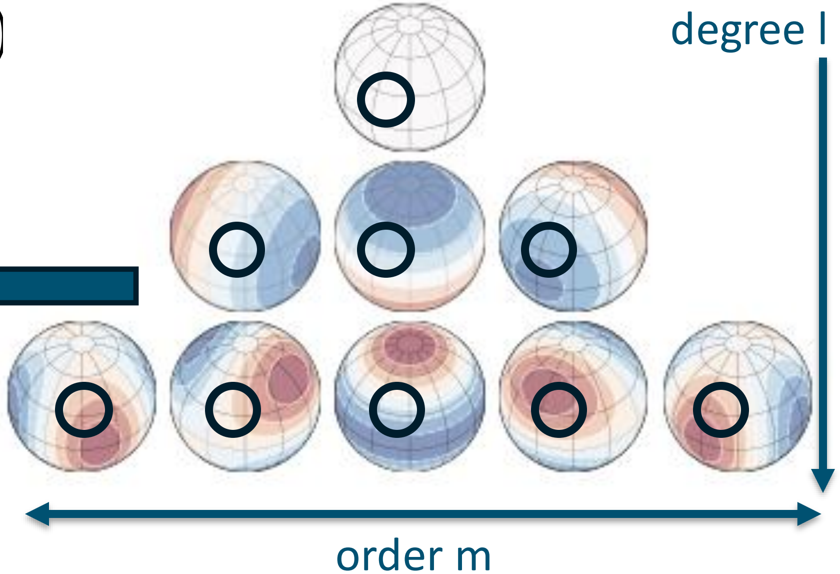
0.5

-0.2 0.4 0.1

0.9 -0.1 -0.2 0.5 0.4

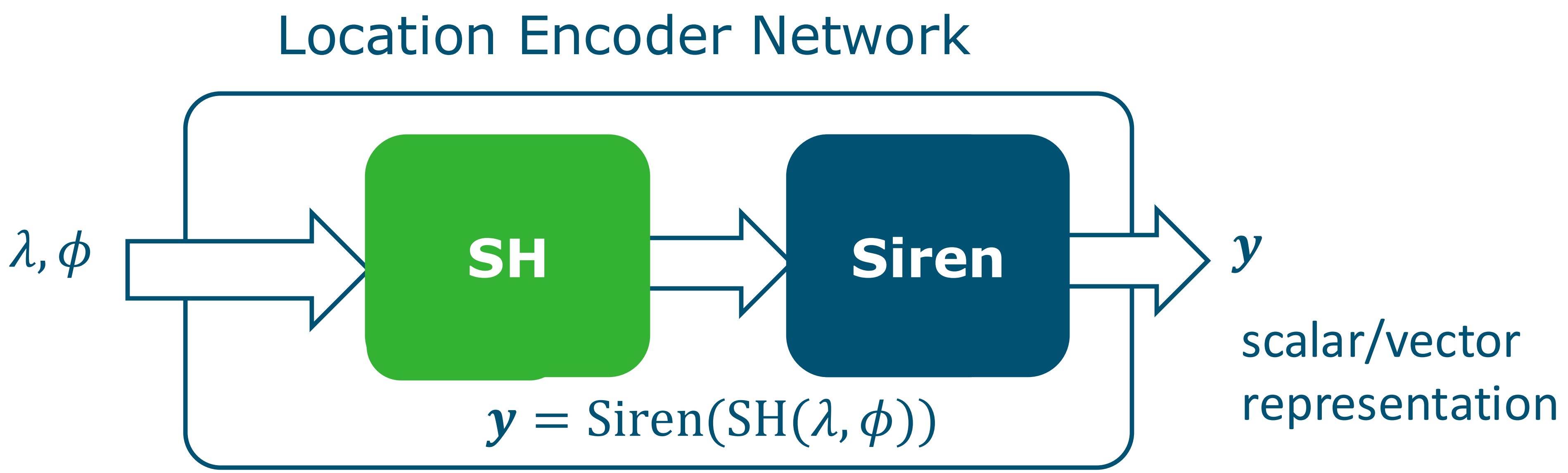


$$f(\lambda, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} w_l^m Y_l^m(\lambda, \phi)$$



Gerstl, M. (2008). Computing the Earth gravity field with spherical harmonics. In From Nano to Space: Applied Mathematics Inspired by Roland Bulirsch (pp. 277-294). Berlin, Heidelberg: Springer Berlin Heidelberg.

Our Proposed Location Encoder: Siren(SH(λ, ϕ))

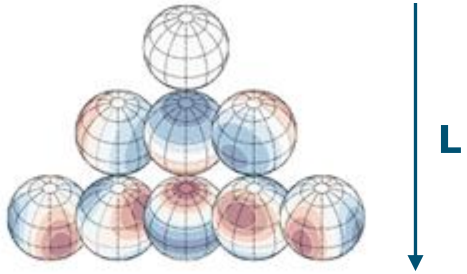


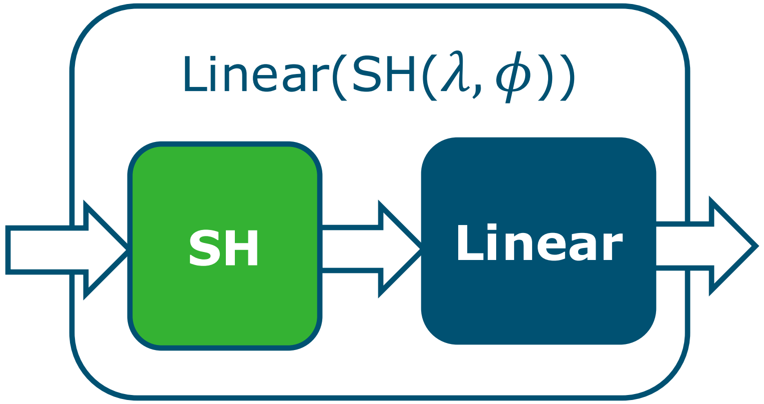
Spherical Harmonics basis functions (non-parametric function)

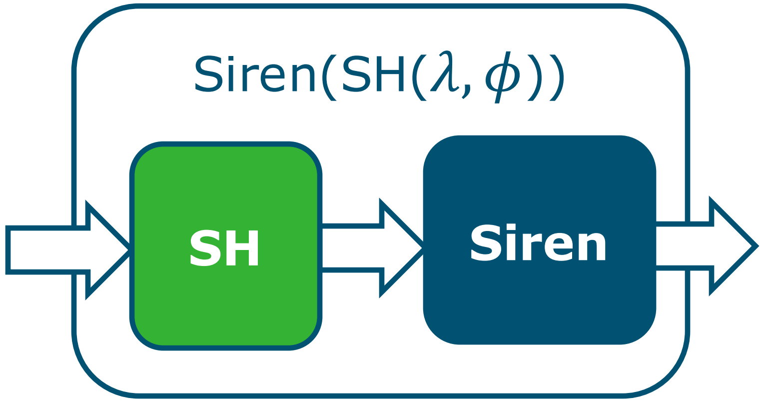
Sinusoidal Representation Networks (with trainable weights)

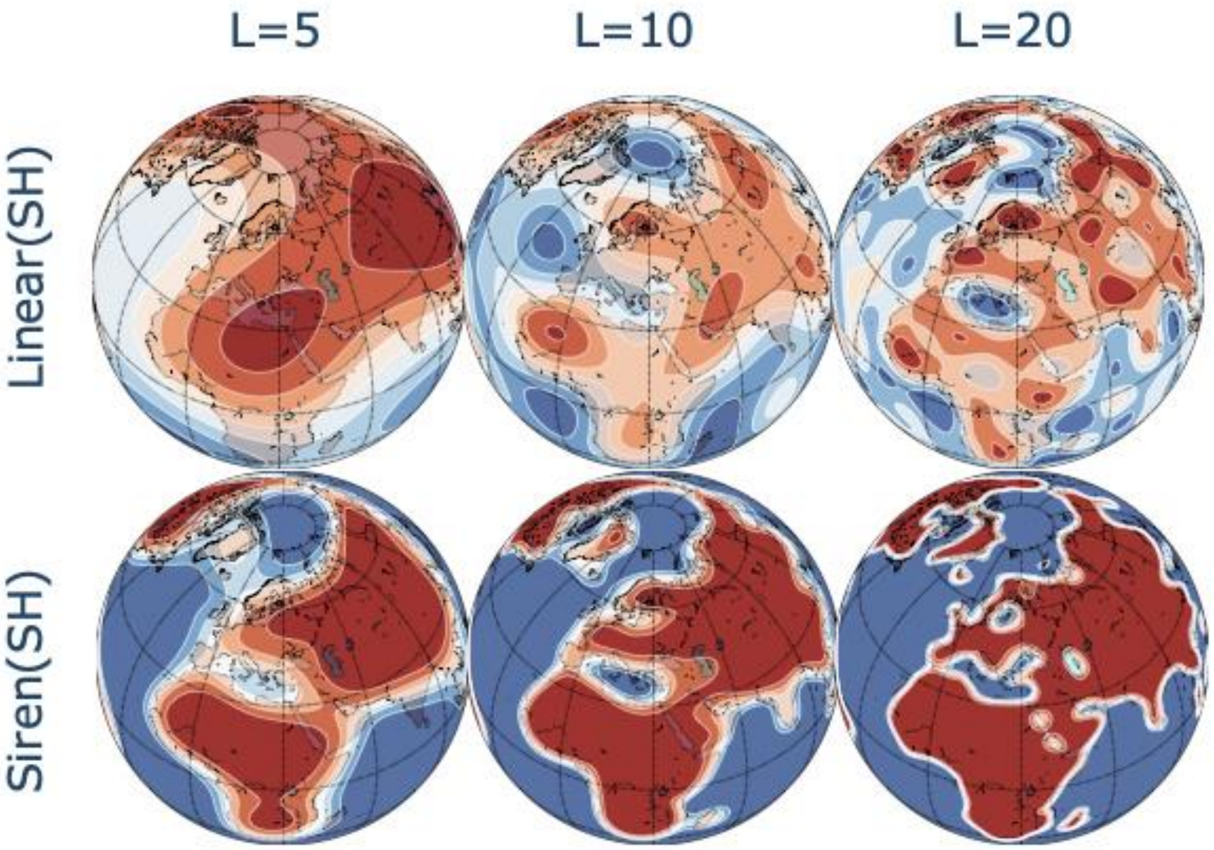
Rußwurm, M., Klemmer, K., Rolf, E., Zbinden, R., & Tuia, D. (2024). **Geographic location encoding with spherical harmonics and sinusoidal representation networks**. International Conference for Learning Representations (ICLR). Spotlight paper (top 5% of submissions)

SirenSH Controlling Smoothness via L







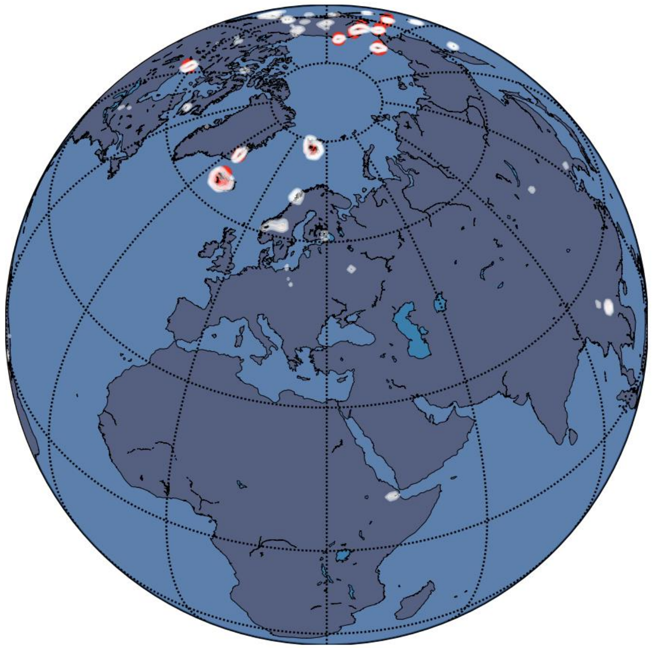


WAGENINGEN UNIVERSITY & RESEARCH

Rußwurm, M., Klemmer, K., Rolf, E., Zbinden, R., & Tuia, D. (2024). Geographic location encoding with spherical harmonics and sinusoidal representation networks. International Conference for Learning Representations.

Results: Orthogonal Representations (SH & Siren) matter

iNaturalist



iNaturalist 2018 accuracy improvement

PE ↓ NN →	LINEAR	FCNET	SIRENNET
DIRECT	-5.9 ± 0.1	$+9.3 \pm 0.3$	$+12.1 \pm 0.1$
CARTESIAN3D	$+0.8 \pm 0.2$	$+11.8 \pm 0.1$	$+12.0 \pm 0.1$
WRAP	-0.1 ± 0.1	$\underline{+12.1 \pm 0.1}$	$+12.1 \pm 0.1$
GRID	$+11.2 \pm 0.1$	$+11.8 \pm 0.2$	$+11.6 \pm 0.4$
THEORY	$+11.5 \pm 0.0$	$+10.8 \pm 0.0$	$+11.4 \pm 0.1$
SPHEREC	$+11.2 \pm 0.1$	$+12.0 \pm 0.2$	$\mathbf{+12.3 \pm 0.1}$
SPHEREC+	$+11.1 \pm 0.2$	$+11.5 \pm 0.3$	$+10.3 \pm 0.4$
SPHEREM	$+7.2 \pm 0.2$	$+11.3 \pm 0.2$	$+10.6 \pm 0.6$
SPHEREM+	$+11.6 \pm 0.1$	$+12.0 \pm 0.1$	$+10.7 \pm 0.2$
SH (ours)	$+10.5 \pm 0.1$	$+12.0 \pm 0.0$	$\mathbf{+12.3 \pm 0.2}$

image-only: 59.2% top-1 accuracy with encoder NN(PE) ↑

Spherical Harmonics work well with all NNs

Siren is work well with all PEs

WAGENINGEN UNIVERSITY & RESEARCH

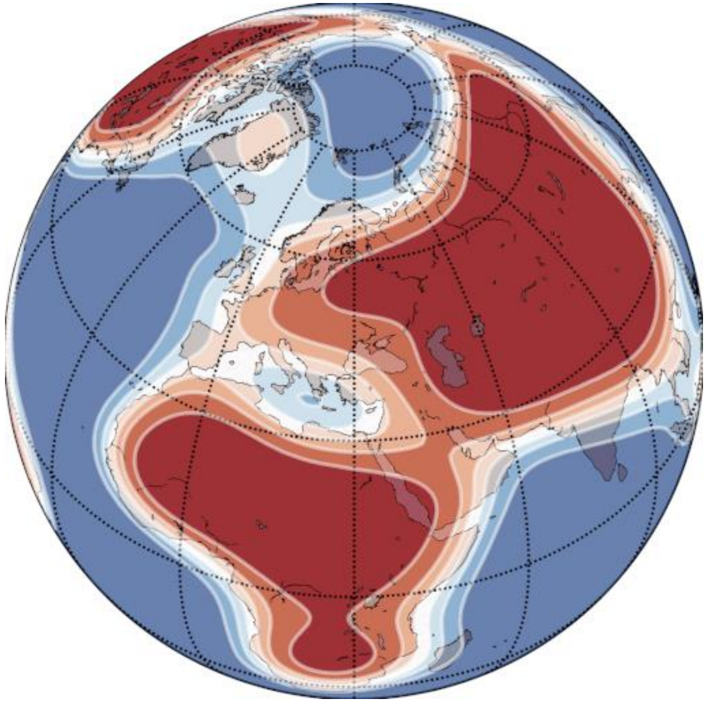
Rußwurm, M., Klemmer, K., Rolf, E., Zbinden, R., & Tuia, D. (2024). Geographic location encoding with spherical harmonics and sinusoidal representation networks. International Conference for Learning Representations.

Location Encoders - Recommendations

To integrate coordinates in a deep neural network:

we can recommend:

1. Siren as Neural Network for **any location encoding problem** and
2. Spherical Harmonic basis functions for **global geographic problems** where the spherical geometry matters



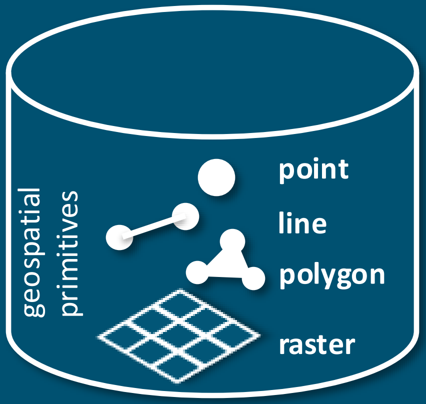
Code: github.com/marccoru/locationencoder
Contact: marc.russwurm@wur.nl

WAGENINGEN UNIVERSITY & RESEARCH

Rußwurm, M., Klemmer, K., Rolf, E., Zbinden, R., & Tuia, D. (2024). **Geographic location encoding with spherical harmonics and sinusoidal representation networks.** International Conference for Learning Representations.

Larger Takeaway: Location Encoders store Geodata

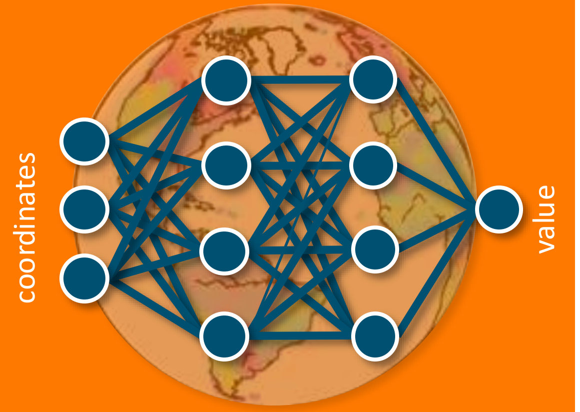
Explicit Geospatial Database



discrete geodata stored in **geospatial database**

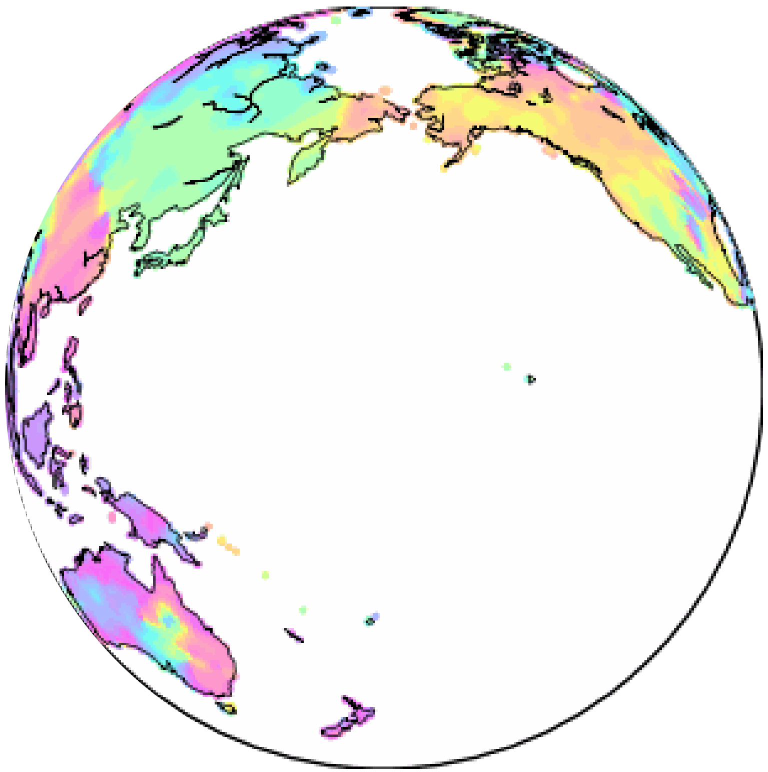
continuous geospatial fields stored in **neural network weights**

Implicit Neural Geo Representation



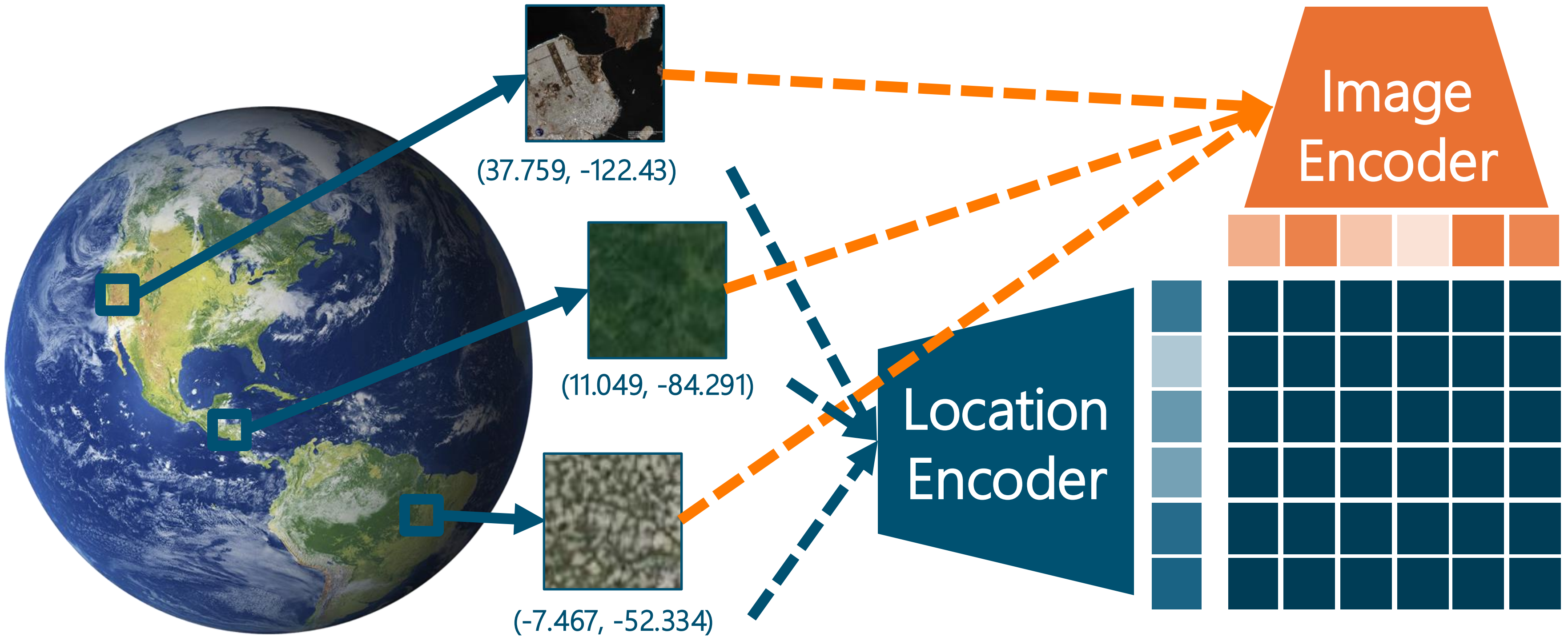
ts

Part 3: Implicit Embedding Models through Geolocalization



WAGENINGEN
UNIVERSITY & RESEARCH

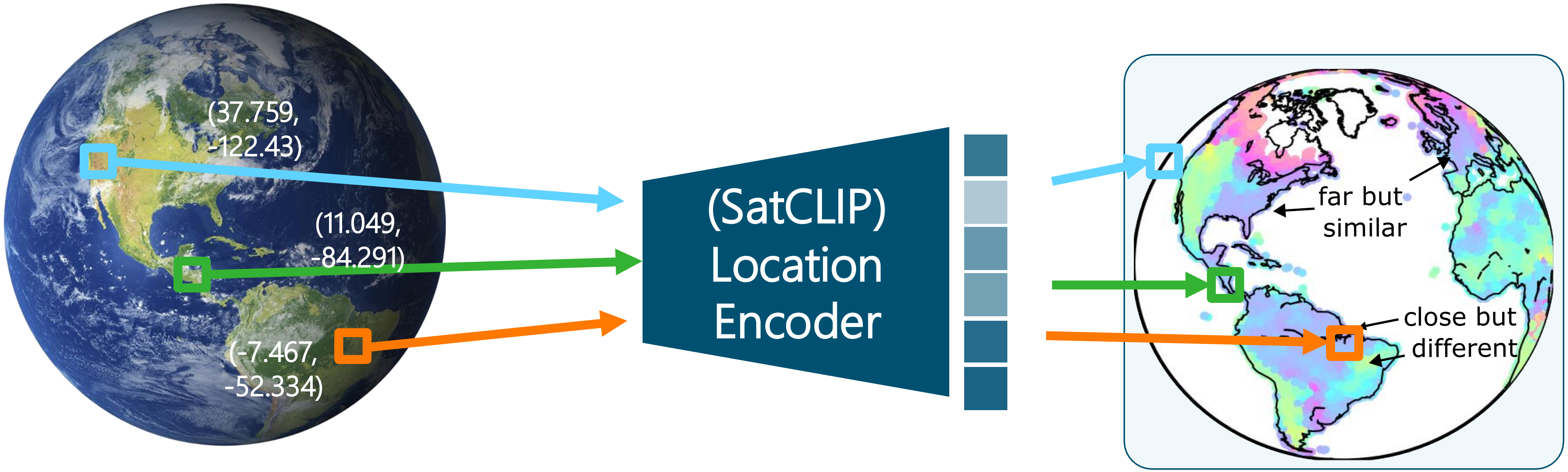
Geolocalization via Satellite Contrastive Location-Image Pretraining (SatCLIP)



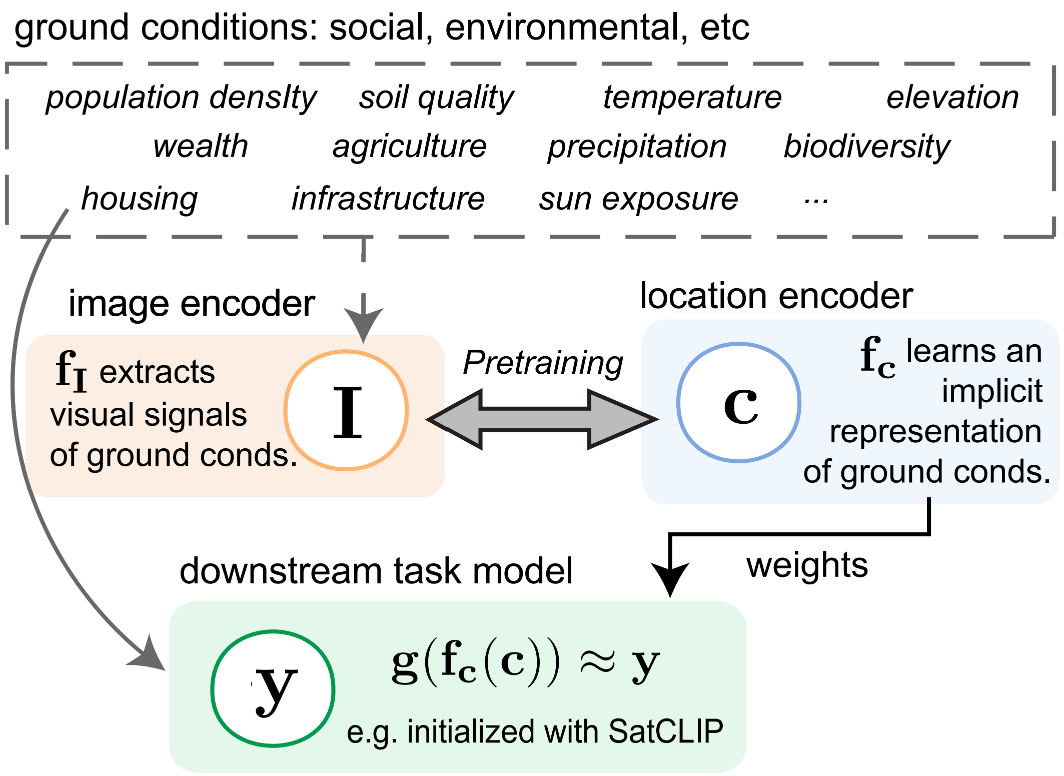
Contrastive learning: match image and location embeddings.

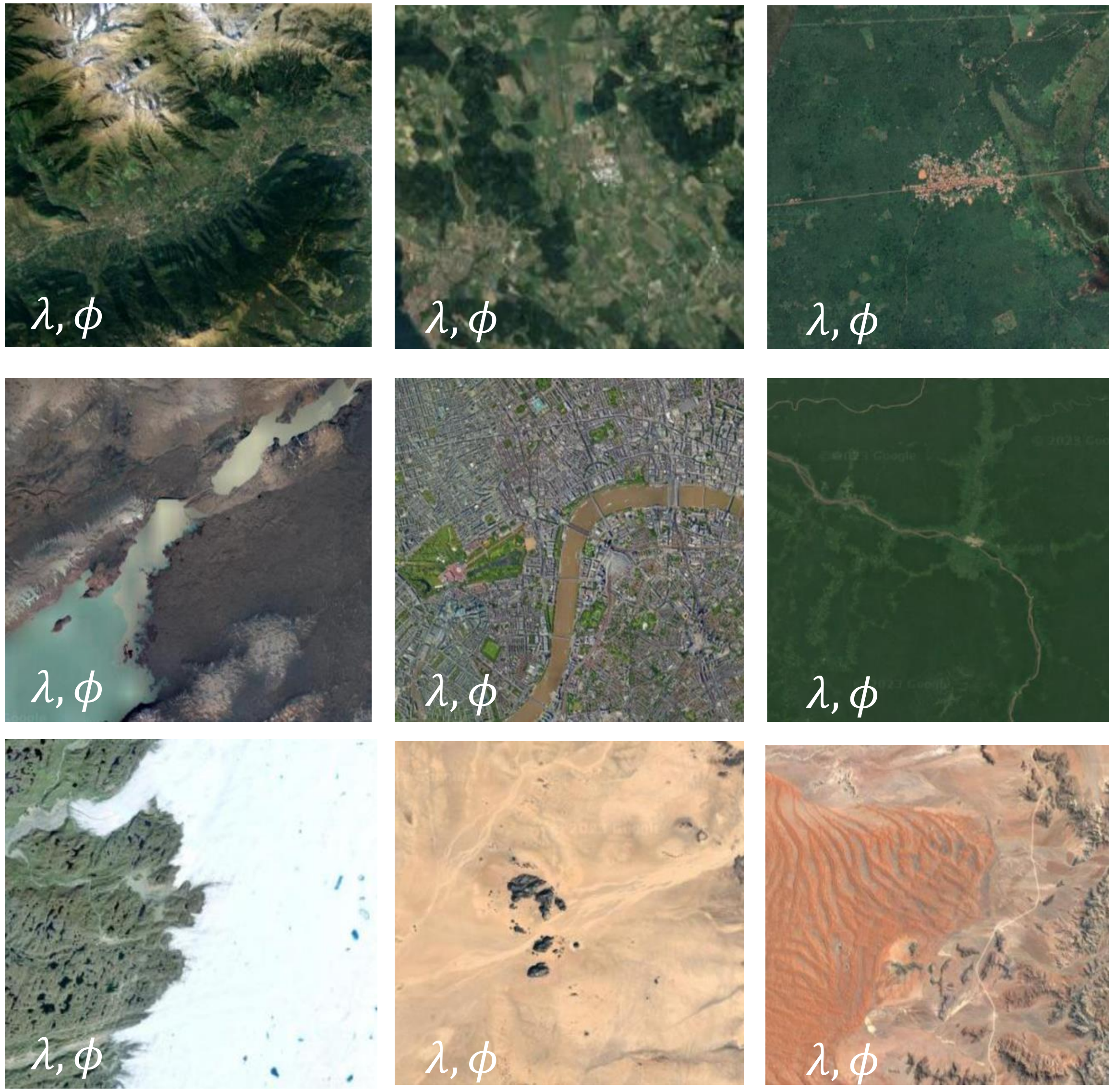
Klemmer, K., Rolf, E., Robinson, C., Mackey, L., & Rußwurm, M. (2025). **Satclip: Global, general-purpose location embeddings with satellite imagery.** AAAI 2025

A simple pre-trained MLP stores an Embedding Field



Intuition behind SatCLIP: distill location-specific patterns



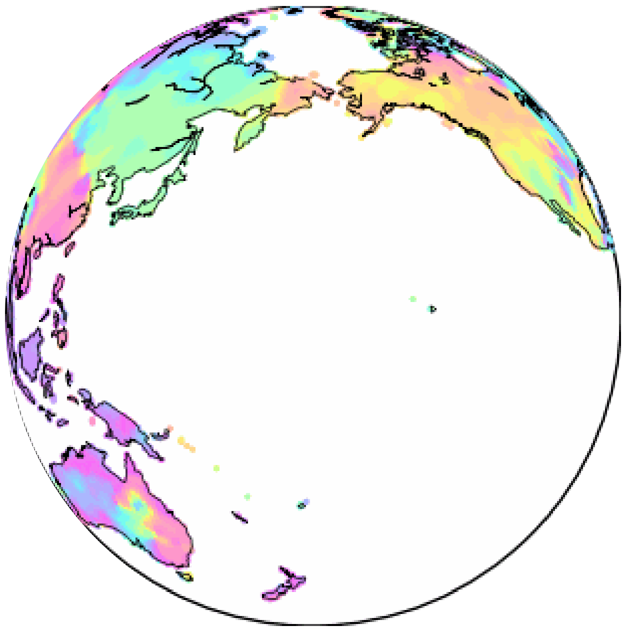


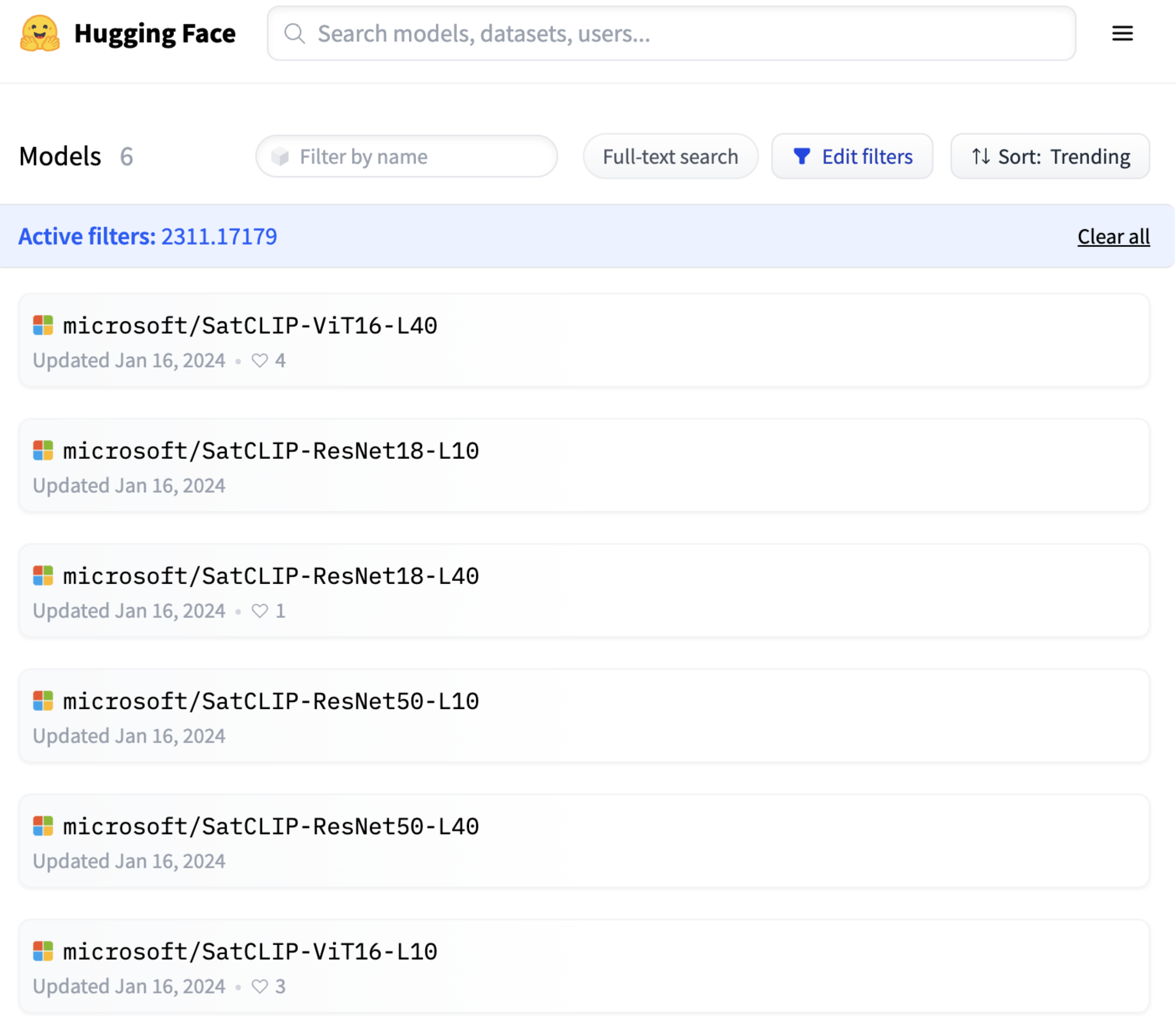
WAGENINGEN UNIVERSITY & RESEARCH

Imagery ©2023 Airbus, CNES / Airbus, Landsat / Copernicus, Maxar Technologies

SatCLIP availability

SatCLIP is a small SirenSH with weights are publicly available on HuggingFace

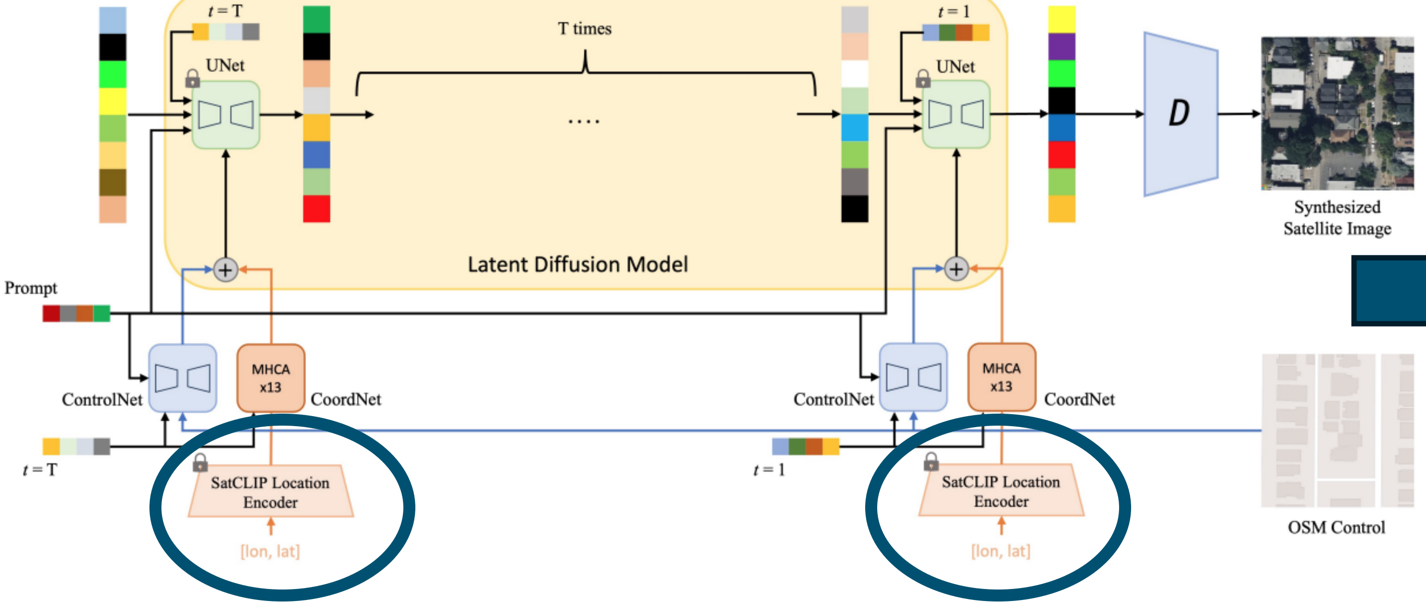




WAGENINGEN
UNIVERSITY & RESEARCH

Conditional Image Generation

SatCLIP is used to guide diffusion models:



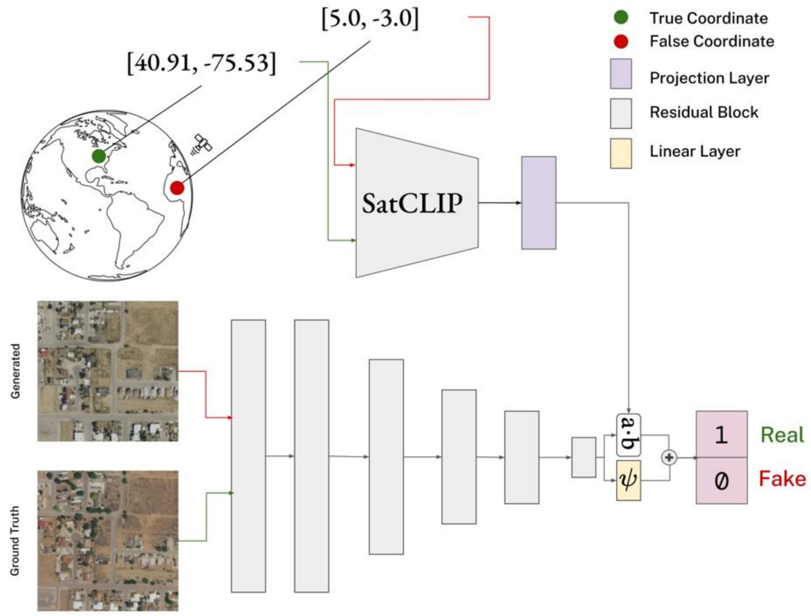
Sastry, Srikumar et al. (2024) GeoSynth: Contextually-aware high-resolution satellite image synthesis. *EarthVision, CVPR.*

Usages of SatCLIP – Location-guided super resolution

Can Location Embeddings Enhance Super-Resolution of Satellite Imagery?

Daniel Panangian* Ksenia Bittner
The Remote Sensing Technology Institute
German Aerospace Center (DLR), Wessling, Germany
{daniel.panangian, ksenia.bittner}@dlr.de

(a) Sentinel-2 (b) Super-resolved (c) Building Segmentation

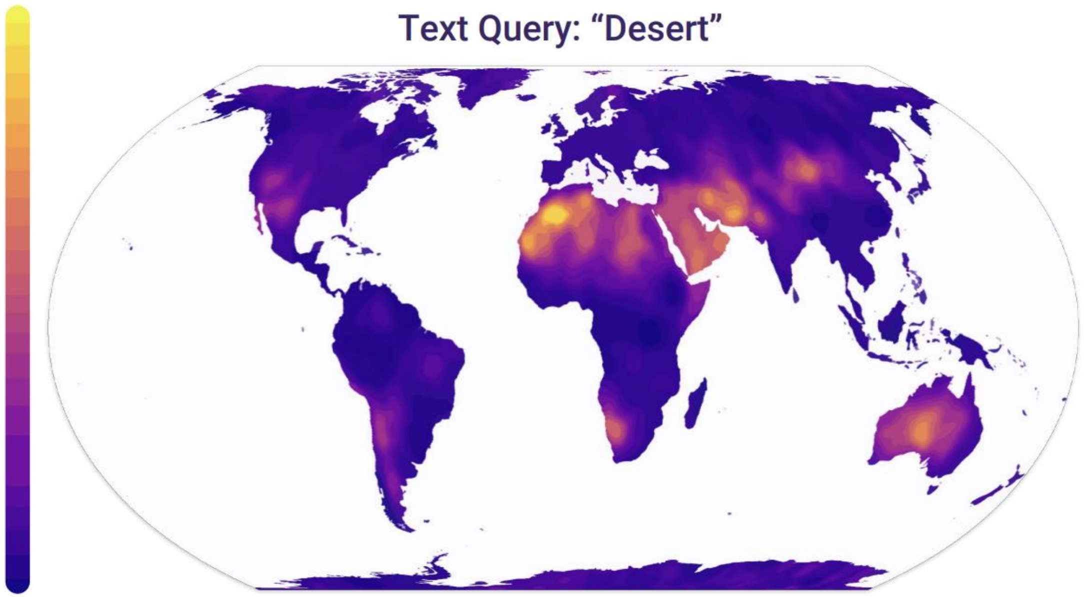


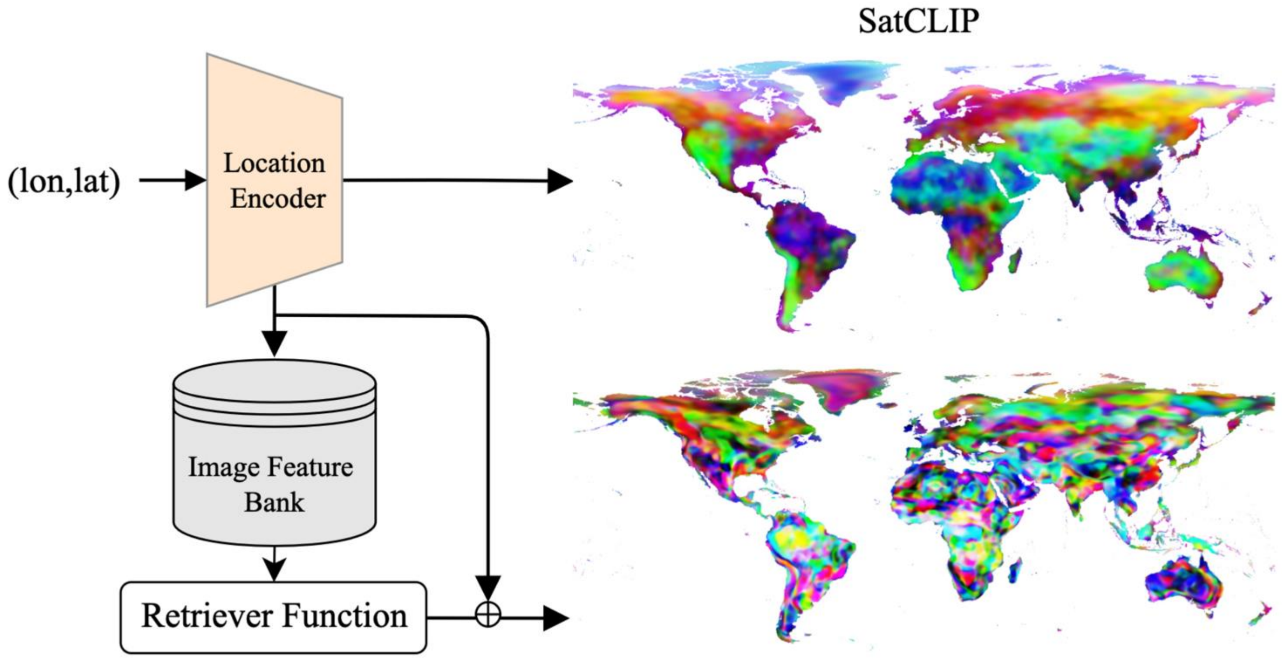
Panangian, D., & Bittner, K. (2025). Can Location Embeddings Enhance Super-Resolution of Satellite Imagery?. arXiv preprint arXiv:2501.15847.

Similar works: GeoCLIP & RANGE

GeoCLIP: CLIP-inspired alignment of ground images for worldwide geolocalization

RANGE: Retrieval Augmented Neural Fields for Multi-Resolution Geo-Embeddings

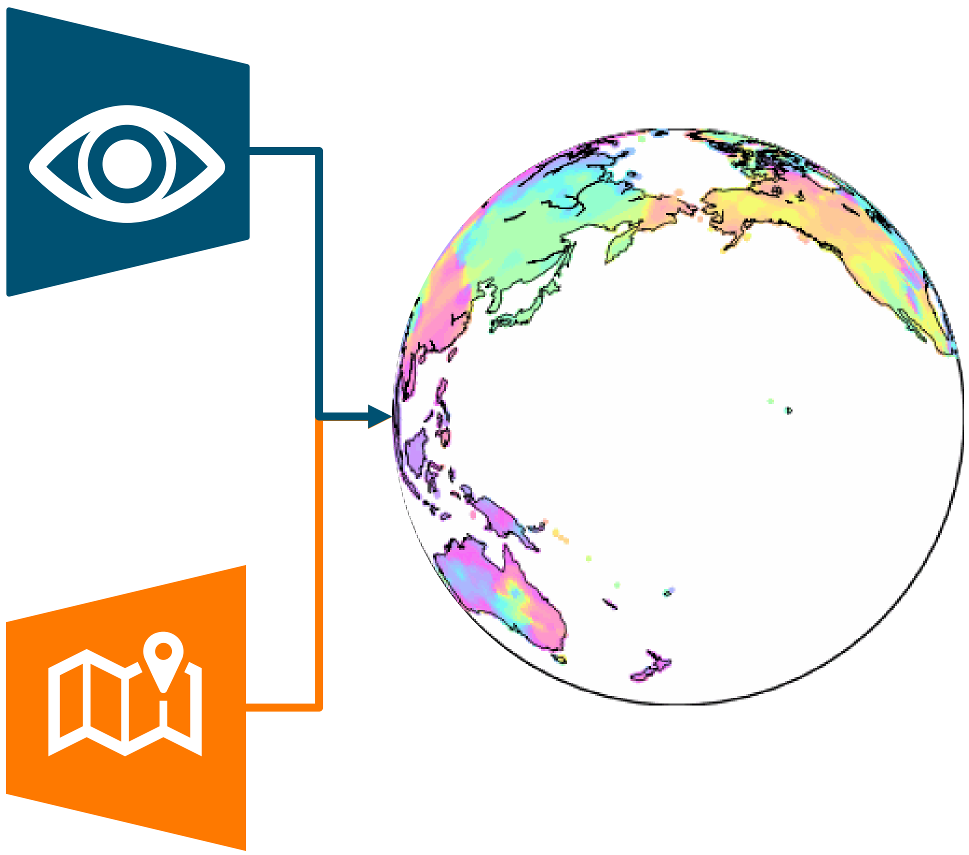




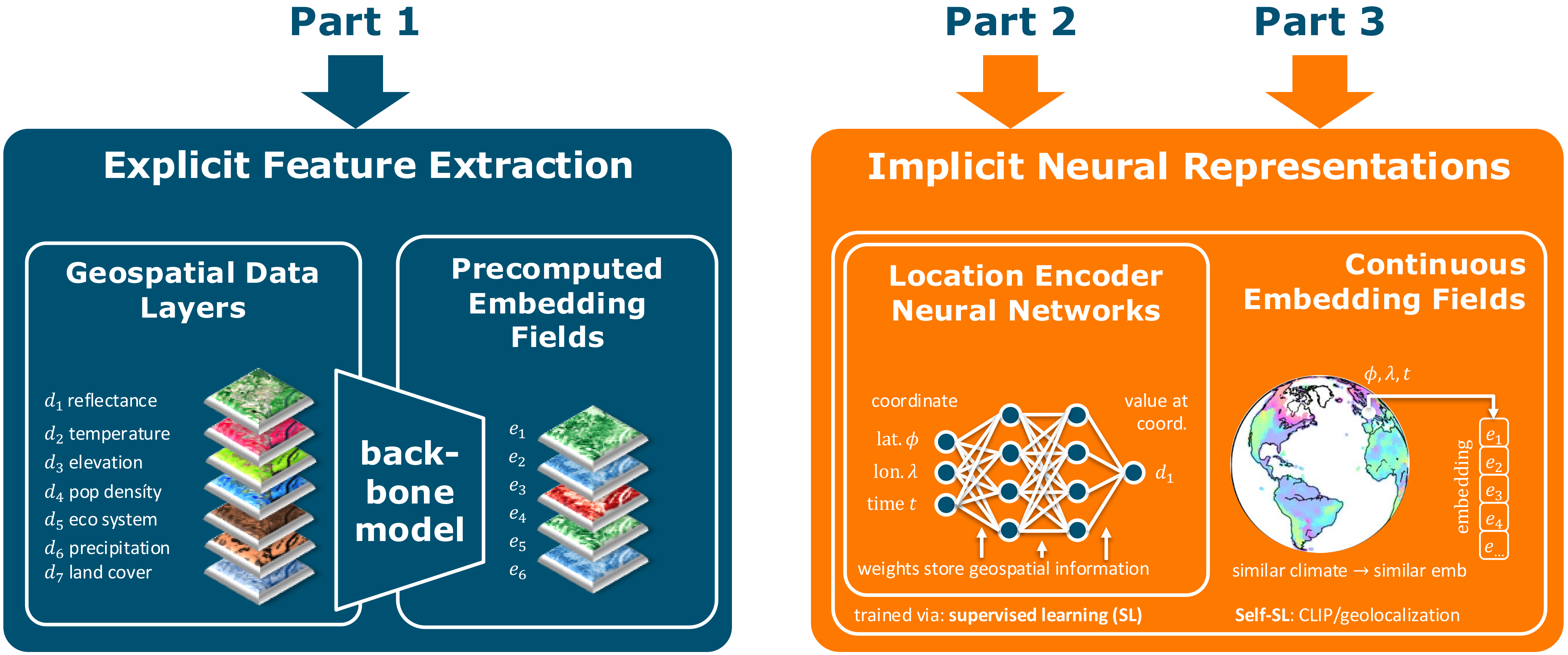
Vivanco Cepeda, Vicente, Gaurav Kumar Nayak, and Mubarak Shah. "Geoclip: Clip-inspired alignment between locations and images for effective worldwide geo-localization." Advances in Neural Information Processing Systems 36 (2023): 8690-8701.

Dhakal, A., Sastry, S., Khanal, S., Ahmad, A., Xing, E., & Jacobs, N. (2025). RANGE: Retrieval Augmented Neural Fields for Multi-Resolution Geo-Embeddings. In Proceedings of the Computer Vision and Pattern Recognition Conference (pp. 24680-24689).

Conclusion: Storing Mental Maps in Neural Nets



Explicit Feat. Extraction/Implicit Neural Representations



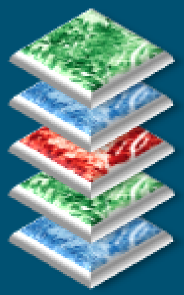
Embedding Databases versus Embedding Models

Explicit Feature Extraction

typically available as

Precomputed Embedding Databases

- created by "foundation" models with (cross-modal) auto-encoding
- available as compressed rasters of "deep features"
- drop-in replacement for raw geodata & build on existing geodata infrastructure (GEE)
- are generally high-resolution



Implicit Neural Representations

typically available as

Continuous Embedding Models

- created by small neural networks through self-supervised geolocalization
- available as pre-trained models (on HuggingFace)
- drop-in model to encode coordinates in neural net architectures
- are generally lower-resolution

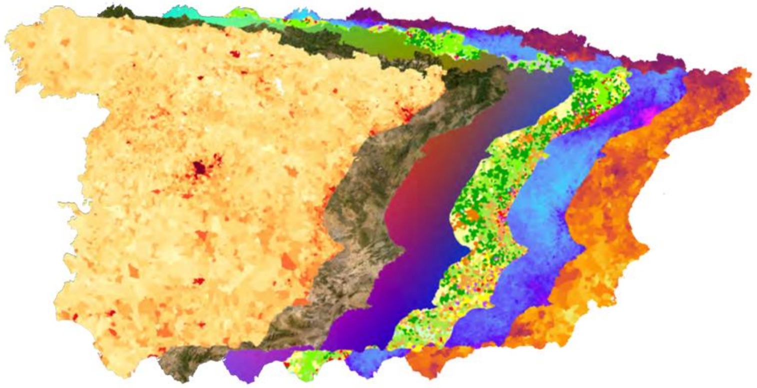
Is there one Embedding Strategy to rule them all? - No

For visual Tasks: Land Cover Classifications Image Feature Encodings are better

For contextual Tasks: Embedding models SatCLIP/GeoCLIP are better

	Model ↓ Task →	CLC % Accuracy	T_{min} R^2	T_{max} R^2	PM2.5 R^2	$E100k_1$ R^2
Expl. Features — BL	Mean reflection	33.53 ± 0.66	0.69 ± 0.01	0.71 ± 0.01	0.38 ± 0.01	0.29 ± 0.01
	Random	24.60 ± 0.38	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Expl. Features — Image Enc.	RN18-MoCo_{fc}	46.93 ± 0.68	0.65 ± 0.01	0.77 ± 0.00	0.44 ± 0.01	0.28 ± 0.01
	RN50-DINO_{fc}	47.50 ± 0.56	0.80 ± 0.00	**0.87** ± 0.00	0.54 ± 0.02	0.37 ± 0.01
	RN50-MoCo_{fc}	46.21 ± 0.61	0.73 ± 0.01	0.83 ± 0.00	0.49 ± 0.01	0.32 ± 0.02
	ViT-DINO	44.83 ± 0.66	0.79 ± 0.00	0.85 ± 0.00	0.49 ± 0.02	0.32 ± 0.02
	ViT-MoCo	**49.69** ± 0.37	0.76 ± 0.01	0.81 ± 0.00	0.42 ± 0.02	0.29 ± 0.01
	MMEarth	44.70 ± 0.69	0.71 ± 0.01	0.80 ± 0.00	0.44 ± 0.02	0.33 ± 0.02
Impl. Emb. Models — Location Enc.	$\text{SatCLIP-RN18}_{L=10}$	30.30 ± 0.38	0.75 ± 0.22	0.73 ± 0.23	0.43 ± 0.45	0.44 ± 0.05
	$\text{SatCLIP-RN18}_{L=40}$	31.91 ± 0.90	**0.86** ± 0.02	0.84 ± 0.02	**0.62** ± 0.03	**0.50** ± 0.01
	$\text{SatCLIP-RN50}_{L=10}$	30.25 ± 0.45	0.82 ± 0.06	0.80 ± 0.03	0.48 ± 0.18	0.43 ± 0.05
	$\text{SatCLIP-RN50}_{L=40}$	32.25 ± 0.64	**0.87** ± 0.01	0.84 ± 0.01	0.53 ± 0.40	**0.51** ± 0.01
	$\text{SatCLIP-ViT}_{L=10}$	30.46 ± 0.81	0.69 ± 0.16	0.77 ± 0.10	-0.07 ± 1.84	0.19 ± 0.65
	$\text{SatCLIP-ViT}_{L=40}$	32.25 ± 0.49	**0.87** ± 0.00	0.84 ± 0.01	**0.63** ± 0.04	**0.50** ± 0.03
	GeoCLIP	30.29 ± 0.82	0.83 ± 0.00	0.82 ± 0.00	0.59 ± 0.01	0.46 ± 0.01

PM2.5: Air Quality
E100k: Disease Probability



van Krieken, L. (2025). ***Comparing image and location encoder models in the context of disease mapping*** (MSc thesis, Wageningen University & Research).

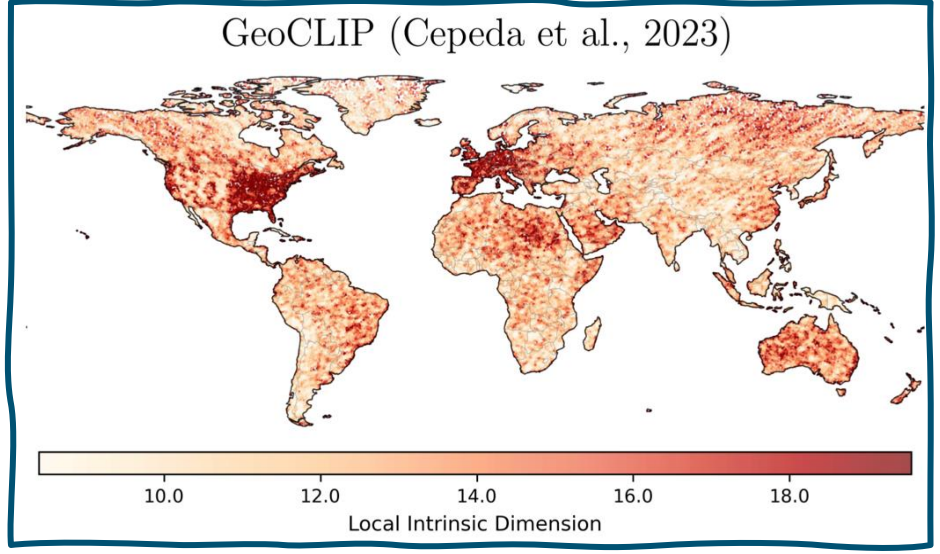
Levien van Krieken
WUR MSc Thesis

Intrinsic Dimension (ID) to Analyze Current Embeddings

Model	*D*	FisherS	MLE	MOM	TLE
Location encoders, Land sampling (100k points)					
SatCLIP–L10	256	5.00	1.96	2.02	2.16
SatCLIP–L40	256	**8.08**	2.03	2.39	2.32
GeoCLIP	512	7.68	**11.21**	**13.02**	**11.53**
CSP–fMoW	256	1.70	5.18	5.23	6.25
CSP–iNat	256	0.92	3.37	4.64	4.14
Image encoders on S2-100K (Klemmer et al., 2025)					
RCF	512	1.64	6.32	5.23	7.10
CROMA	768	**9.79**	19.57	17.00	20.30
DOFA	768	3.32	15.58	13.78	16.20
ResNet18	512	6.32	16.14	12.27	16.80
ResNet50	2048	6.42	16.27	13.18	17.00
ResNet152	2048	7.60	**20.72**	**17.50**	**21.50**
ViT-Small	384	3.33	18.53	15.80	19.20
ScaleMAE (RGB)	1024	2.96	10.16	8.90	11.00
ResNet18 (RGB)	512	0.92	10.85	8.70	11.70
ResNet50 (RGB)	2048	0.92	9.92	8.10	10.80

The intrinsic dimension is much lower than the ambient dimension D

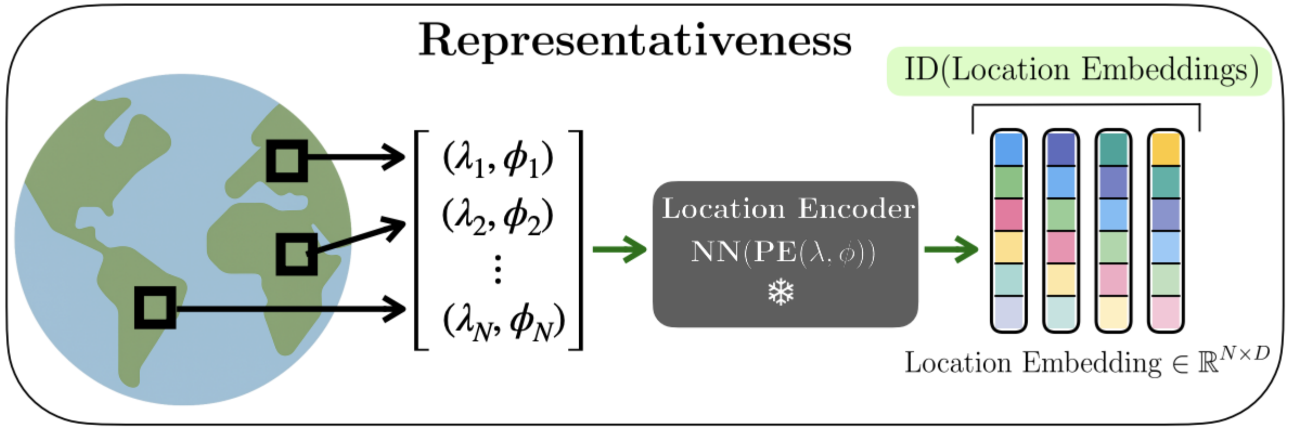
The intrinsic dimension reveals spatial artifacts due to training data biases

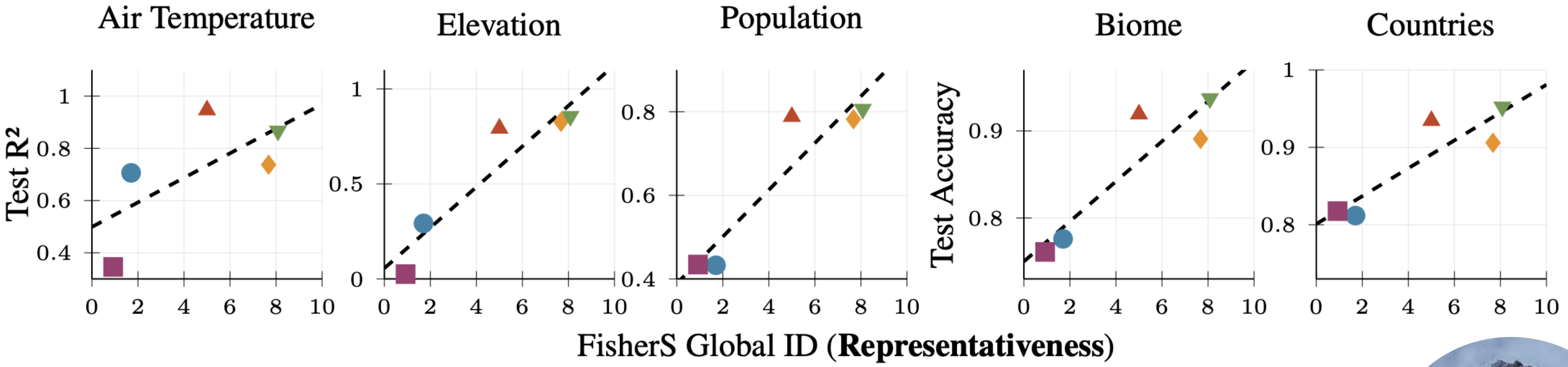


WAGENINGEN UNIVERSITY & RESEARCH

Rao, A., Rußwurm, M., Klemmer, K., & Rolf, E. (2025). **Measuring the Intrinsic Dimension of Earth Representations**. *arXiv preprint*

Representativeness: Higher Intrinsic Dimension Correlates with Task Performance

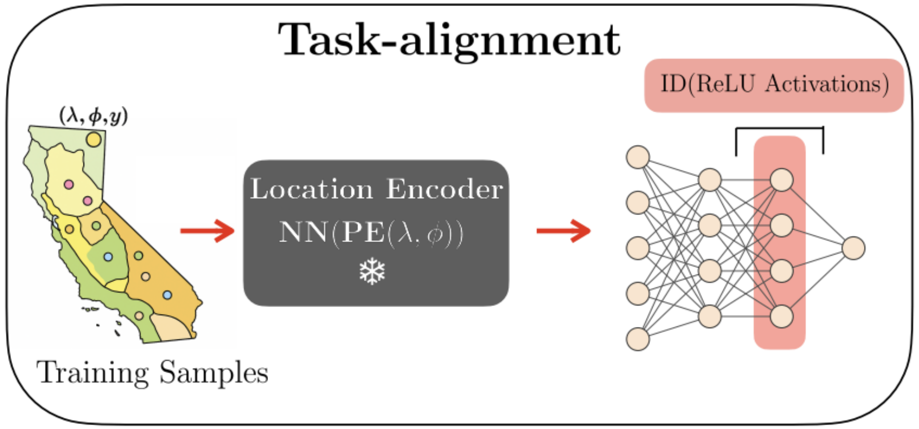


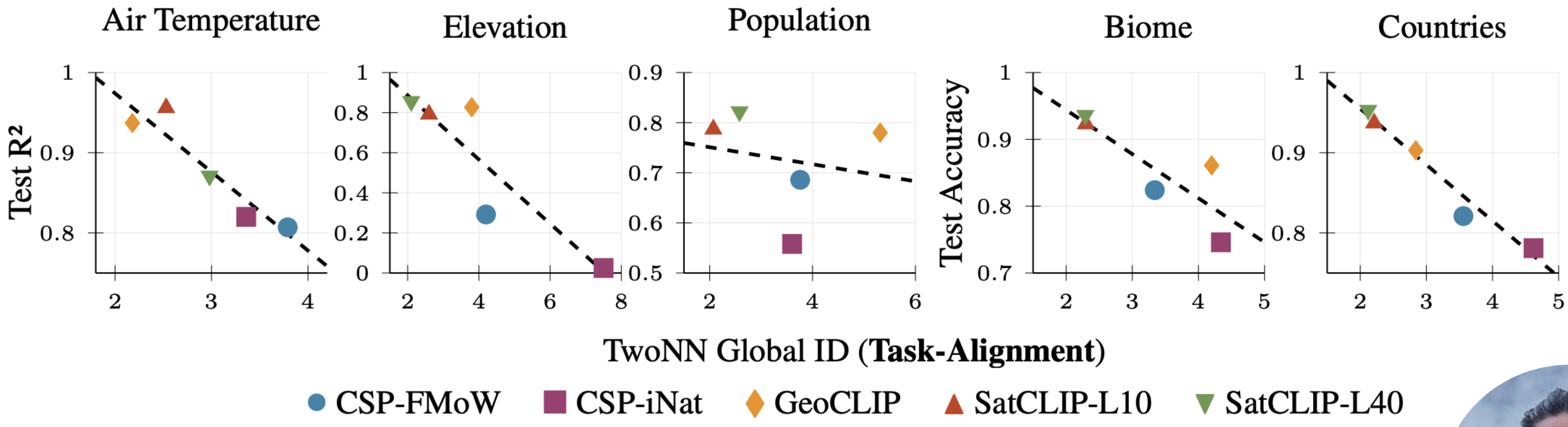


Rao, A., Rußwurm, M., Klemmer, K., & Rolf, E. (2025). **Measuring the Intrinsic Dimension of Earth Representations**. *arXiv preprint*

WAGENINGEN UNIVERSITY & RESEARCH

Task-alignment: Lower Intrinsic Dimension After Fine-tuning Correlates with Task Performance

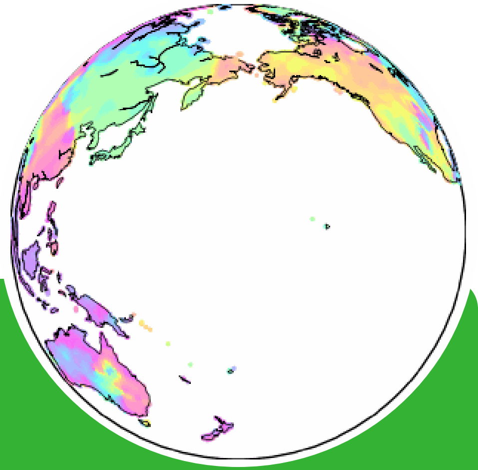




WAGENINGEN
UNIVERSITY & RESEARCH

Rao, A., Rußwurm, M., Klemmer, K., & Rolf, E. (2025). **Measuring the Intrinsic Dimension of Earth Representations**. *arXiv preprint*

Thank you, contacts & collaborators



Geographic Location Encoding:
Rußwurm et al., 2024, ICLR
https://marcrusswurm.com/locationencoder

More technical
45 minute
talk at the SSL4EO
Summer School 2024

Marc Rußwurm,
marc.russwurm@wur.nl

Assistant Professor
Wageningen University, Netherlands

Collaborators: Esther Rolf, Konstantin Klemmer



SatCLIP:
Klemmer et al., 2025, AAAI
https://github.com/microsoft/satclip
https://arxiv.org/pdf/2311.17179.pdf

Intrinsic Dimension:
Rao et al., 2025, ArXiv
https://github.com/arjunarao619/GeoINRID
https://arxiv.org/abs/2511.02101

WAGENINGEN UNIVERSITY & RESEARCH
HARVARD UNIVERSITY
EPFL
Microsoft