Global Geographic Location Encoding with Implicit Neural (Geo)Representations

Marc Rußwurm

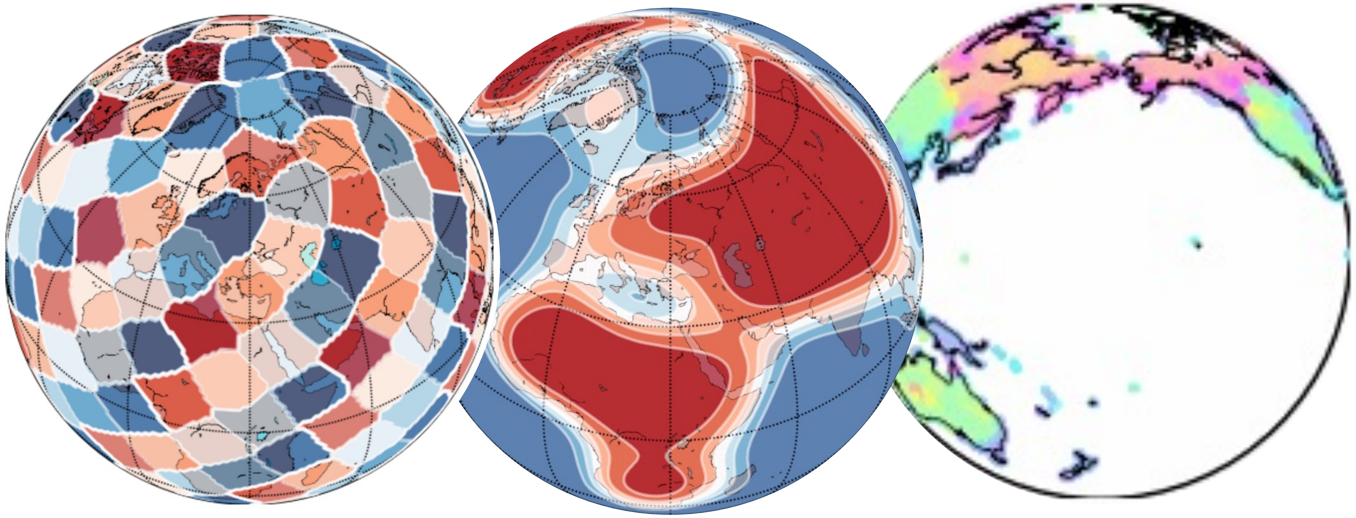
Assistant Professor
Machine Learning and Remote Sensing
Wageningen University

collaborators: Konstantin Klemmer, Esther Rolf, Robin Zbinden, Devis Tuia, Lester Mckay, Caleb Robinson

ELLIT Focus Period

Linköping

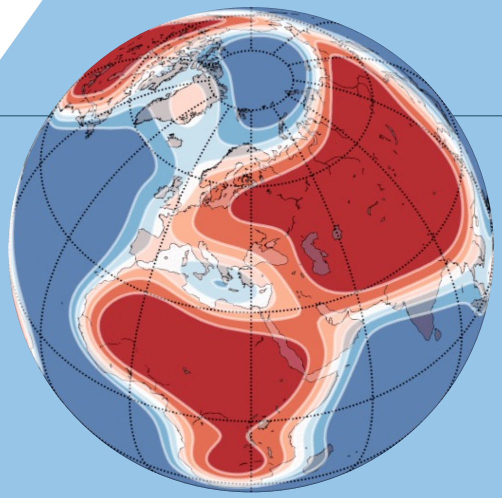
Oct 15th 2024



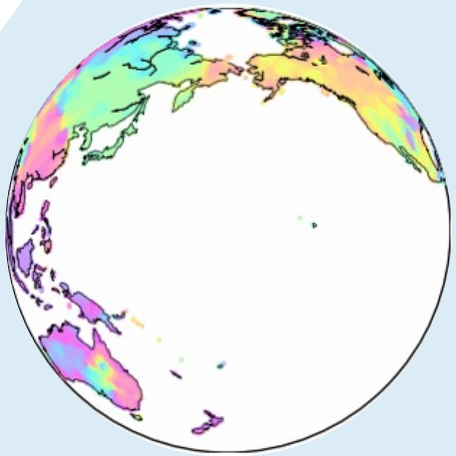
WAGENINGEN UNIVERSITY & RESEARCH HARVARD UNIVERSITY EPFL Microsoft

Outline

1 Spatial Modeling meets Implicit Neural Representations



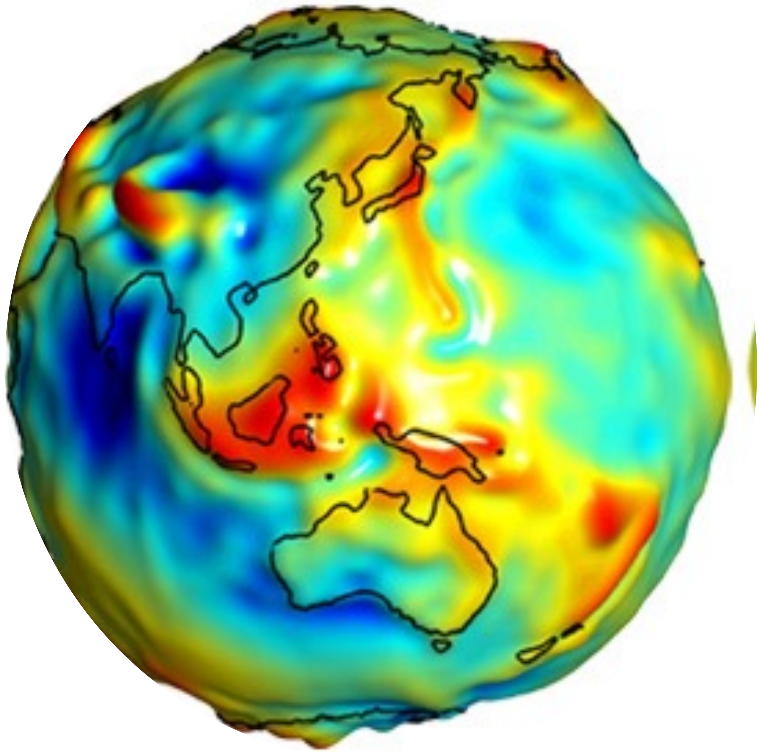
2 Siren(SH) Location Encoder
Rußwurm et al., ICLR 2024



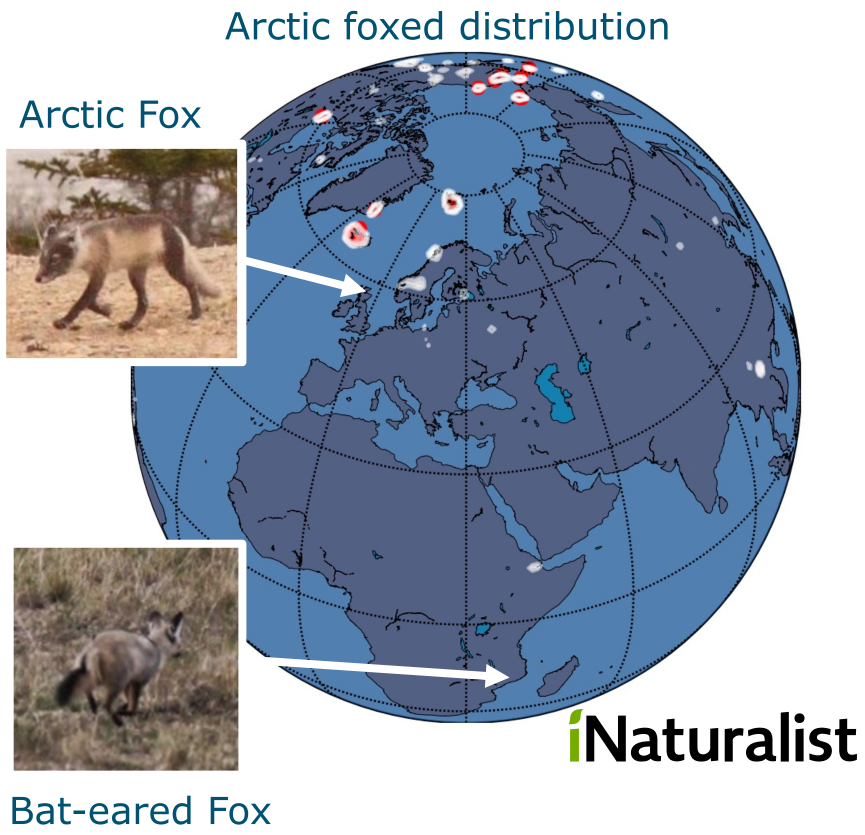
3 SatCLIP Encoder
Klemmer et al., 2024 – In review

Research Fields using Geolocated Data

representing **Earth's Gravity**



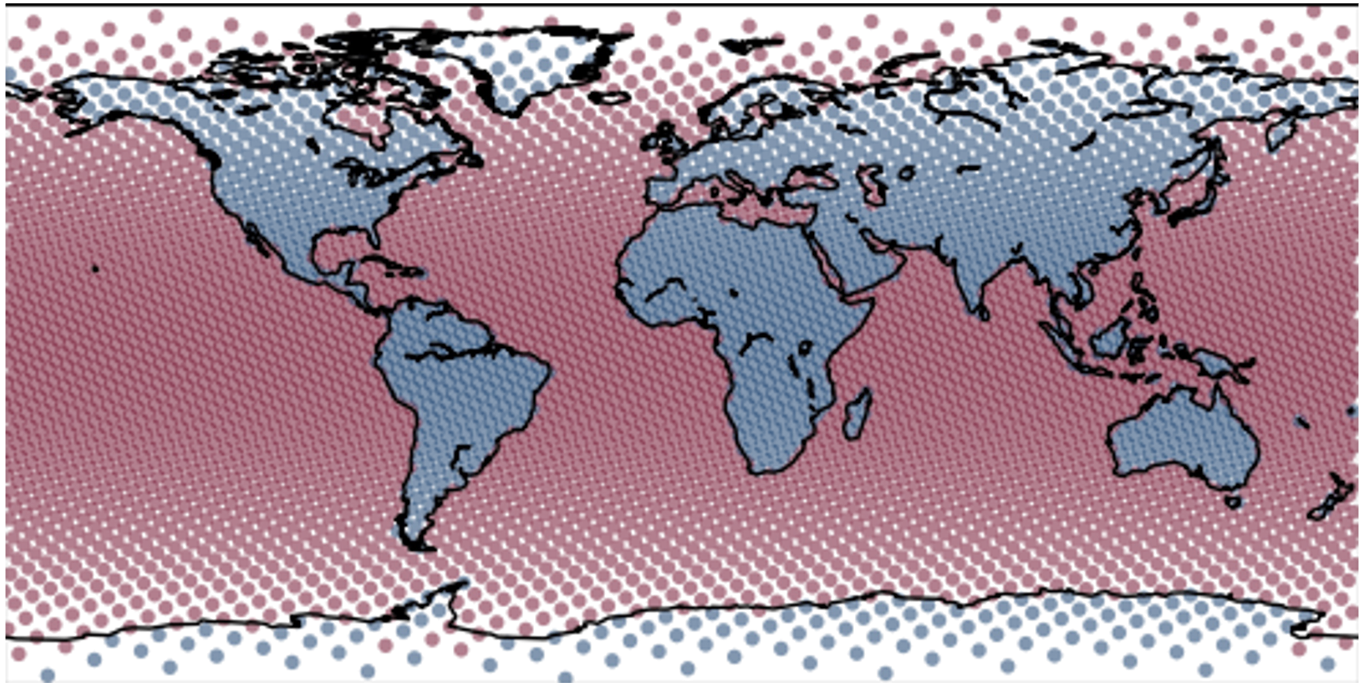
Species Mapping
Cole et al., 2023



Toy Example: Land-Ocean Classification

- Output y: 2 classes –
  land 0 and ocean 1
- Input: longitude, latitude

- Train: 10k points (random)
- Validation: 10k points (random)
- Test: 10k (grid)

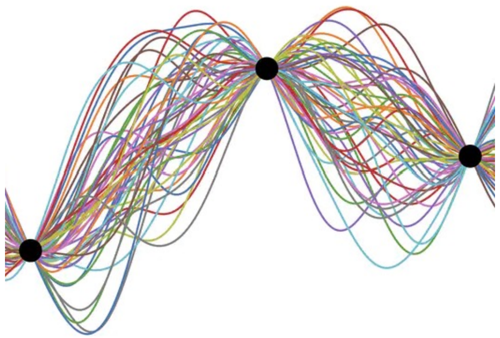


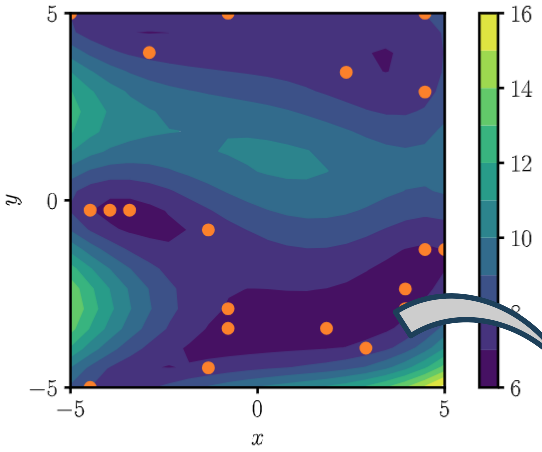
we learn a continuous function over space

Spatial Modeling

we learn a continuous function over space

In traditional geostatistics, we use various interpolation techniques, for examples Gaussian Processes (or "Kriging"):





problem → computationally expensive $O(n^2)$ with many points

because → we need to calculate similarities to „training" points

Location **c**

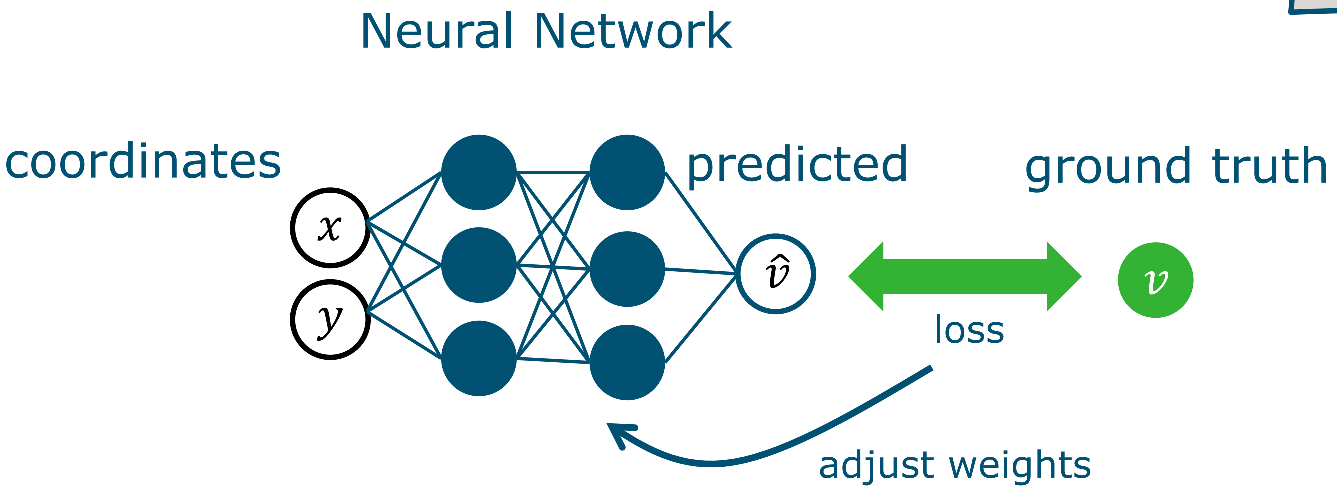
Gaussian Process

$P(y|\boldsymbol{c})$

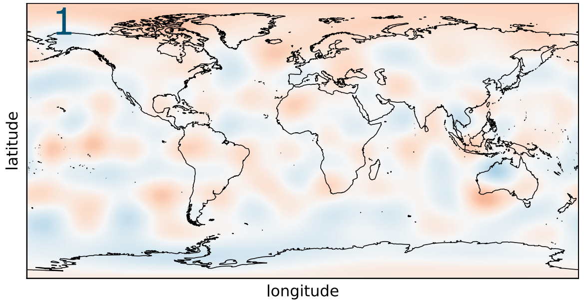
Temperature y:
-13.3 °C

Idea: Implicitly encode data in a neural network

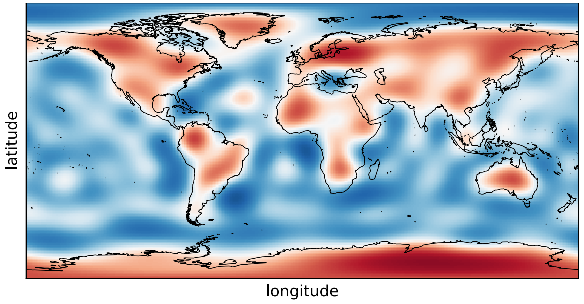
we learn a continuous function over space



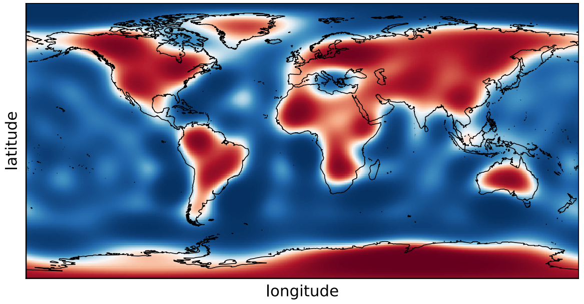
Epoch 1



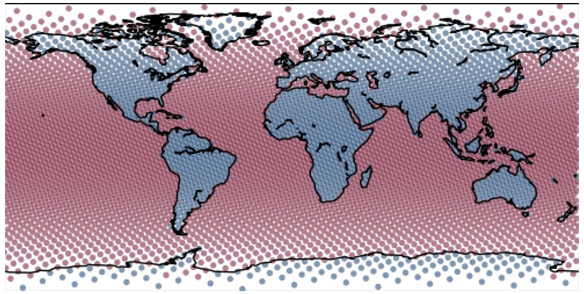
Epoch 100



Epoch 400

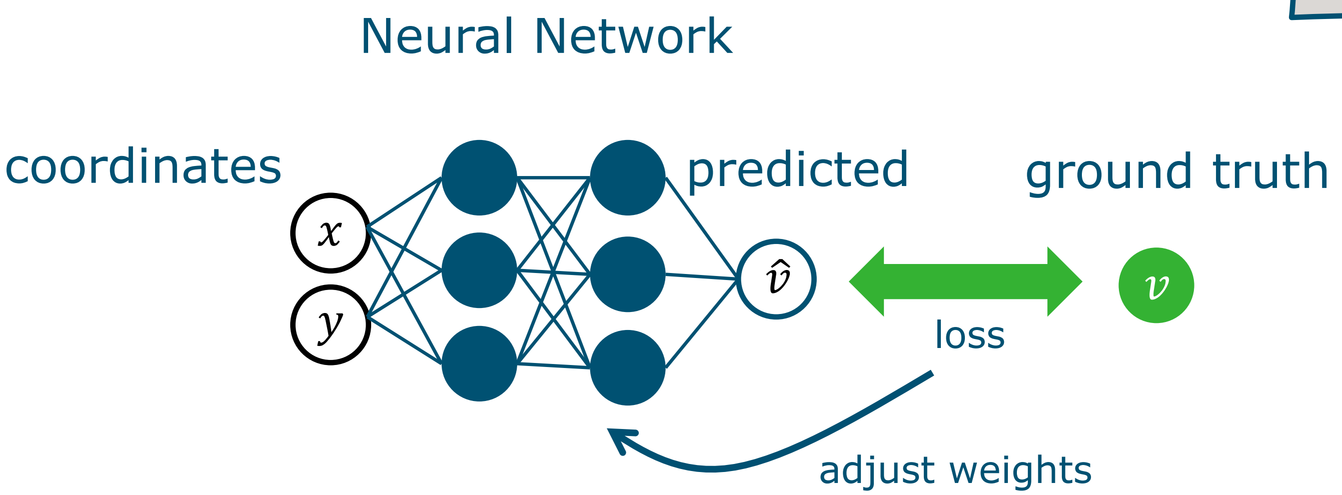


Ground truth:
Land-Ocean Classification



Idea: Implicitly encode data in a neural network

we learn a continuous function over space



we learn an **implicit neural representation** of the training data

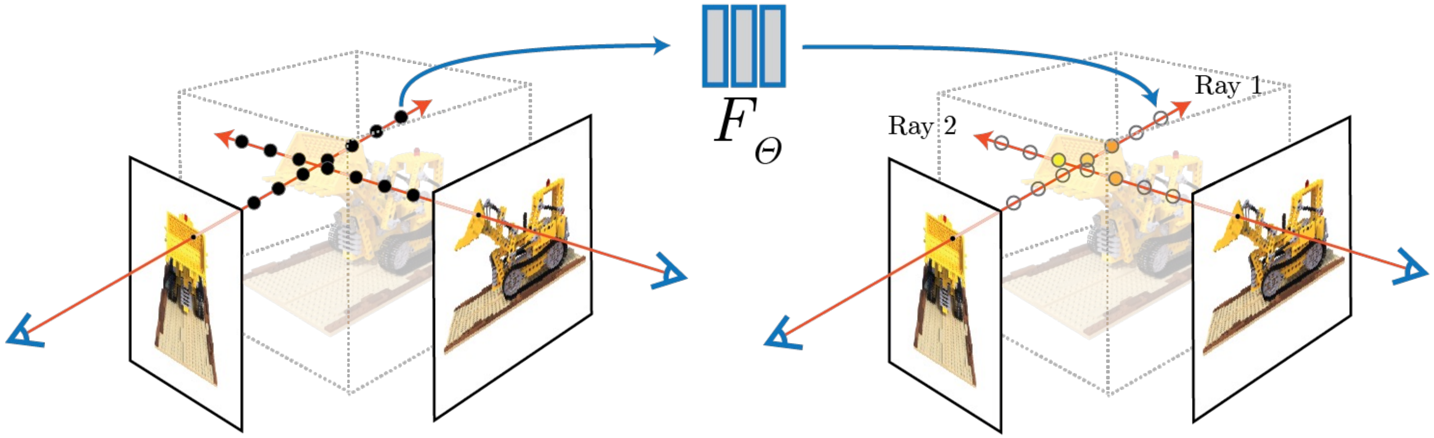
benefit: computationally linear $O(n)$ both in training and inference

benefit: very compatible with other deep models

Implicit neural representations are common in Vision

Neural Implicit Radiance Fields (NeRFs)

as prominent examples

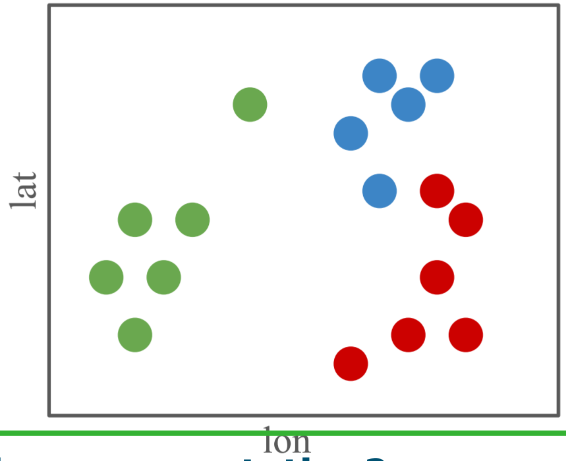


Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2021). Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, *65*(1), 99-106.

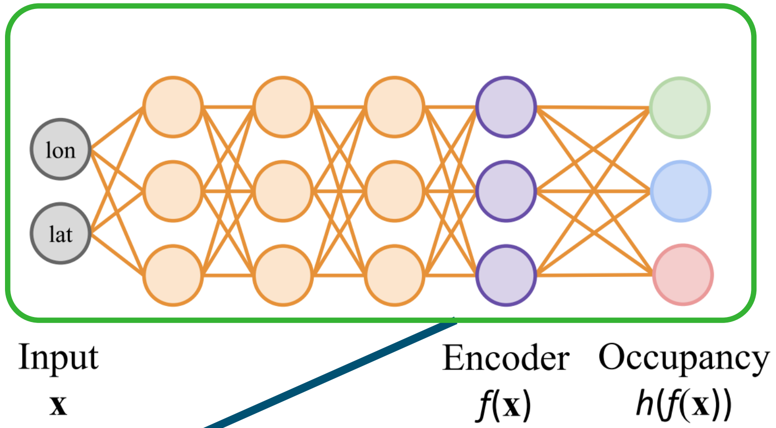
Implicit Neural Representations for Species Mapping

Input Data

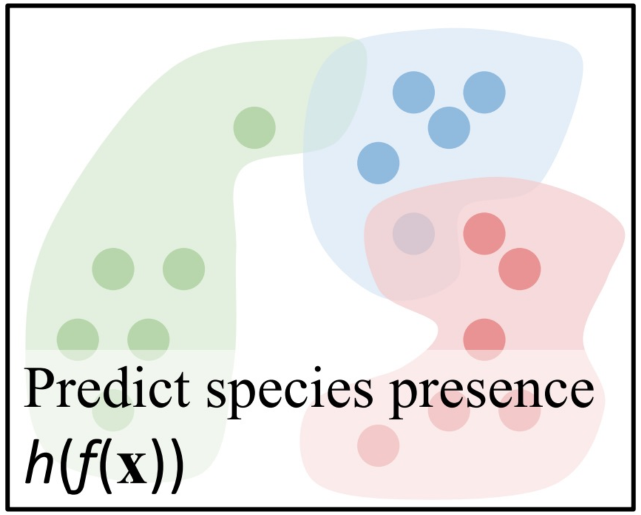
Species Presence



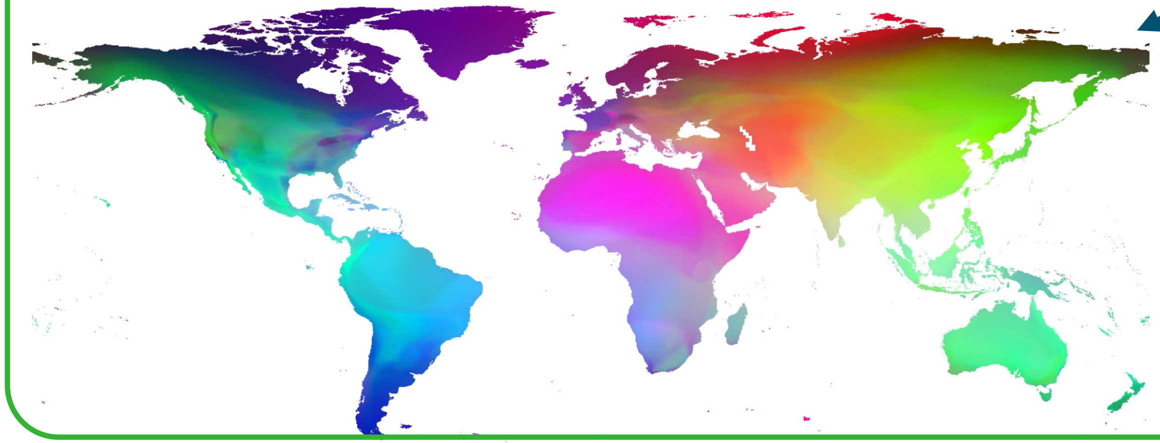
Species Range Model



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)

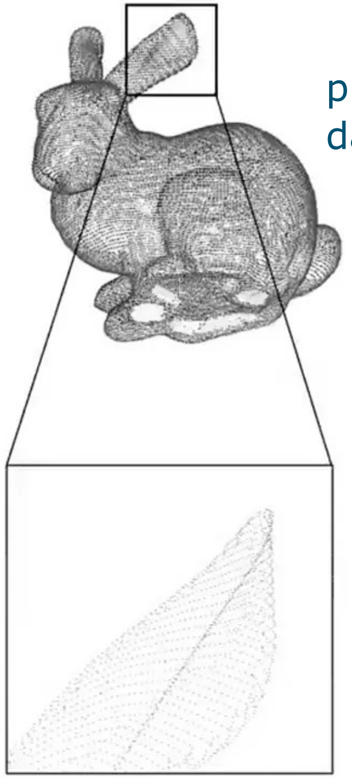
Sinusoidal Representation Networks (Siren)

images as
discrete grid of pixels

Images

shapes as
discrete points

Shapes

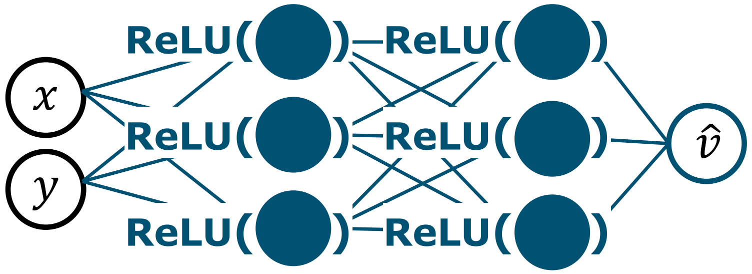


problem-specific
data structures

implicit neural
representation

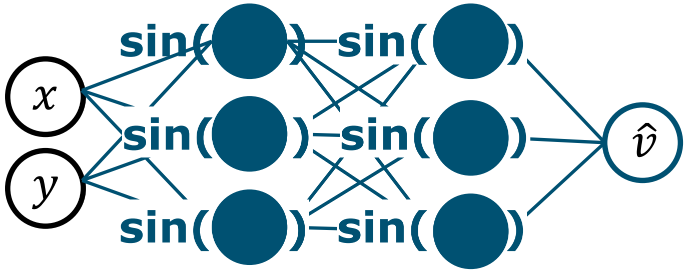
Neural Network:

Multi-layer Perceptron (MLP)
with ReLU activations



Sitzmann et al., 2020

Sirens
a MLP with sine activations

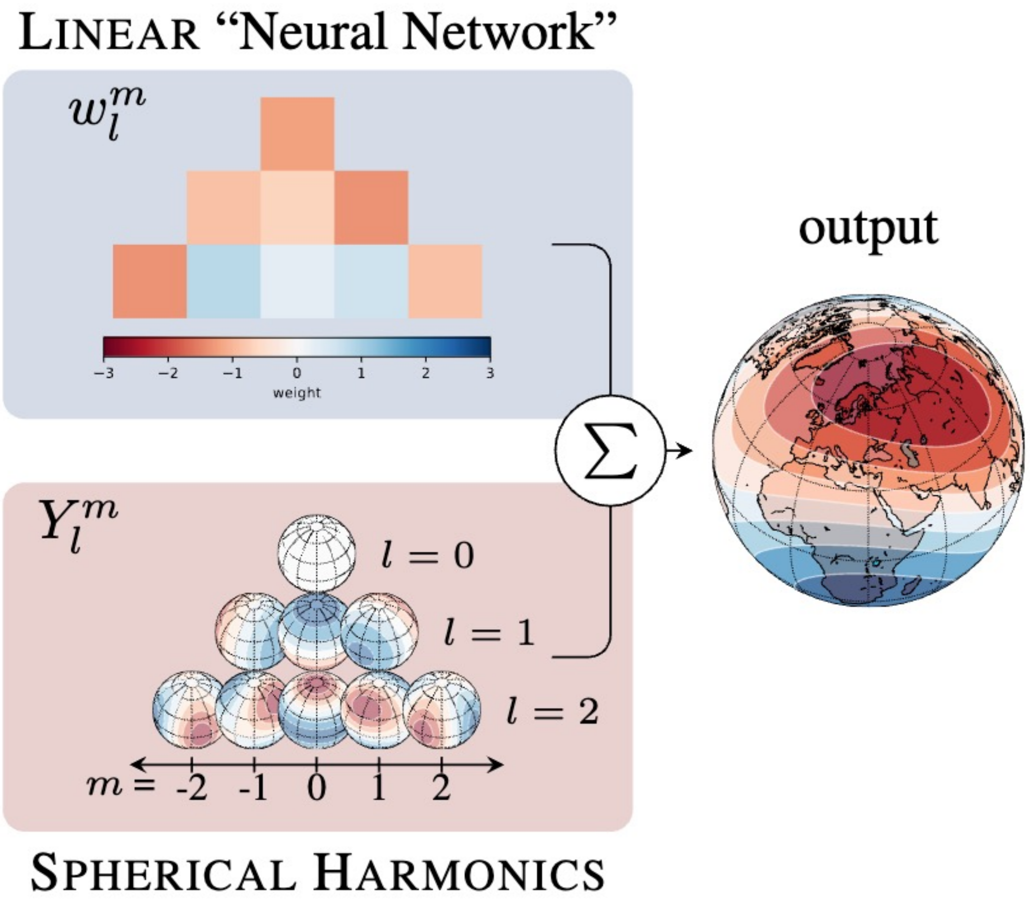


Vincent Sitzmann, Julien N. P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. 2020.
Implicit Neural Representations with Periodic Activation Functions. NeurIPS 2020

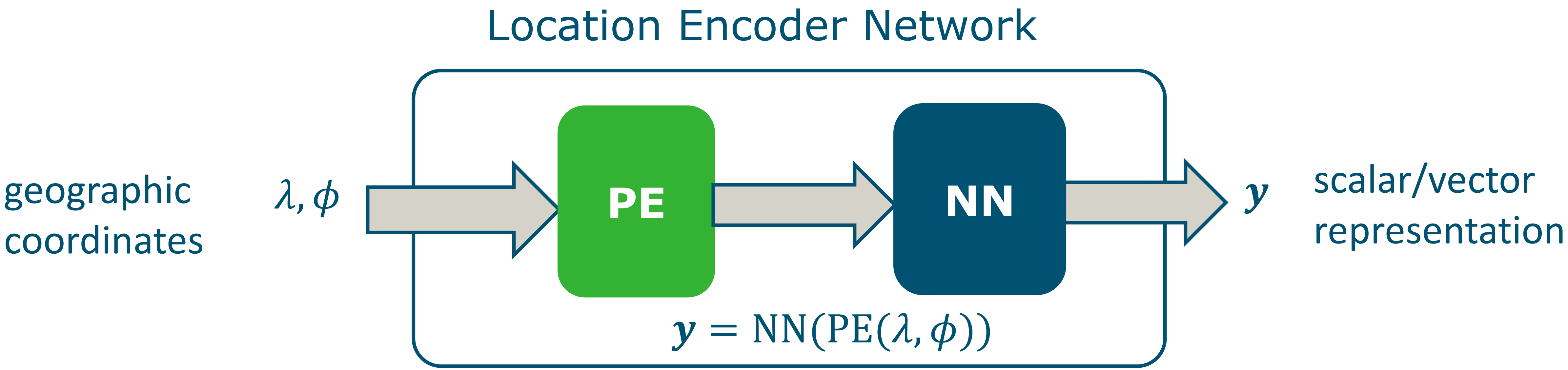
Geographic Location Encoding with Spherical Harmonics and Sinusoidal Representation Networks

International Conference on Learning Representations 2024

Marc Rußwurm, Konstantin Klemmer, Esther Rolf, Robin Zbinden, Devis Tuia



Location Encoder: Positional Embedding & Neural Network

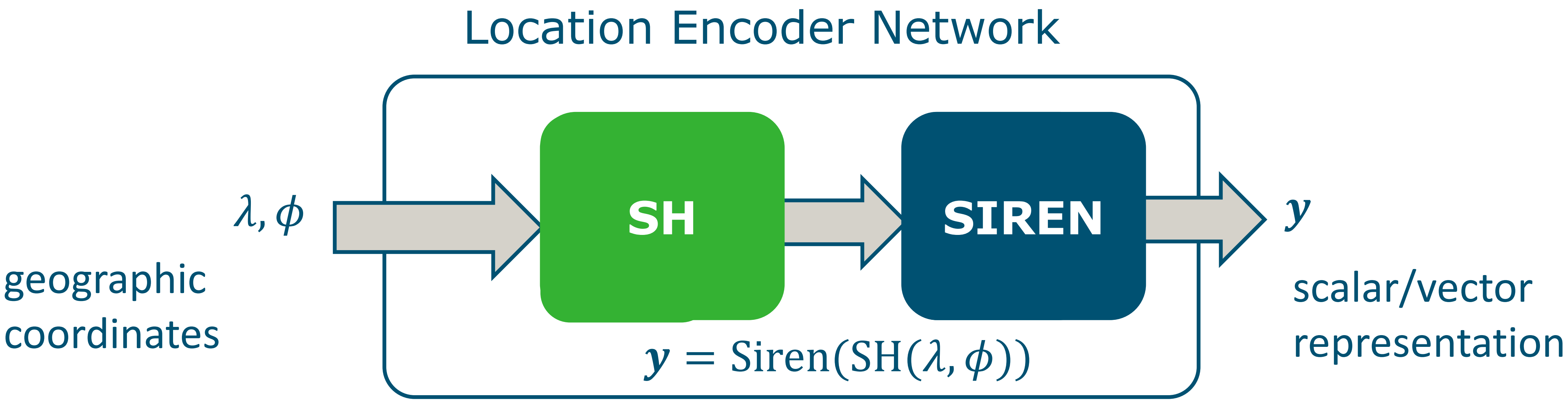


Positional Embedding PE
(non-parametric function)

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.

Our Proposition: Siren(SH(λ, ϕ))



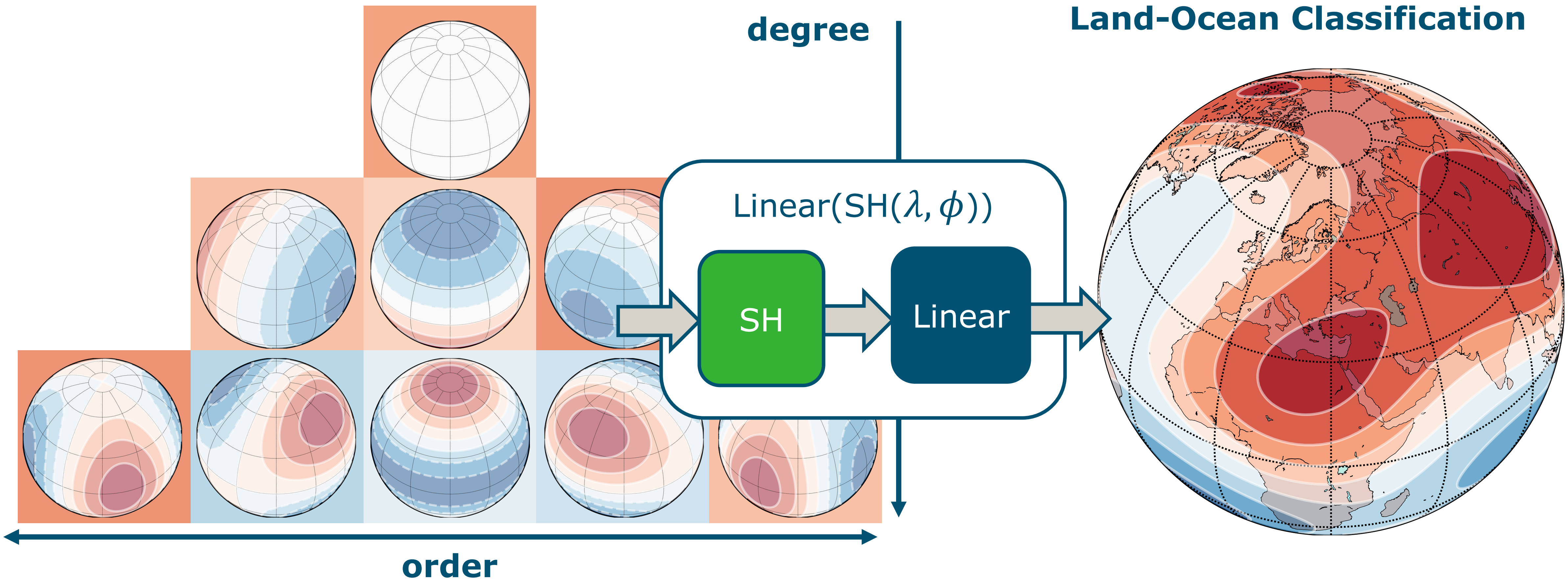
Spherical Harmonics basis functions
(non-parametric function)

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.

Spherical Harmonic (SH)

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



Earth's Gravity Field

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

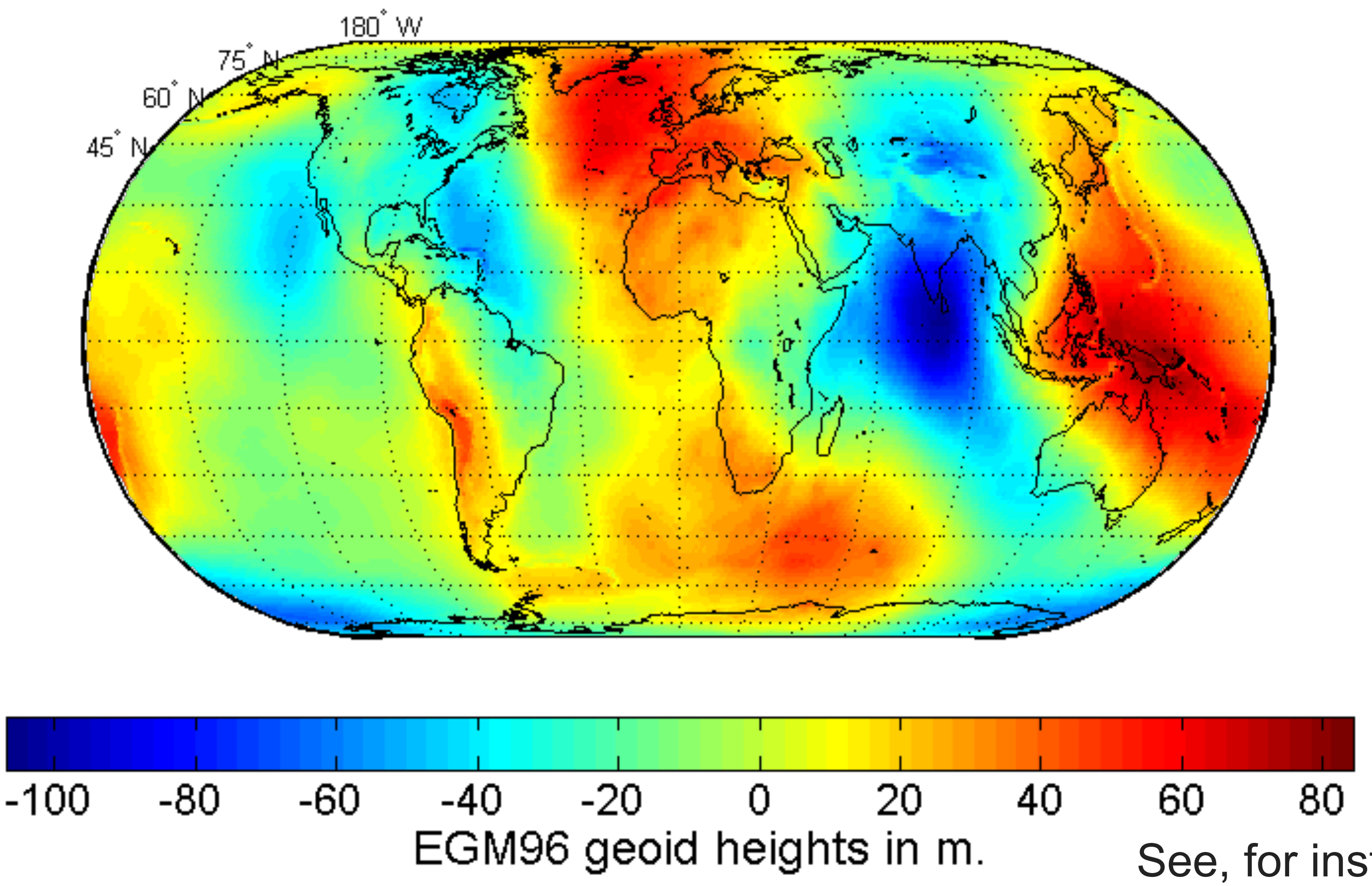
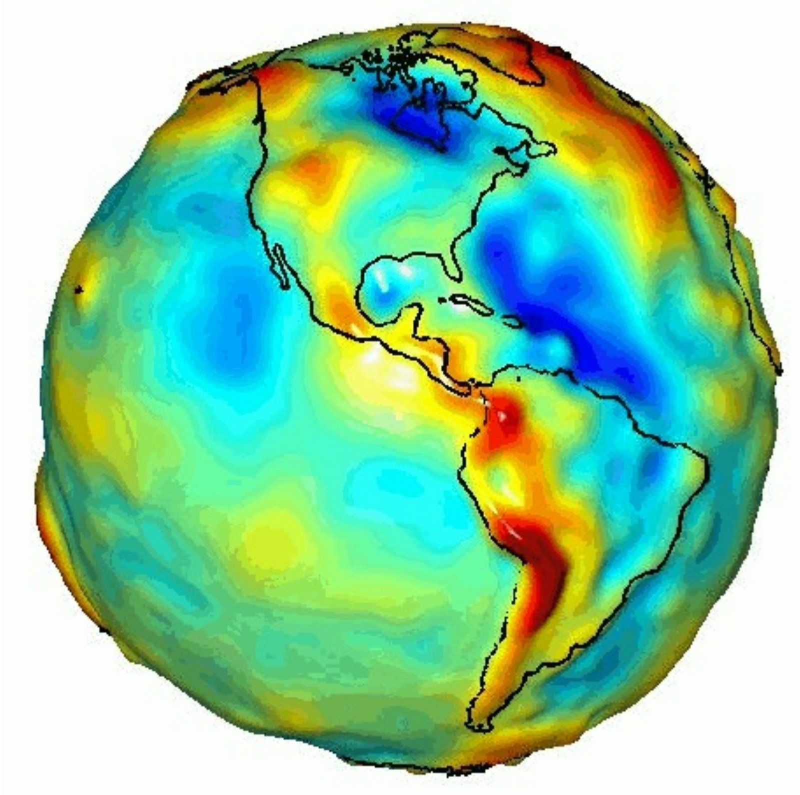
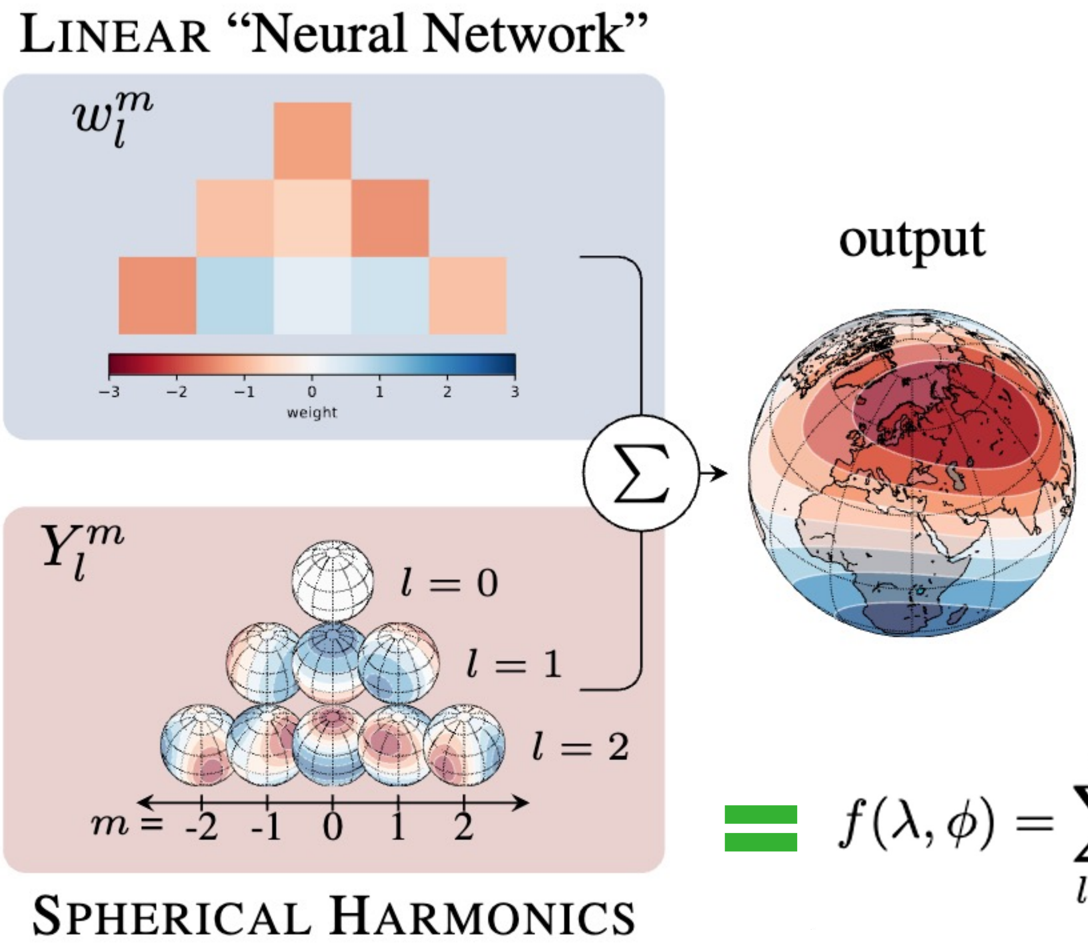


Image source wikipedia

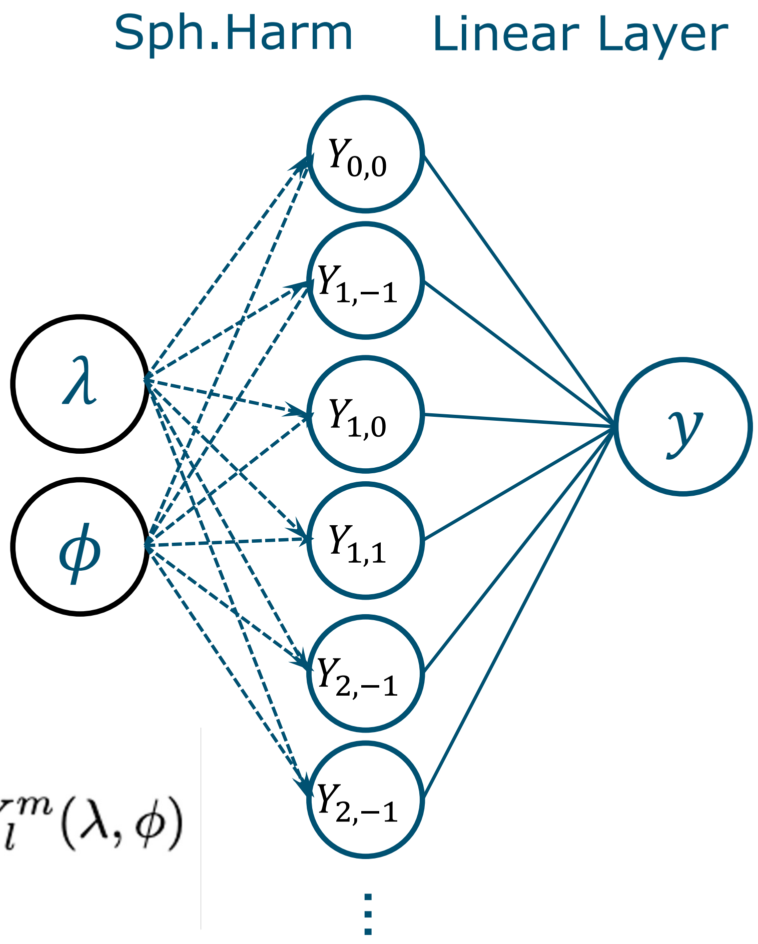


See, for instance, Pail, Roland, et al. "First GOCE gravity field models derived by three different approaches." *Journal of Geodesy* 85 (2011): 819-843.

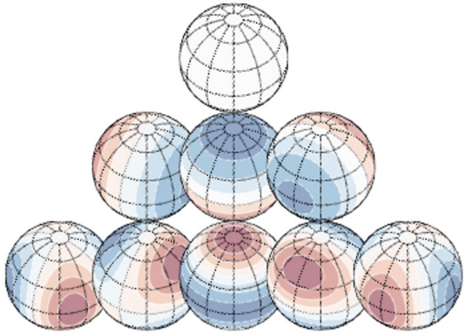
Spherical Harmonics as Positional Encoder

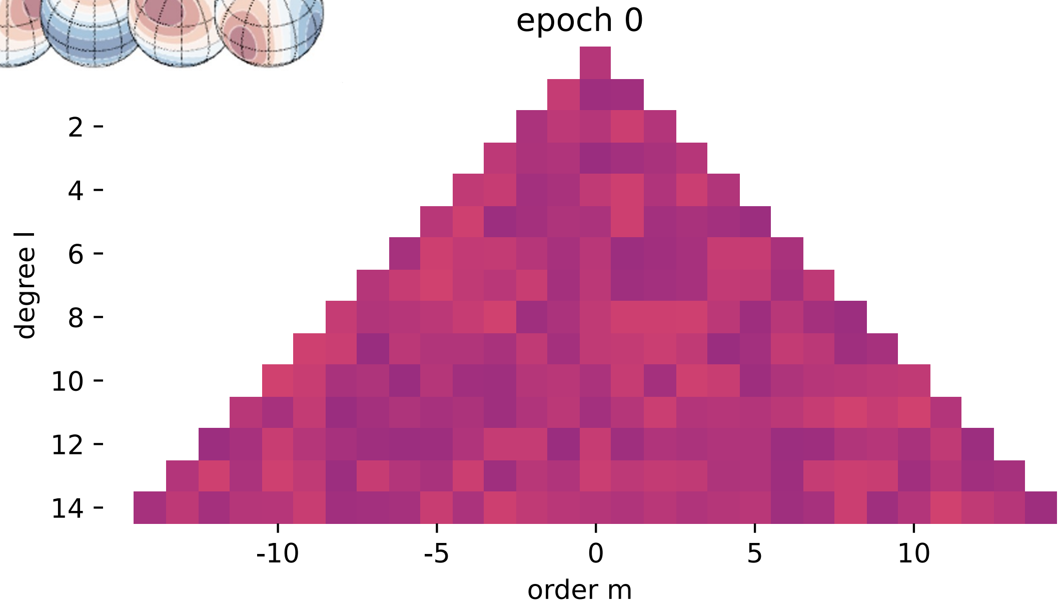


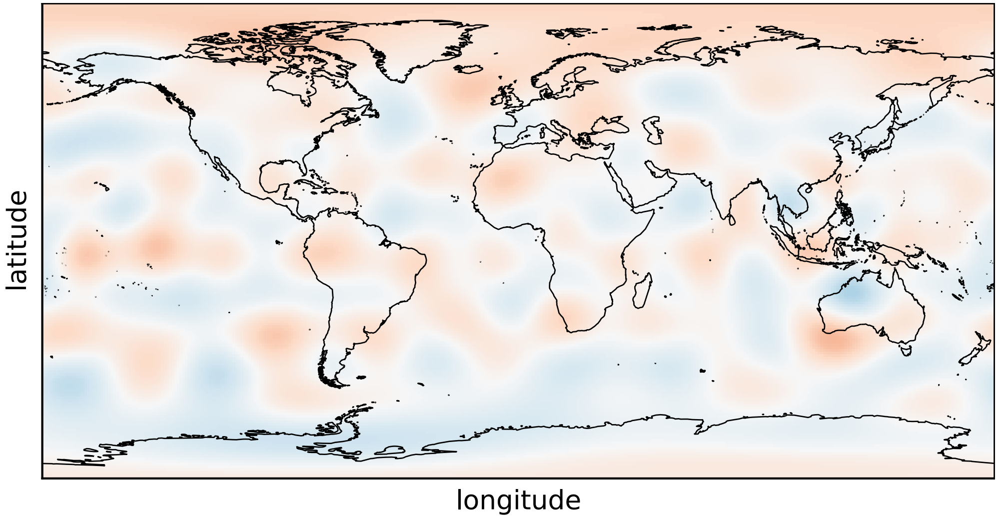
$$= \quad f(\lambda, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} w_l^m Y_l^m(\lambda, \phi)$$



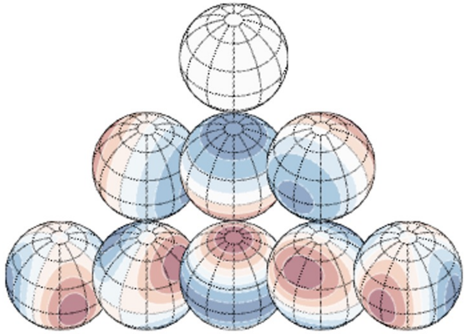
Training Epoch 0

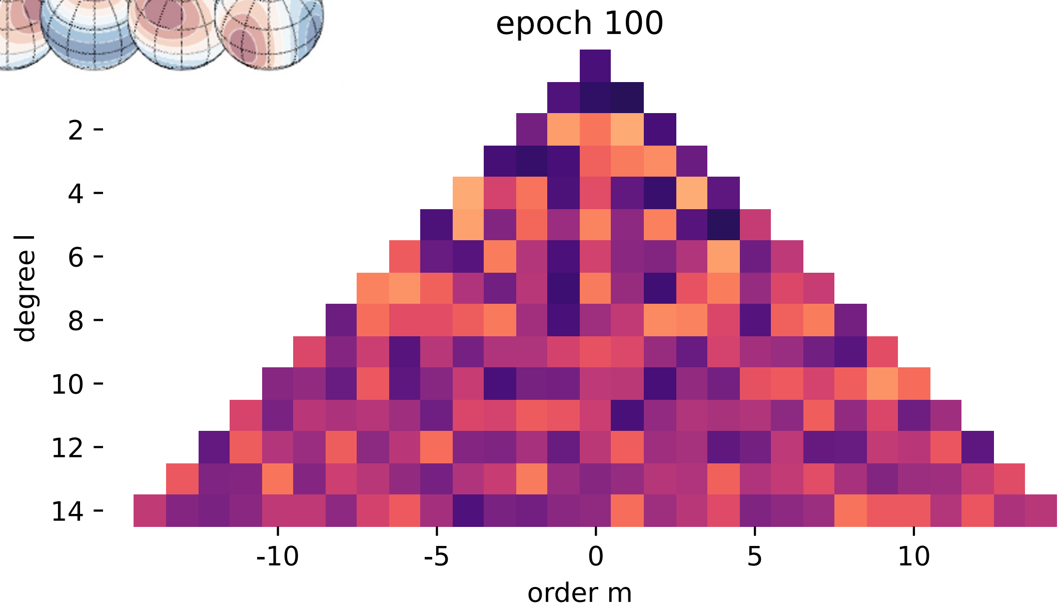


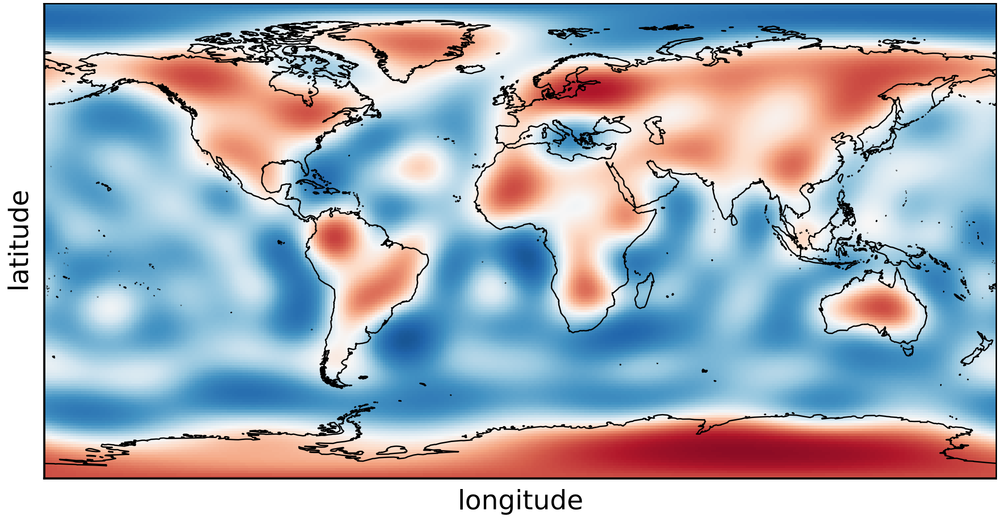




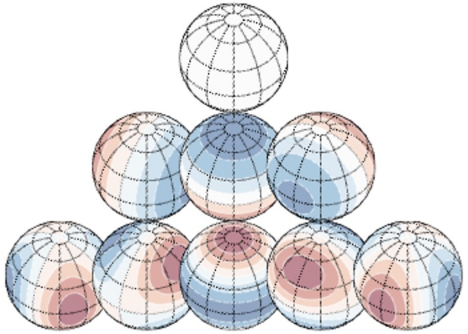
Training Epoch 100

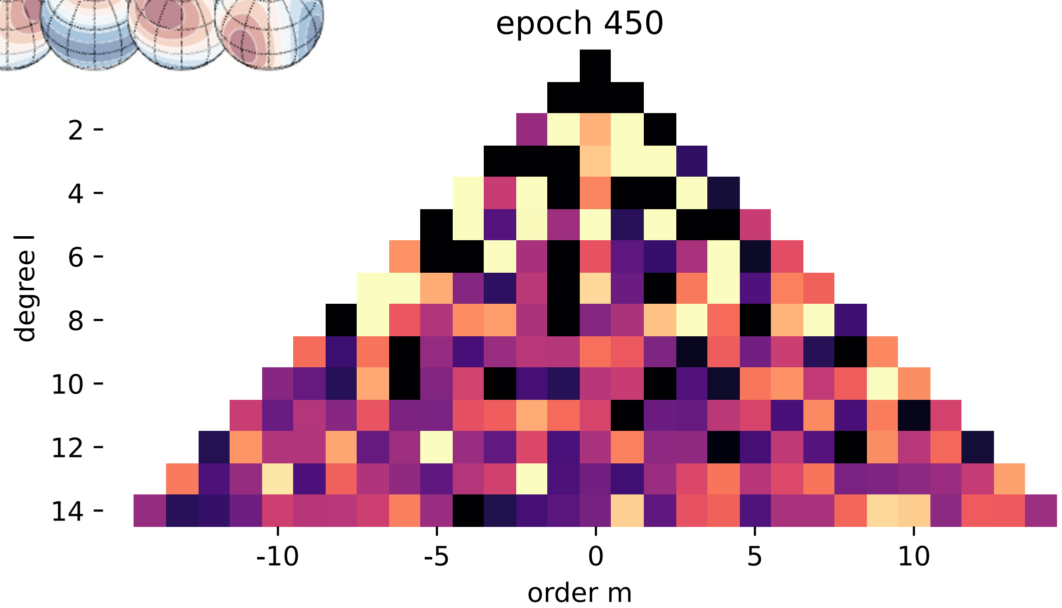


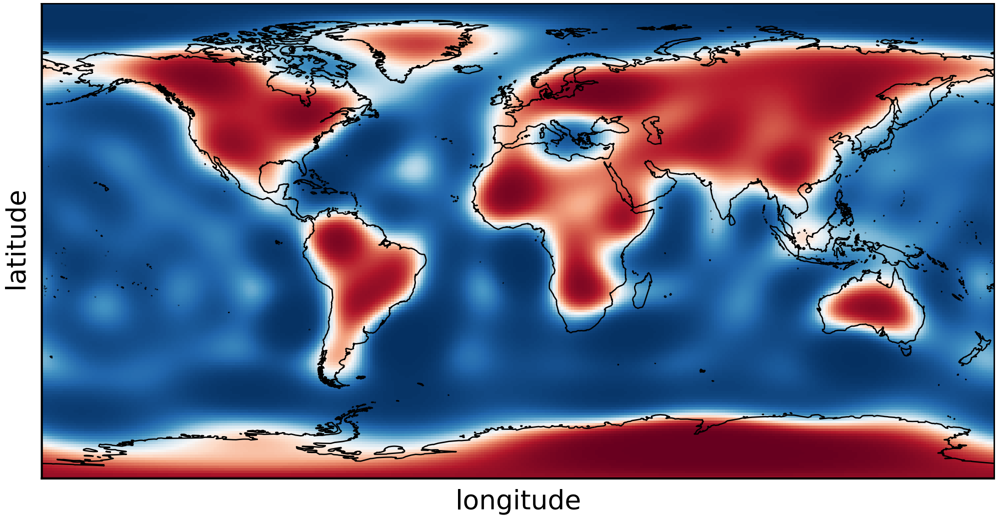




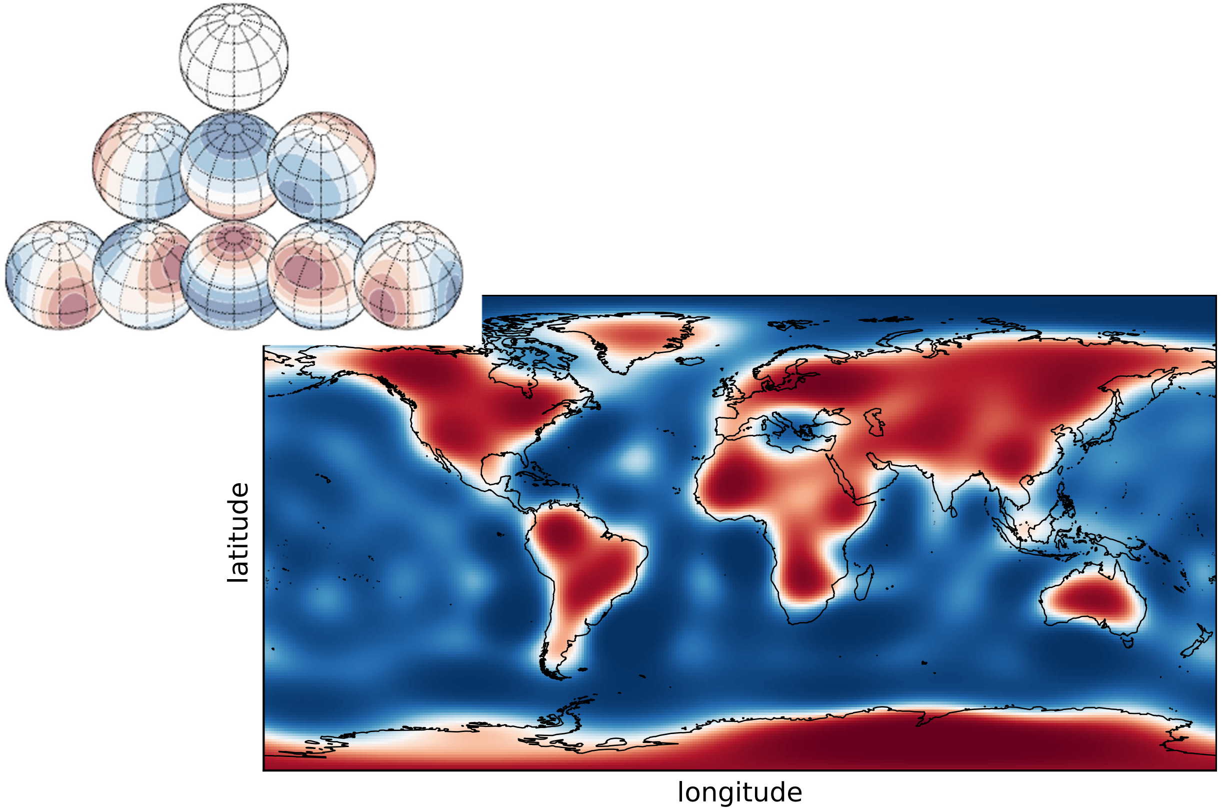
Training Epoch 450



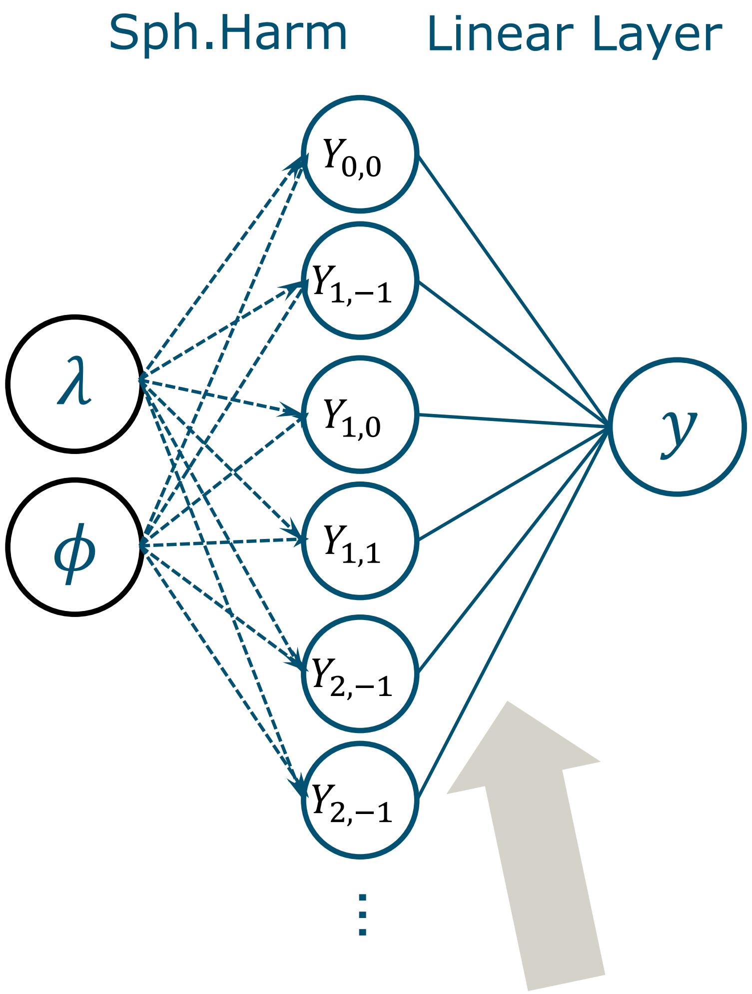




Training Epoch 450

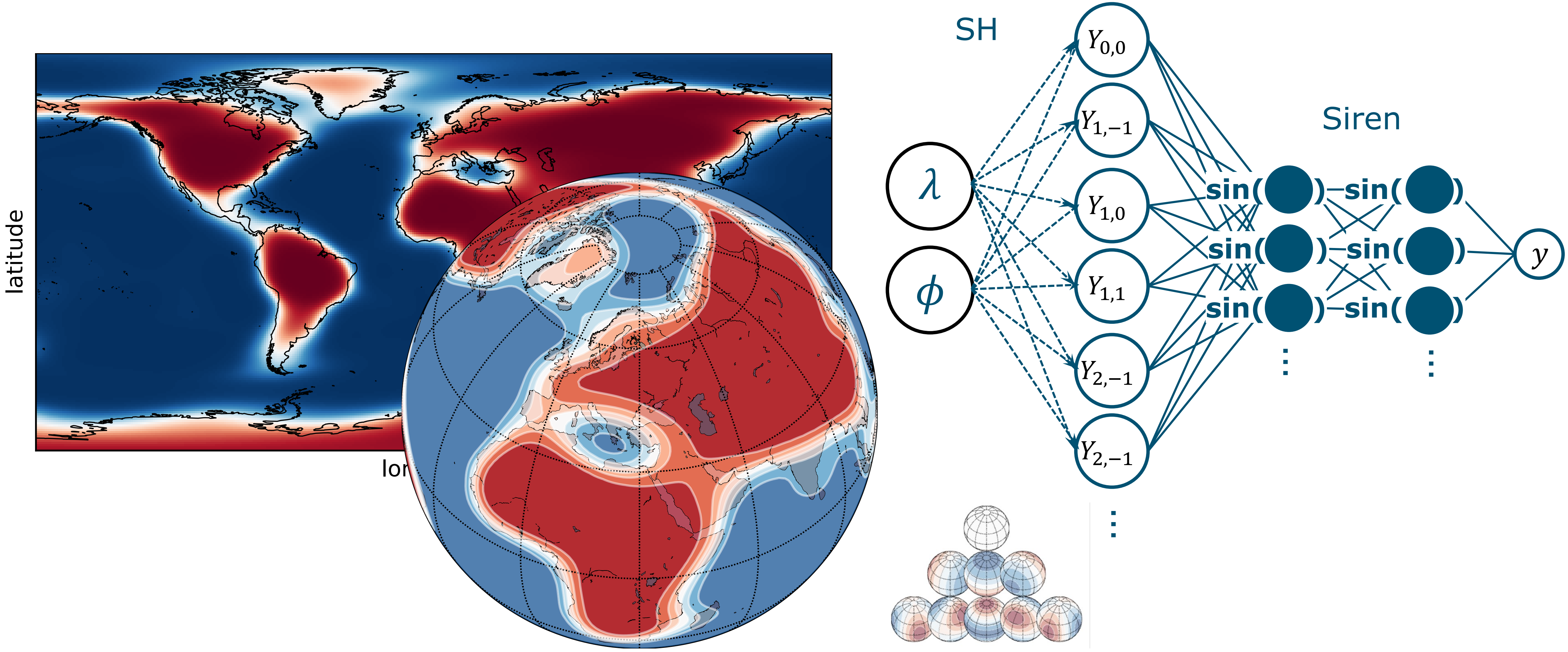


$$= \quad f(\lambda,\phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} w_l^m Y_l^m(\lambda,\phi)$$

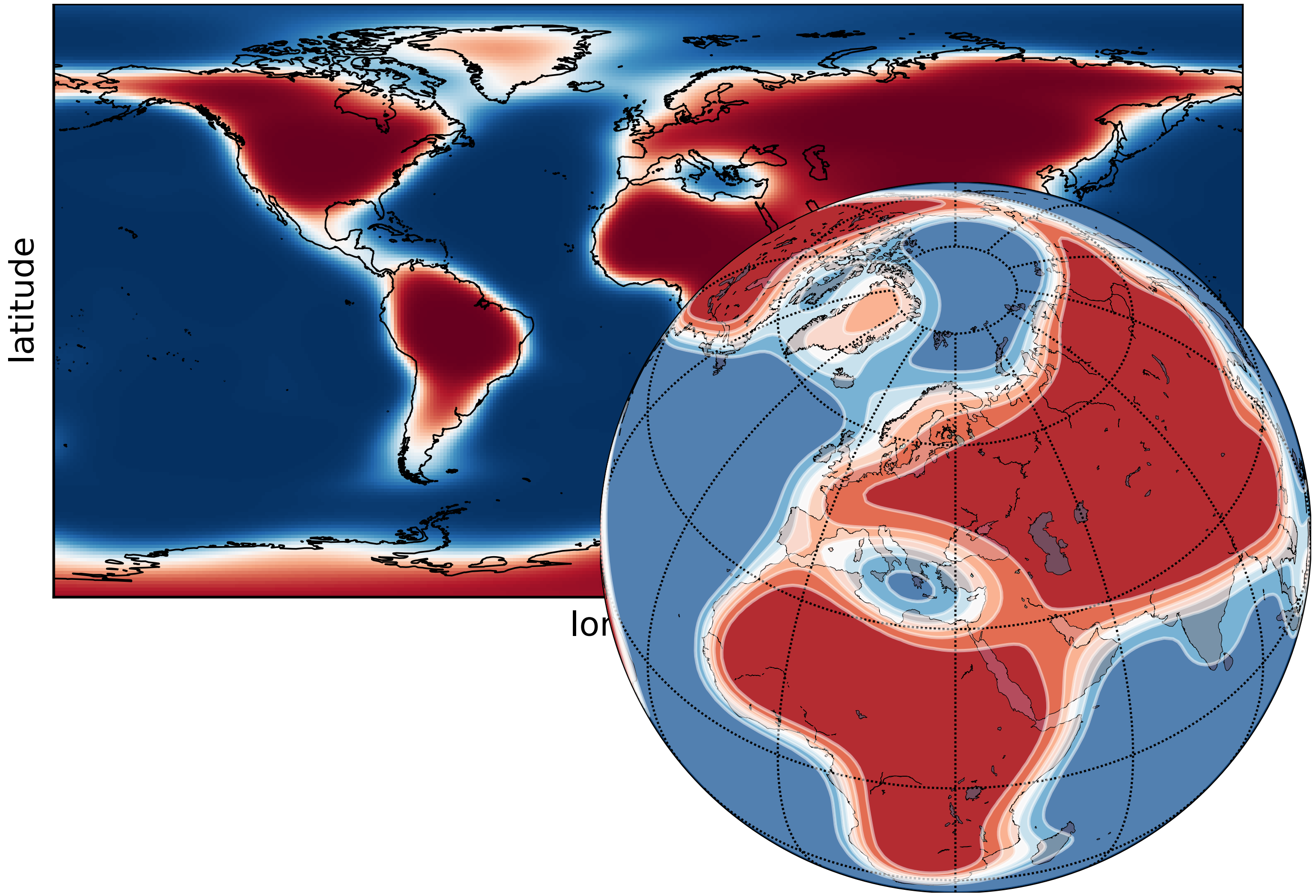


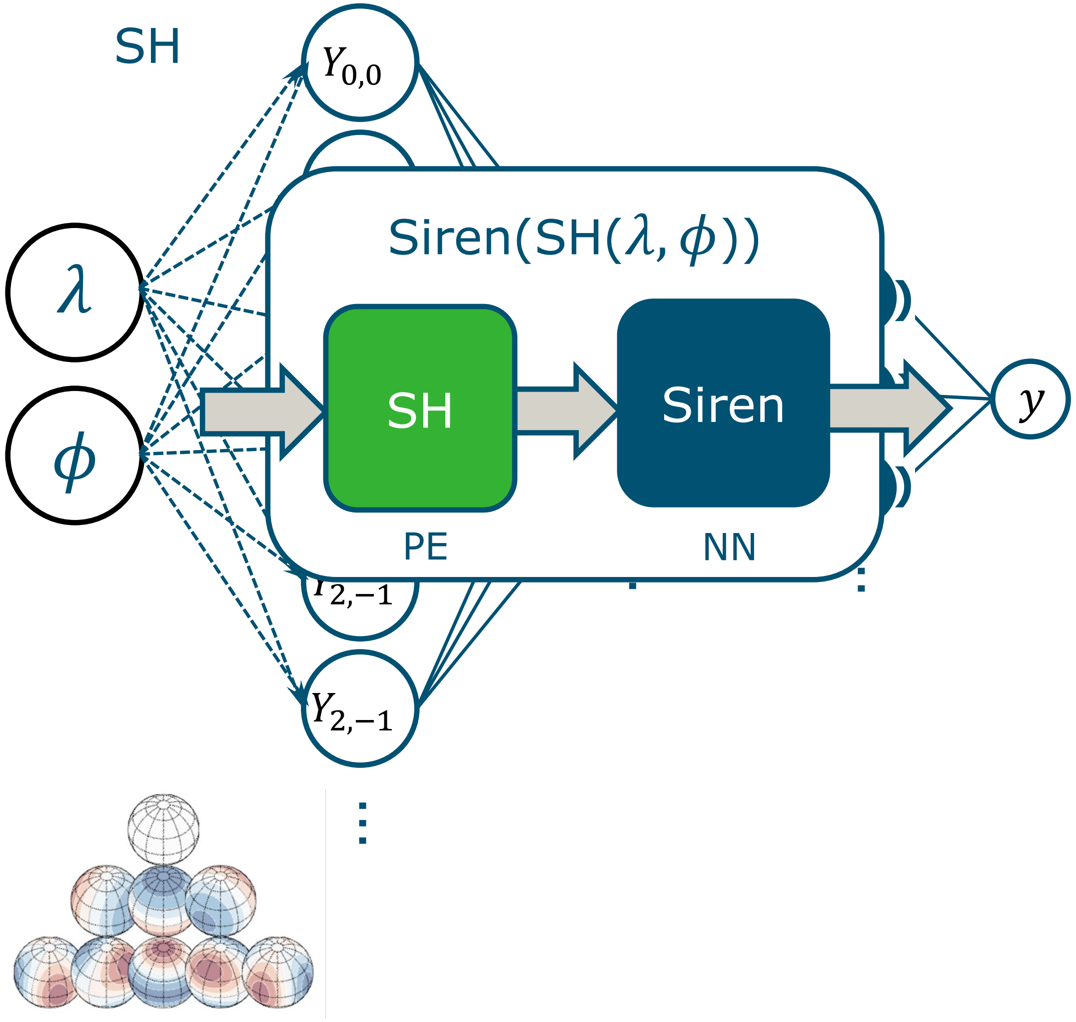
only a **single linear** layer with few trainable parameters

Adding Siren adds additional capacity

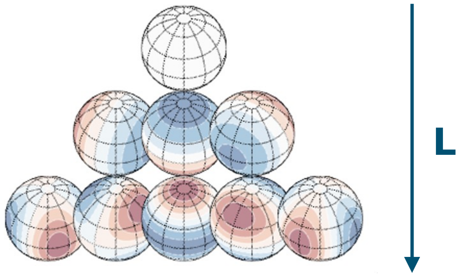


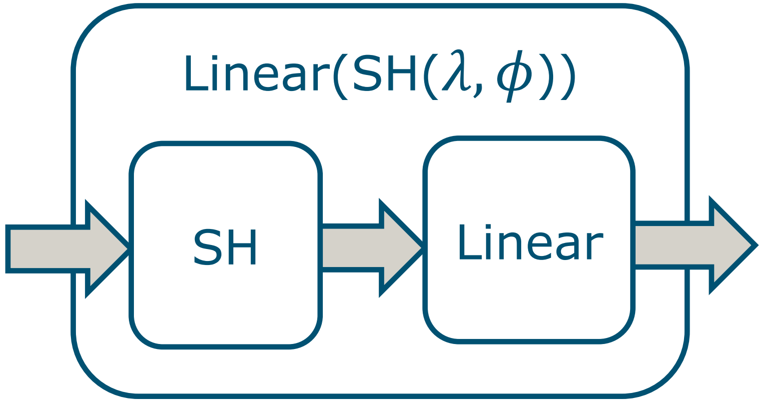
Adding Siren helps resolution

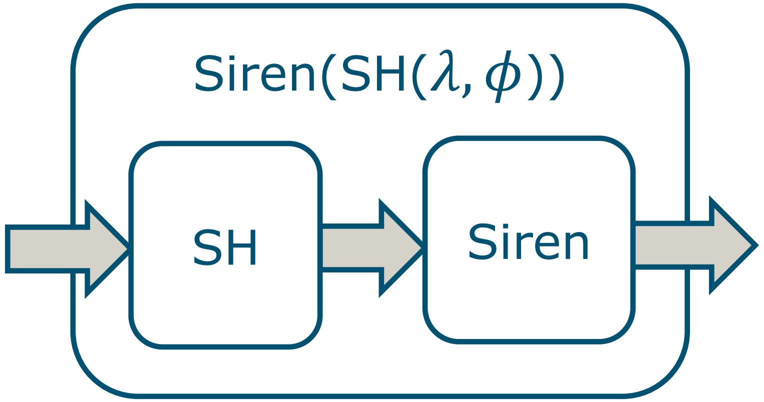


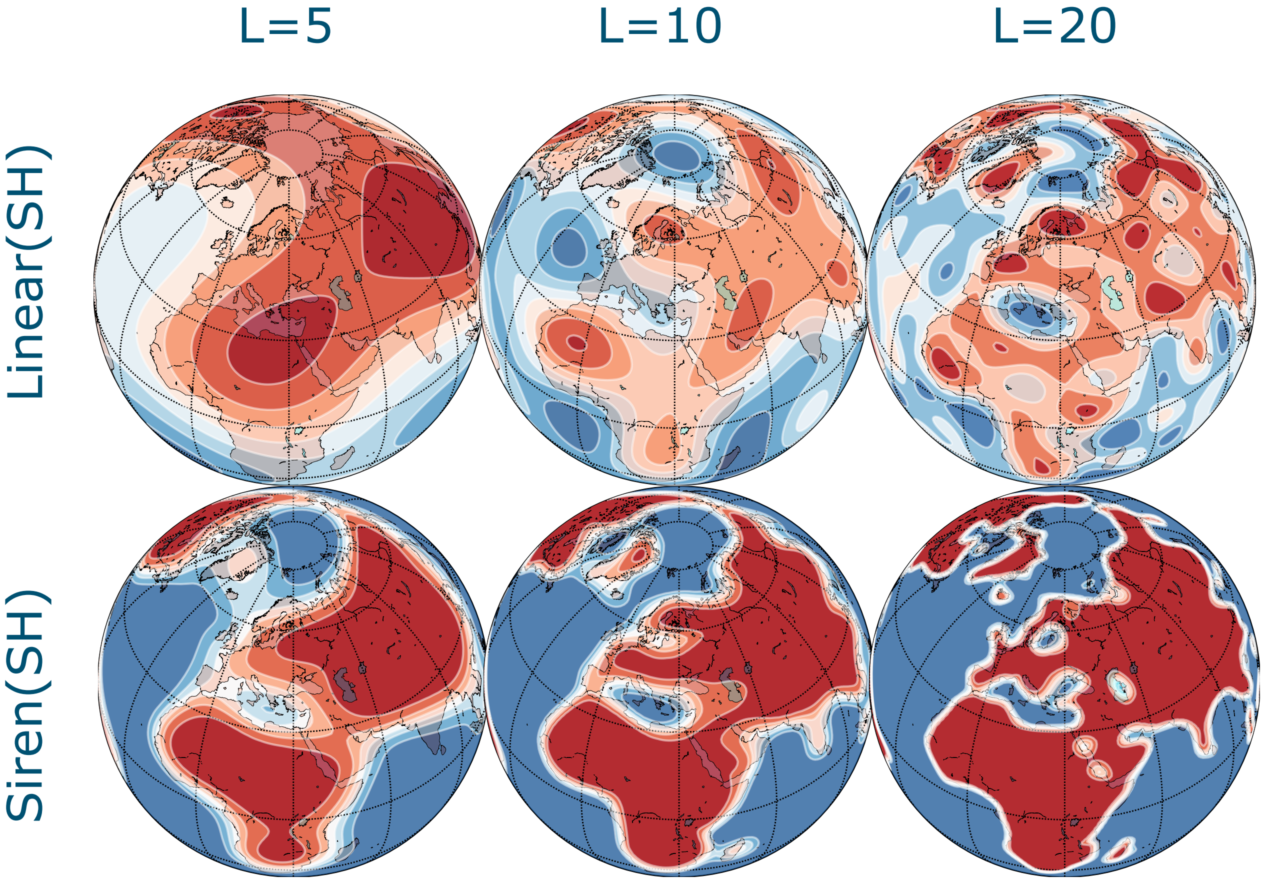


Controling Spatial Smoothness via L



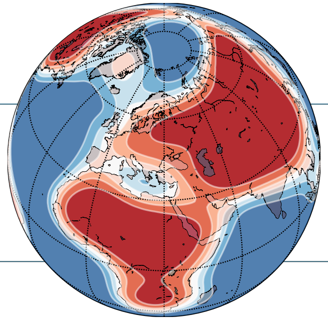


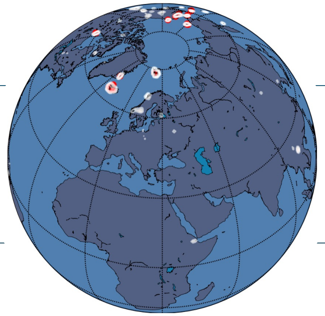




For Linear(SH): max resolution (in degree) $f_{\max} = \frac{180^\circ}{2L}$

Quantitative Results





Land-Ocean classification accuracy

PE ↓ NN →	Linear	FcNet	SirenNet
Direct	71.4 ± 0.0	90.3 ± 0.7	95.1 ± 0.3
Cartesian3D	70.5 ± 3.5	92.7 ± 0.3	92.8 ± 0.3
Wrap	74.4 ± 0.3	93.2 ± 0.3	95.2 ± 0.2
Grid	81.7 ± 0.1	95.1 ± 0.1	95.5 ± 0.2
Theory	86.9 ± 0.1	94.9 ± 0.2	95.5 ± 0.1
SphereC	79.6 ± 0.2	95.0 ± 0.3	95.2 ± 0.1
SphereC+	84.6 ± 0.2	95.3 ± 0.1	95.5 ± 0.1
SphereM	74.0 ± 0.0	89.1 ± 0.1	88.3 ± 0.4
SphereM+	81.9 ± 0.2	92.1 ± 0.3	93.7 ± 0.1
SH (ours)	$\underline{94.4 \pm 0.1}$	$\mathbf{95.9 \pm 0.1}$	$\underline{95.8 \pm 0.1}$

iNaturalist 2018 accuracy improvement

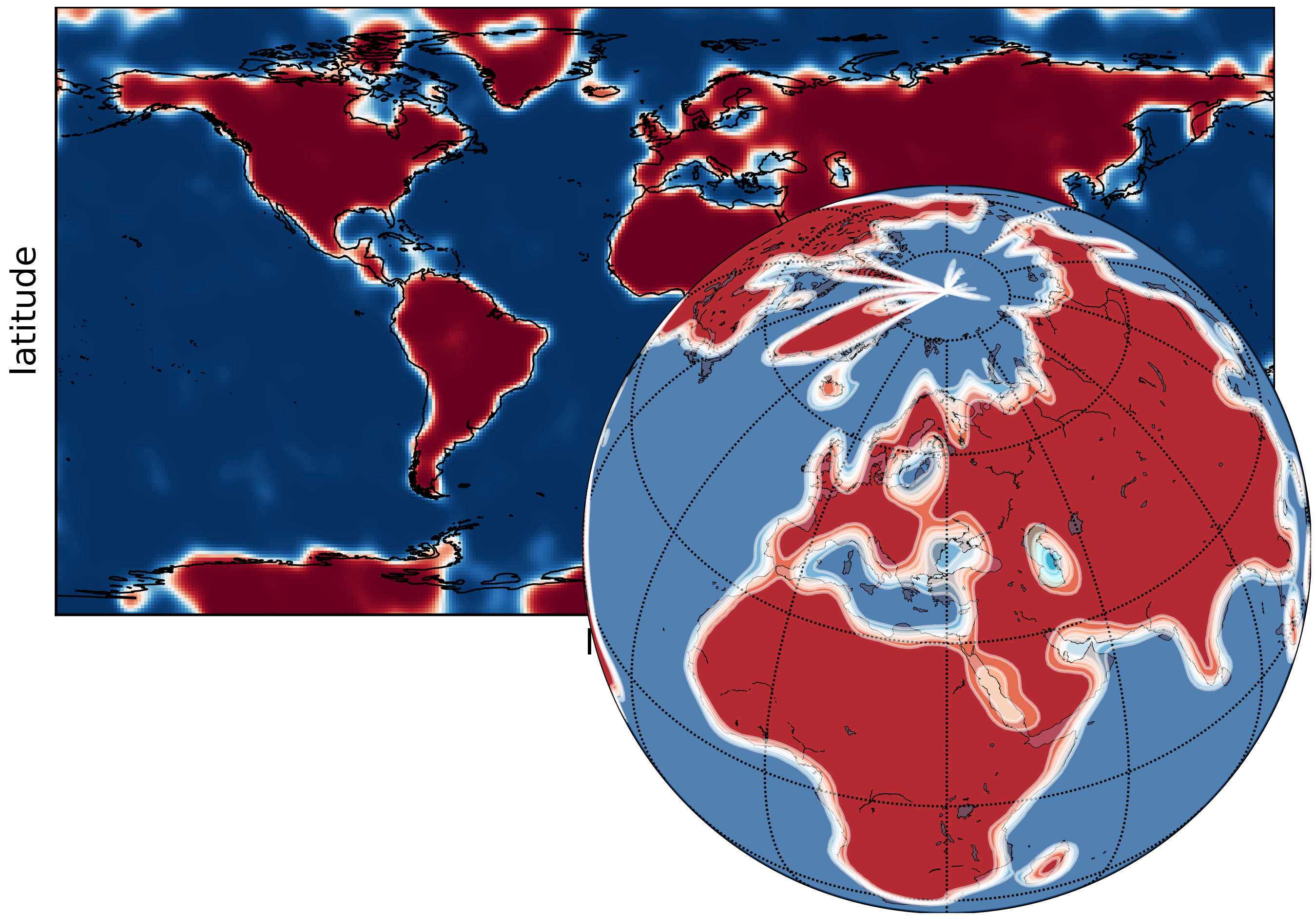
PE ↓ NN →	Linear	FcNet	SirenNet
Direct	-5.9 ± 0.1	$+9.3 \pm 0.3$	$+12.1 \pm 0.1$
Cartesian3D	$+0.8 \pm 0.2$	$+11.8 \pm 0.1$	$+12.0 \pm 0.1$
Wrap	-0.1 ± 0.1	$\underline{+12.1 \pm 0.1}$	$+12.1 \pm 0.1$
Grid	$+11.2 \pm 0.1$	$+11.8 \pm 0.2$	$+11.6 \pm 0.4$
Theory	$+11.5 \pm 0.0$	$+10.8 \pm 0.0$	$+11.4 \pm 0.1$
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)	$+10.5 \pm 0.1$	$+12.0 \pm 0.0$	$\mathbf{+12.3 \pm 0.2}$

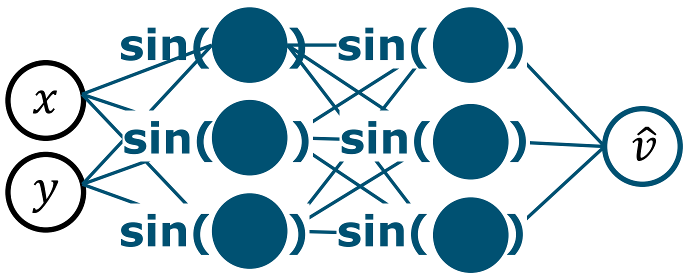
image-only: 59.2% top-1 accuracy with encoder NN(PE) ↑

Spherical Harmonics work well with all NNs

Siren is work well with all PEs

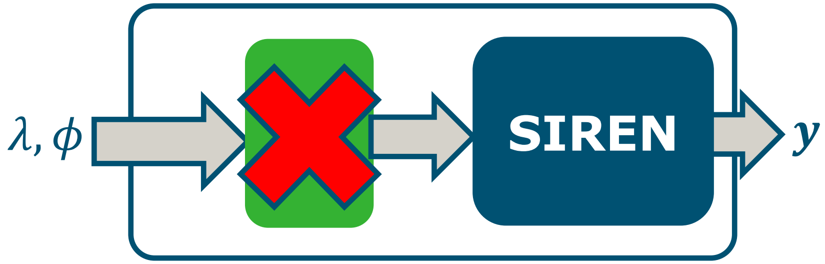
Siren works with without positional embedding (on moderate latitudes)



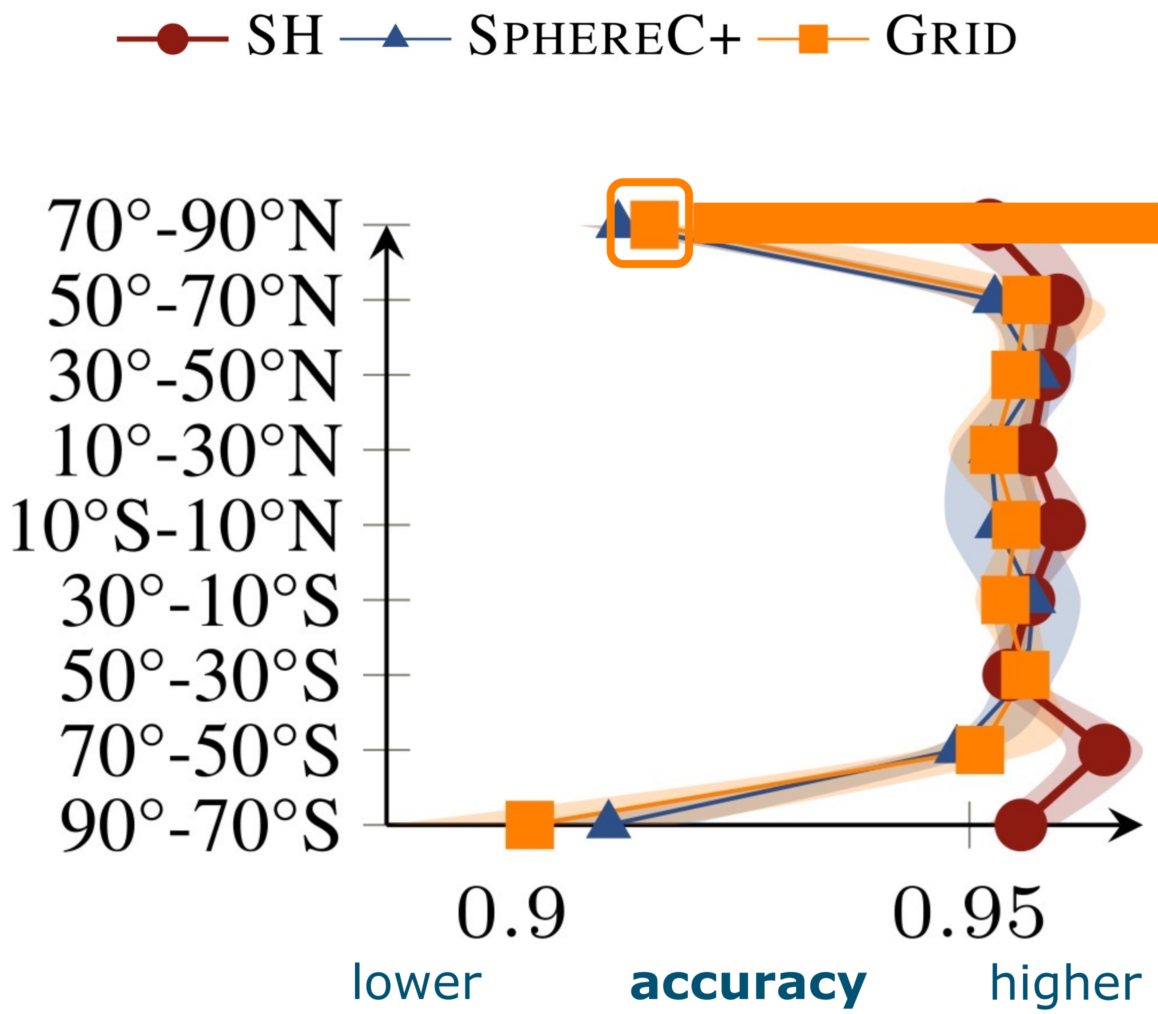


with Siren

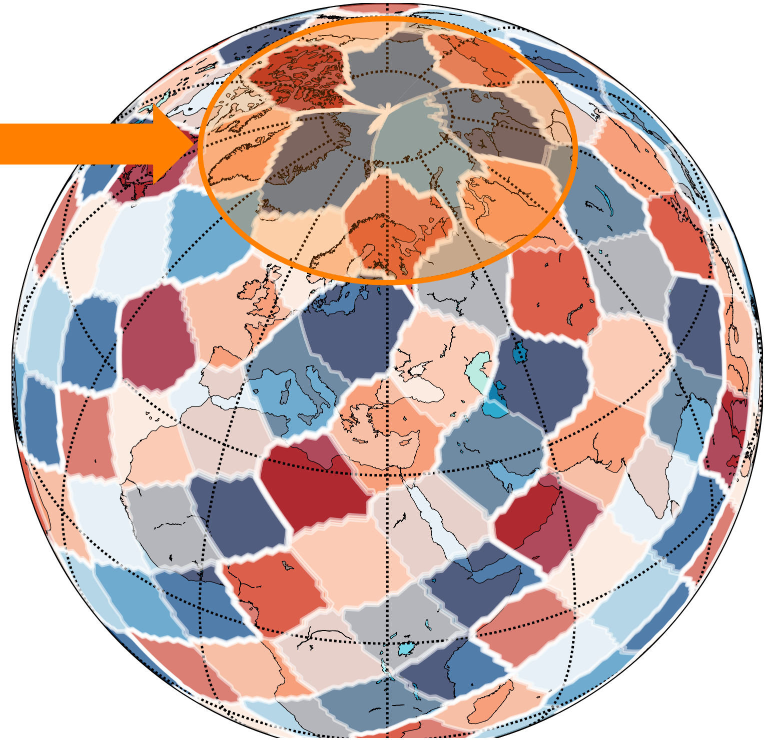
No positional embedding is needed (in moderate latitudes)



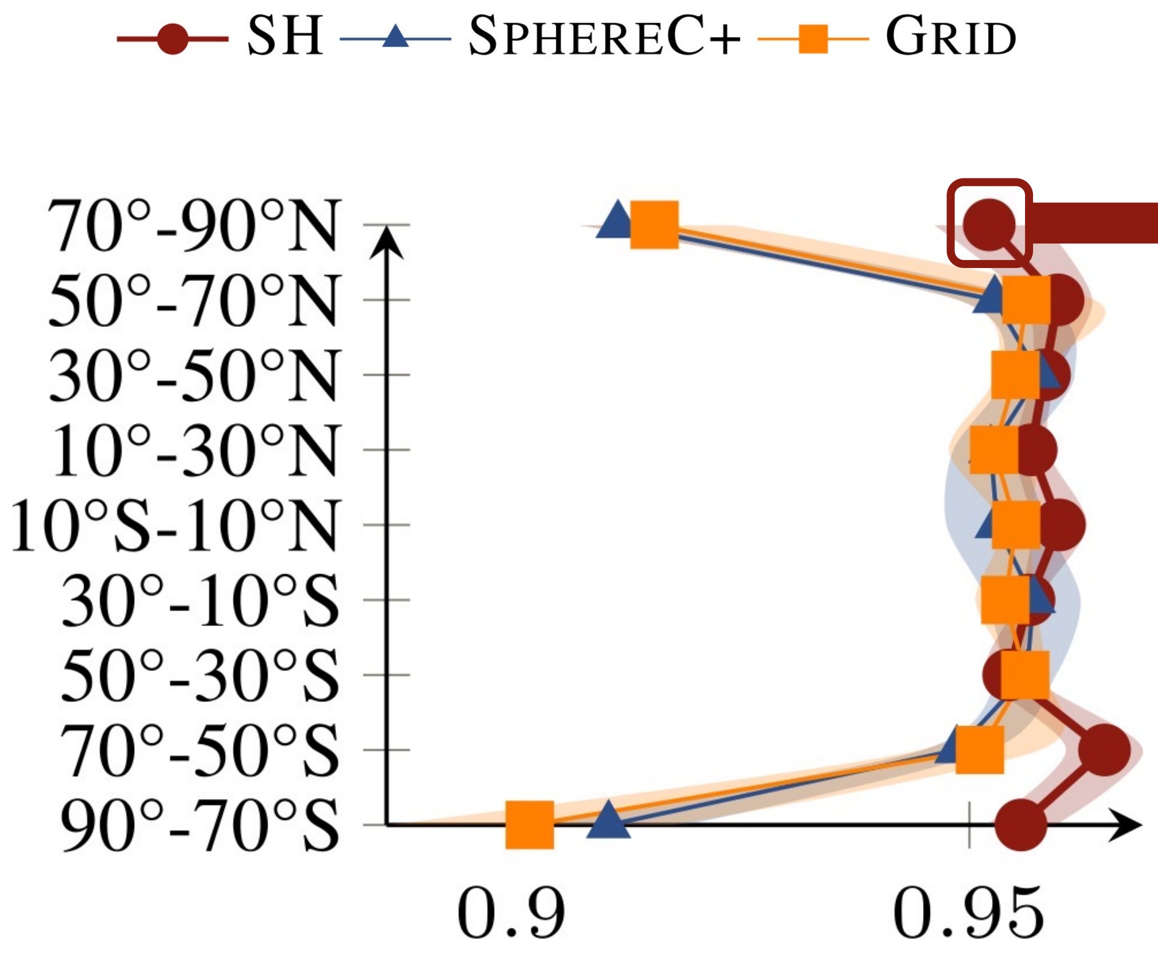
Longitudinal Accuracy – Checkerboard Classification



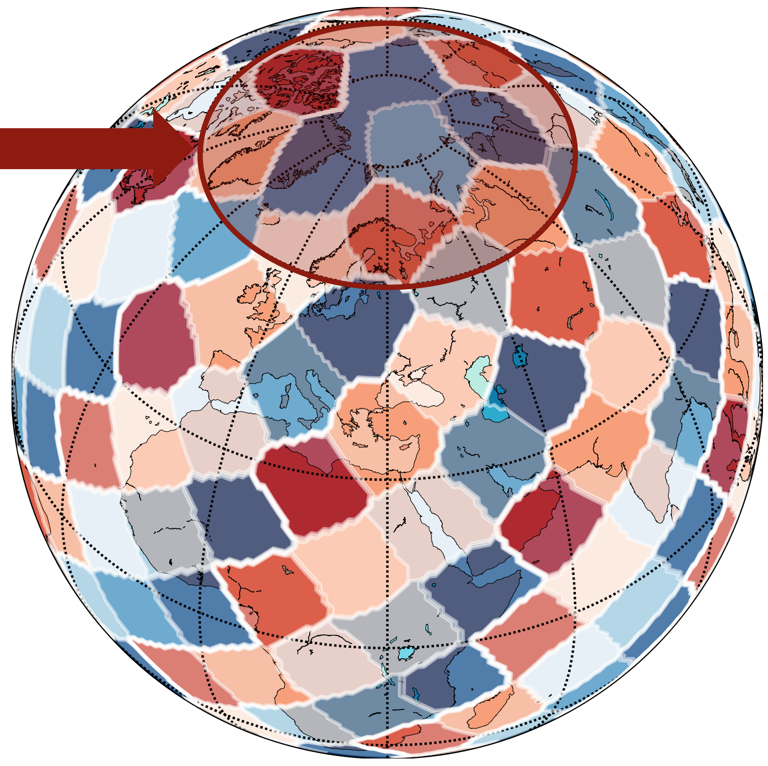
Siren(Grid)



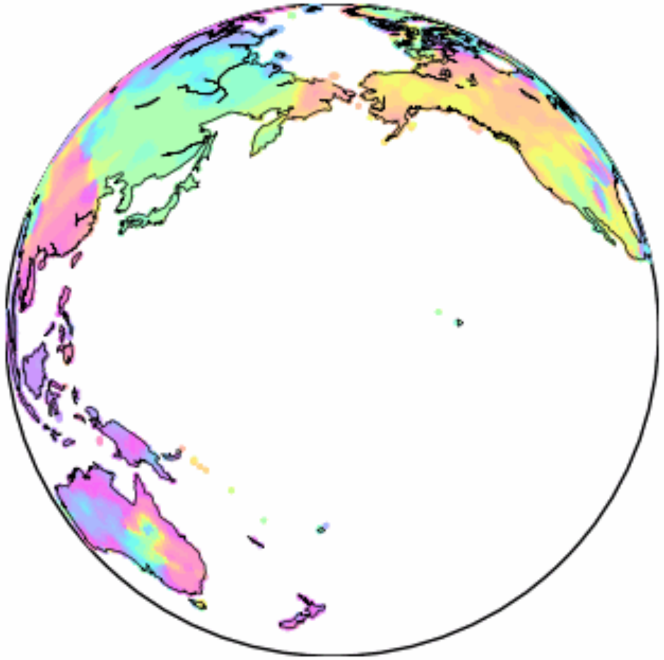
Longitudinal Accuracy – Checkerboard Classification



Siren(SH)



A location encoder for a
general representation of location
(according to Satellite images)



SatCLIP: Global, General-Purpose Location Embeddings with Satellite Imagery

https://arxiv.org/abs/2311.17179

Konstantin Klemmer, Esther Rolf, Caleb Robinson, Lester Mackey, Marc Rußwurm

Next Step – learn a descriptive vector of any location λ, ϕ

Previous Part: Location Encoding

how to store spatial data in a location encoder representation?

how can we semantically describe a location?

1. with Siren(SH) as location encoder
2. strong-supervised loss function like cross-entropy/mean squared error

we always need labelled training/support data

technically the same application field as classic methods like interpolation, Gaussian Processes/Kriging

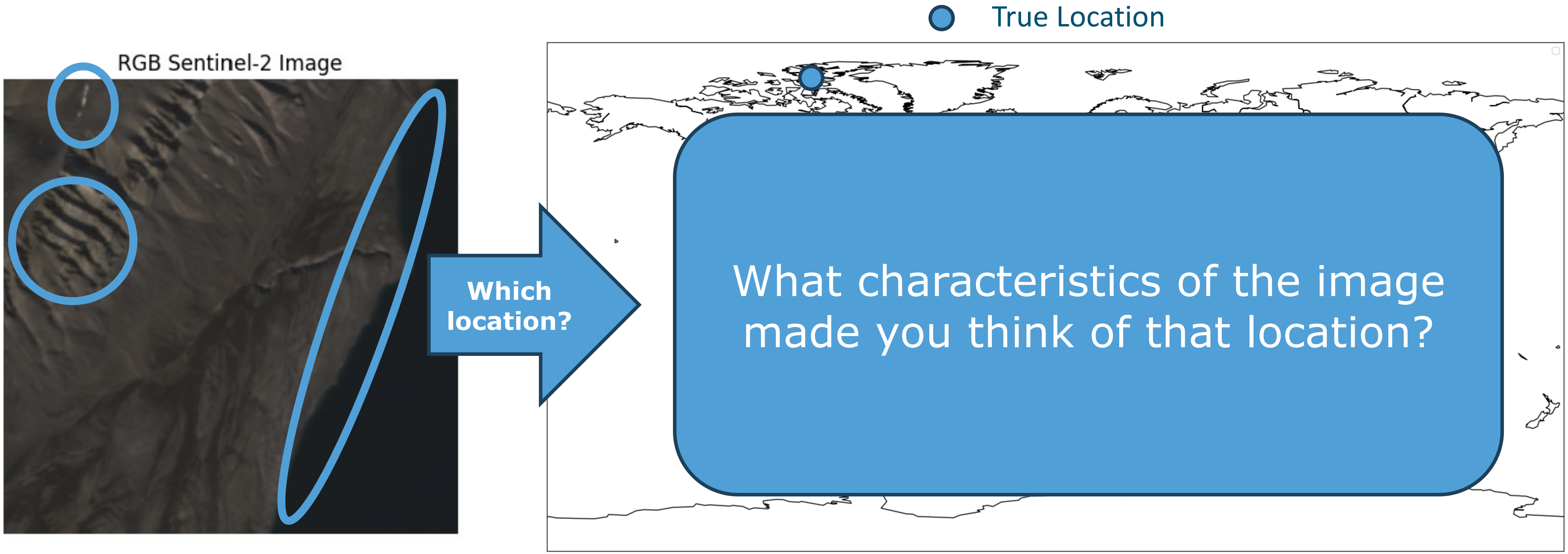
1. with Siren(SH) as location encoder
2. trained with a **loss function that defines what we mean by „semantic description"**

Geolocalization (i.e., Geoguesser) as contrastive pretext task

from above:

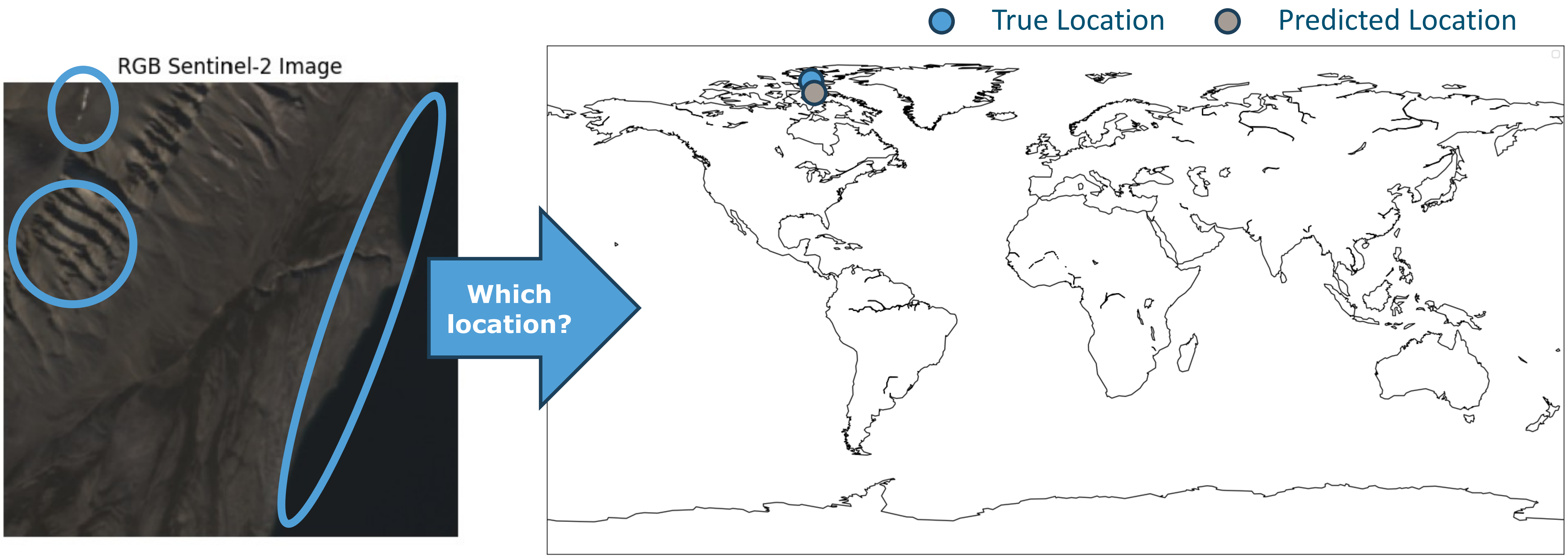
Geolocation pre-text task to extract location description

But how can we **pretrain location encoders without labels**?

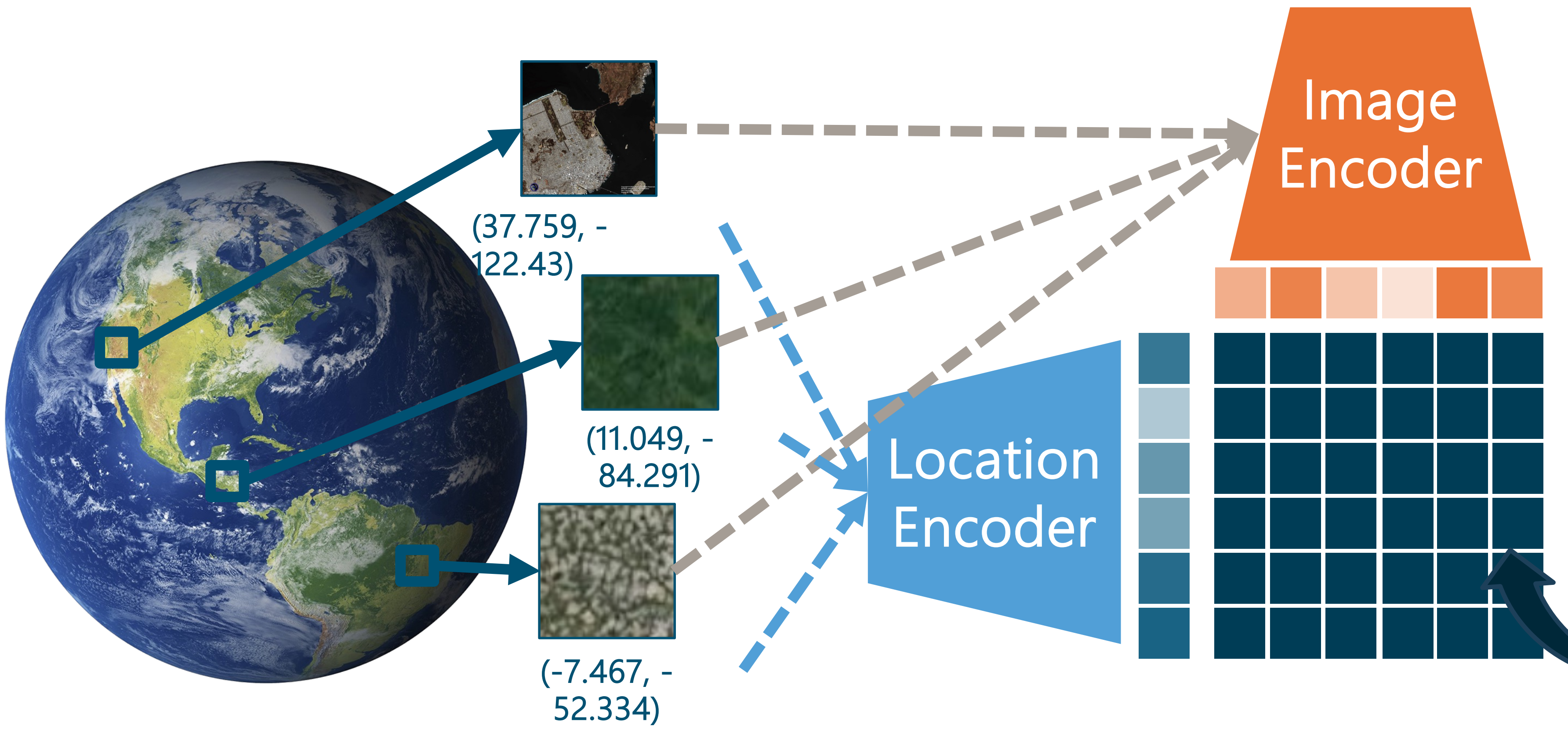


Geolocation pre-text task to extract location description

But how can we **pretrain location encoders without labels**?



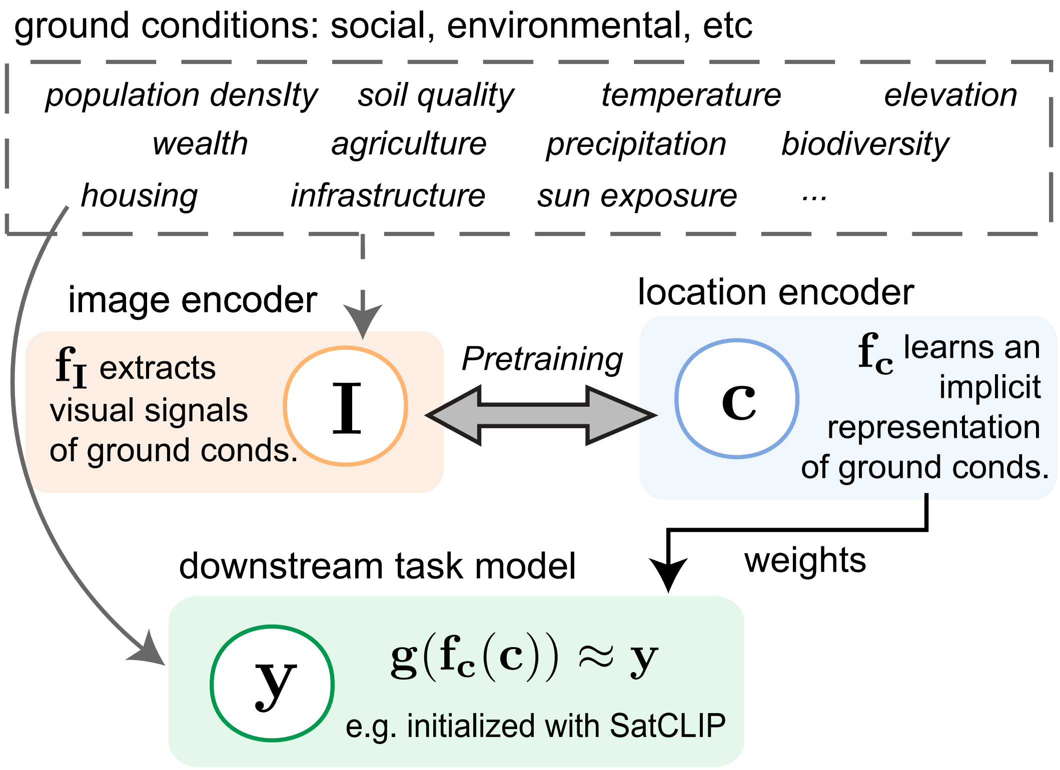
Implementation as Contrastive Location-Image Pretraining

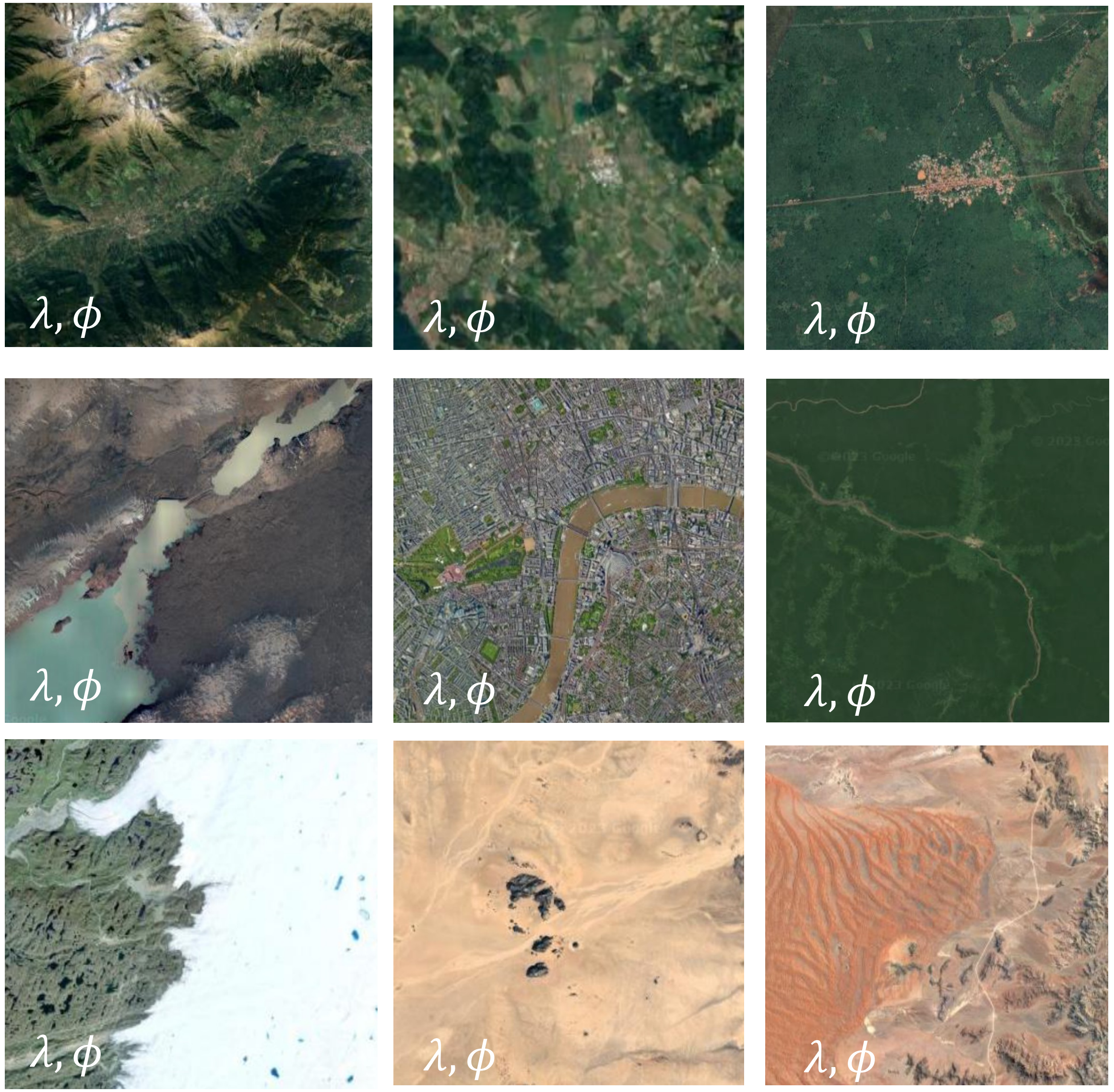


Contrastive learning: match image and location embeddings.

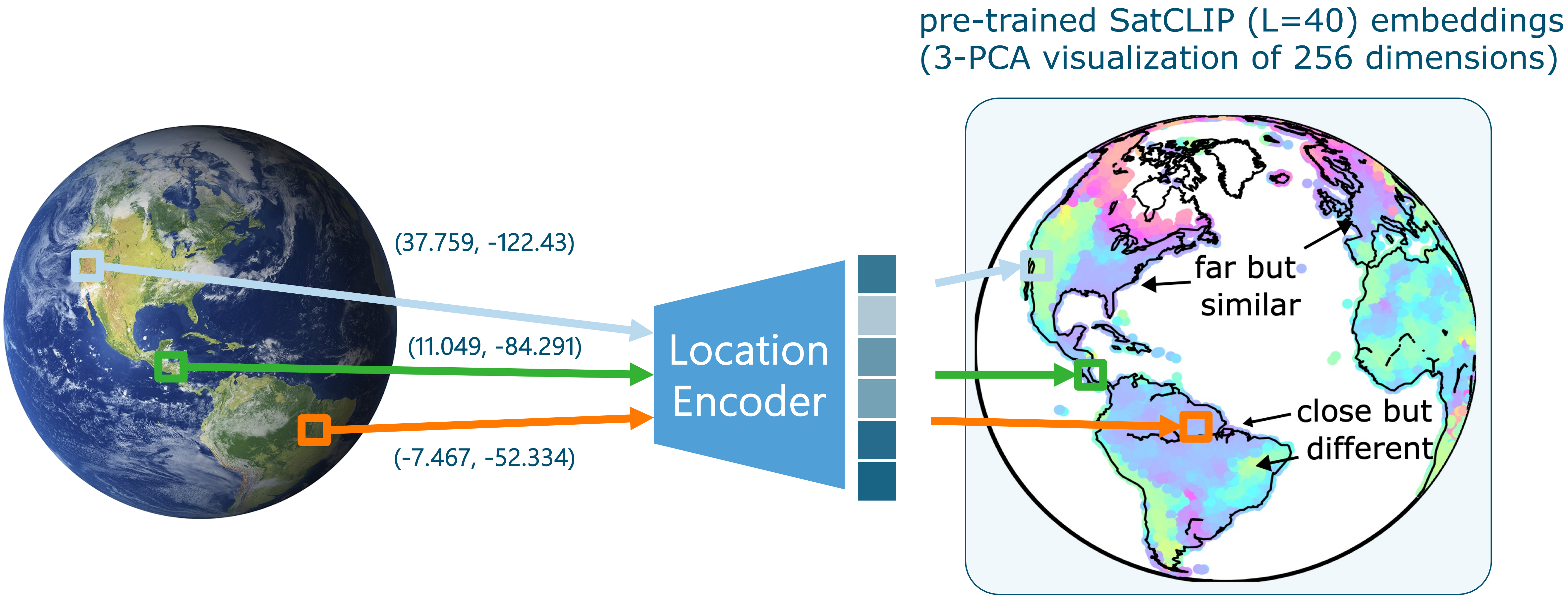
Same training objective as in Contrastive Language-Image Pretraining (CLIP)
Radford et al., 2021 Learning transferable visual models from natural language supervision. ICML

Intuition behind SatCLIP: distill location-specific patterns

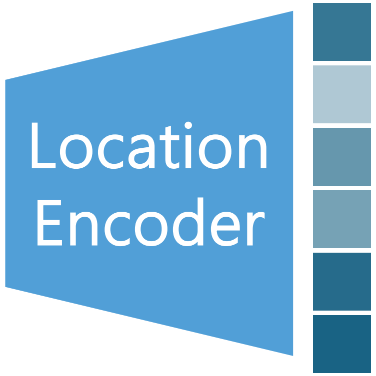


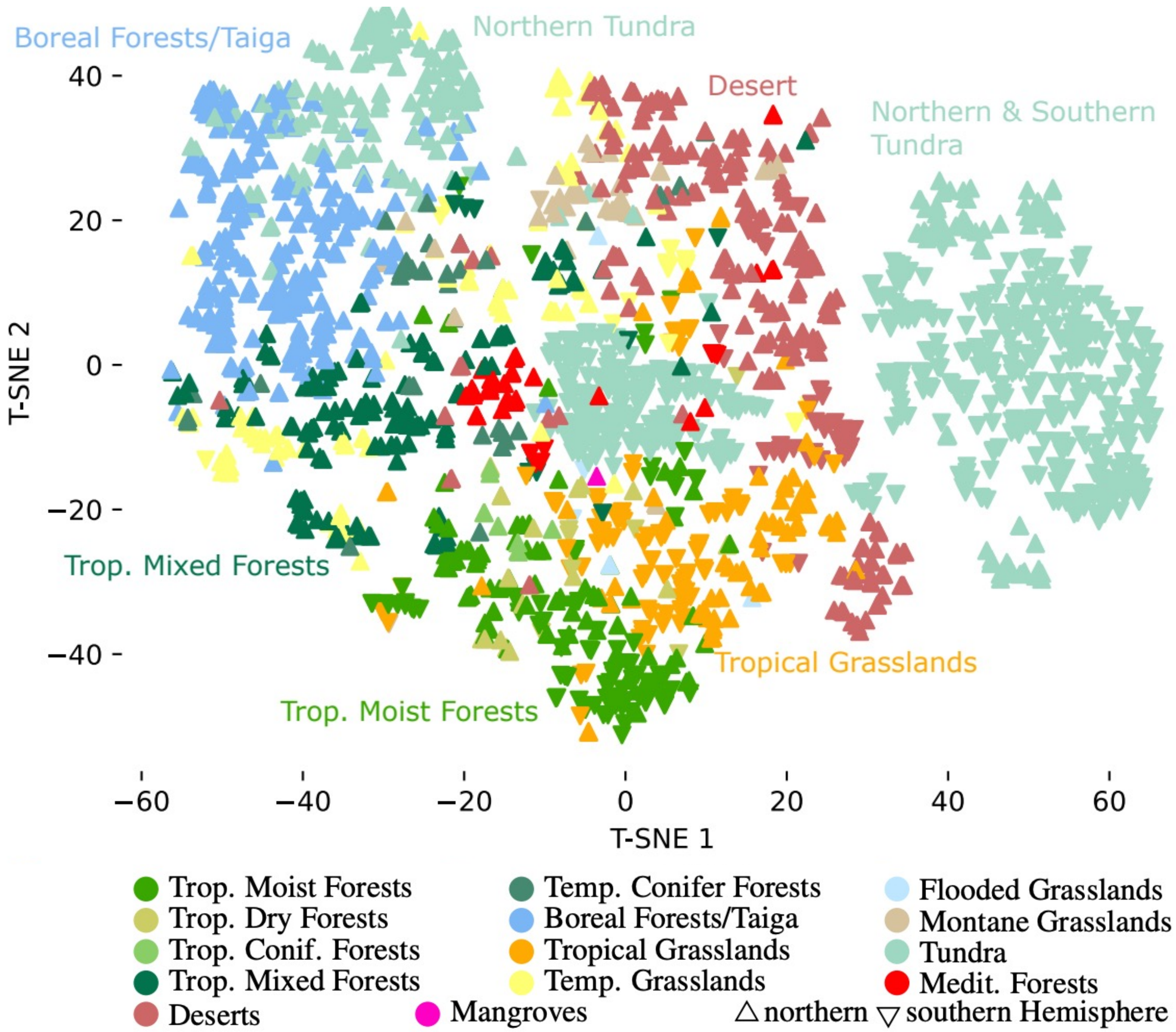


Light-weight Implicit Neural Geo-representation (INGR)



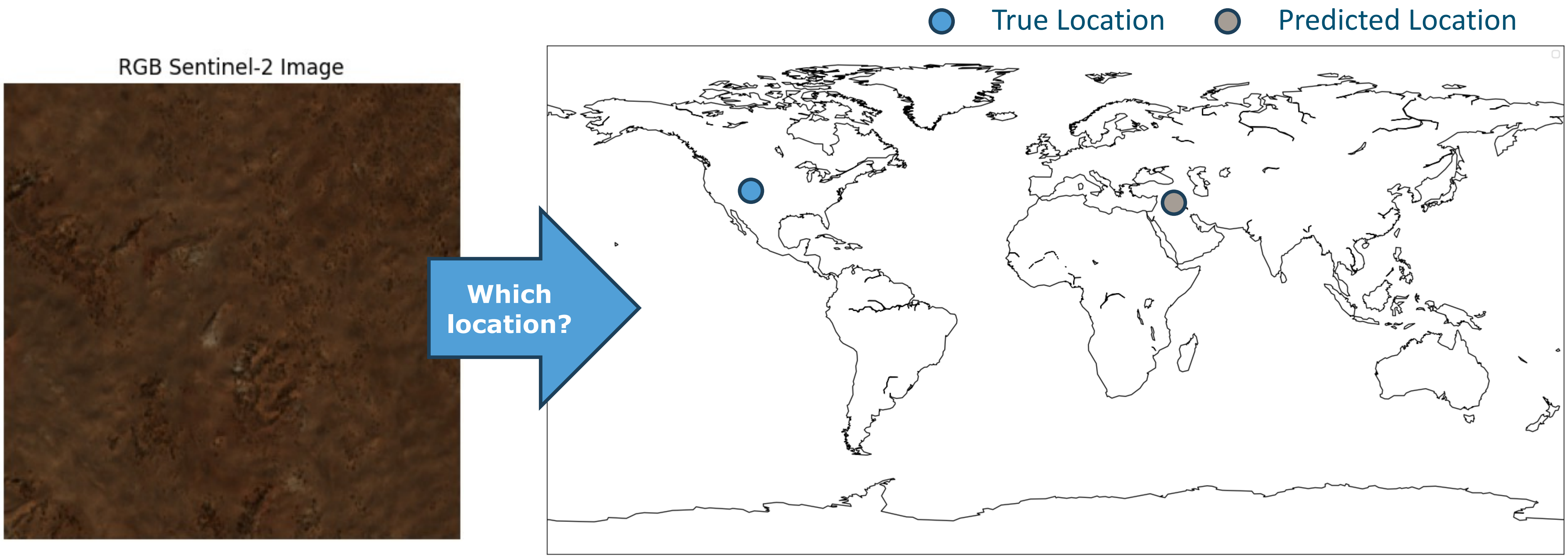
SatCLIP embeddings cluster-well with Biomes





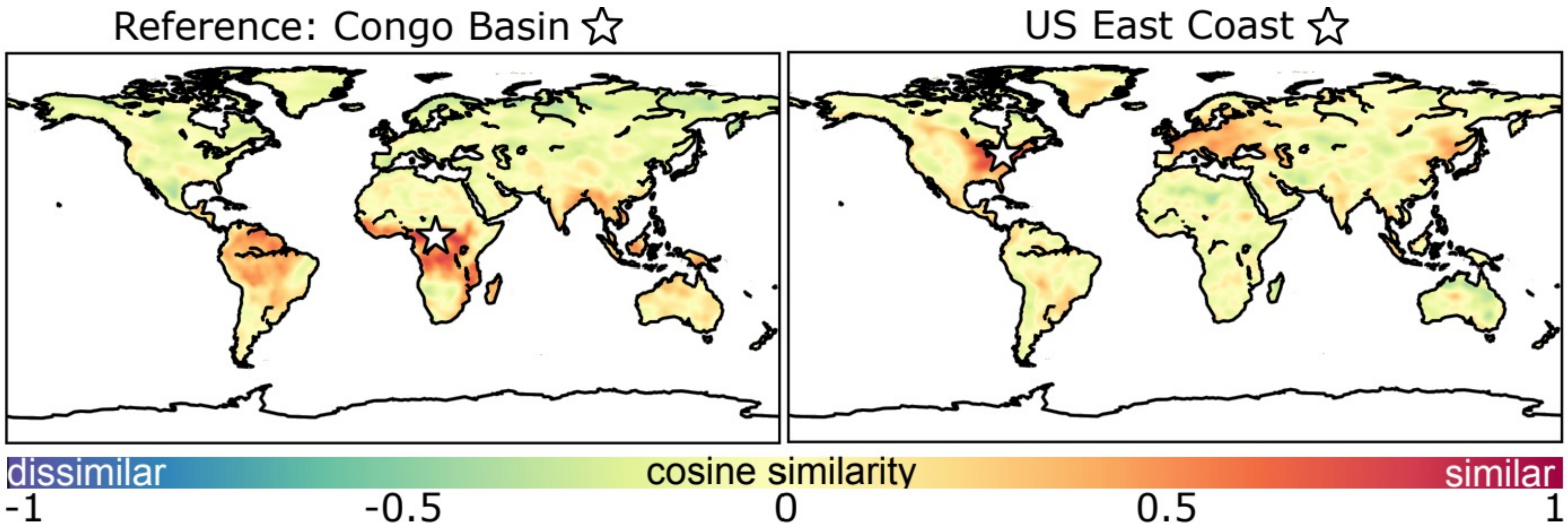
Spatial similarity leads to confusions!

If locations in different parts of the world have similar images, SatCLIP **aligns the embedding space**.

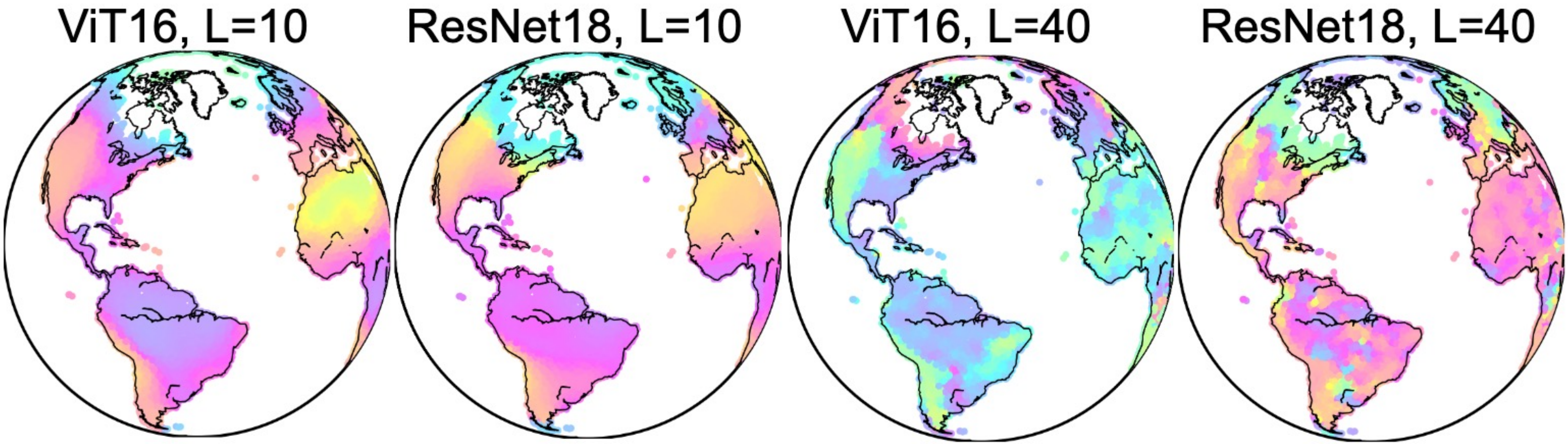


Spatial Similarity – Challenging Tober’s rule

By learning similarities between global locations based on visual similarity



Pre-trained models at two “smoothnesses” L10, L40




```
SatCLIP-ResNet18-L10: wget https://satclip.z13.web.core.windows.net/satclip/satclip-resnet18-l10.ckpt
SatCLIP-ResNet18-L40: wget https://satclip.z13.web.core.windows.net/satclip/satclip-resnet18-l40.ckpt
SatCLIP-ResNet50-L10: wget https://satclip.z13.web.core.windows.net/satclip/satclip-resnet50-l10.ckpt
SatCLIP-ResNet50-L40: wget https://satclip.z13.web.core.windows.net/satclip/satclip-resnet50-l40.ckpt
SatCLIP-ViT16-L10: wget https://satclip.z13.web.core.windows.net/satclip/satclip-vit16-l10.ckpt
SatCLIP-ViT16-L40: wget https://satclip.z13.web.core.windows.net/satclip/satclip-vit16-l40.ckpt
```

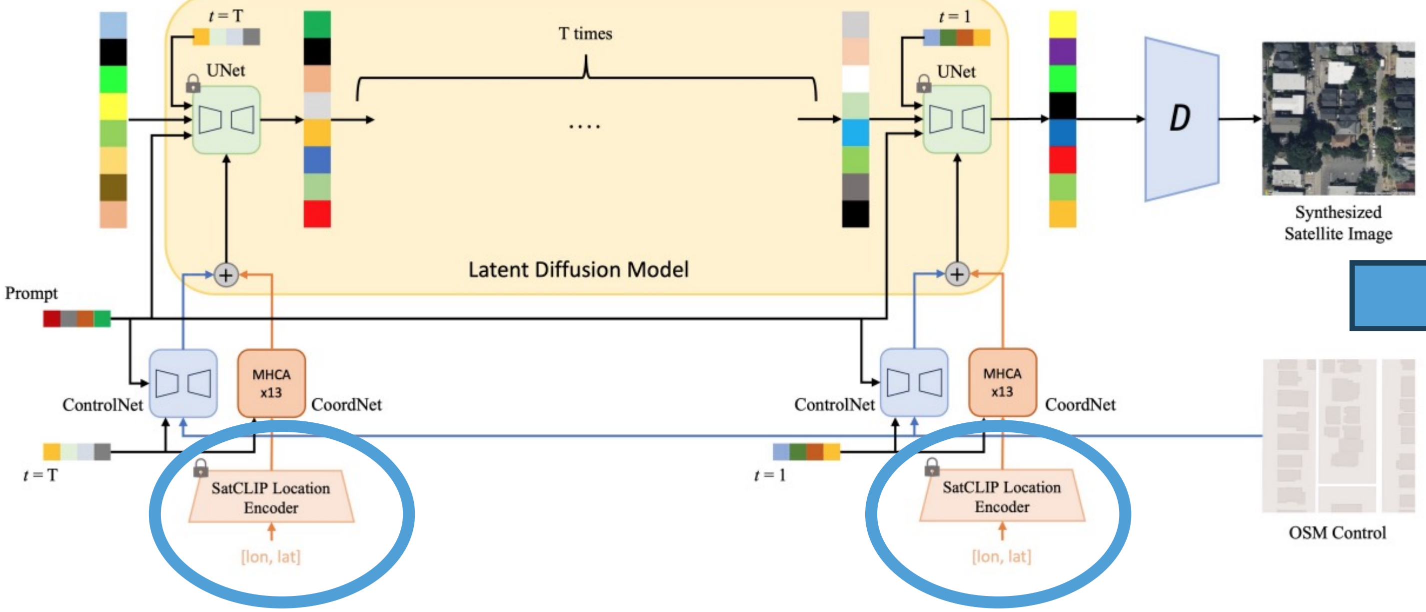
More implementation info on https://github.com/microsoft/satclip

Usages of SatCLIP – Microsoft Bing Maps AI Team

Models trained on Western countries fail to segment buildings in Africa.

Usages of SatCLIP – GeoSynth

Researchers are already **adapting SatCLIP**! For example, to guide diffusion models:



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

Ongoing Master Thesis – Disease Mapping in Spain

SatCLIP embedding as a proxy for environmental variables (visible in satellite images) for disease modeling.

WAGENINGEN UNIVERSITY & RESEARCH

upna
Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa

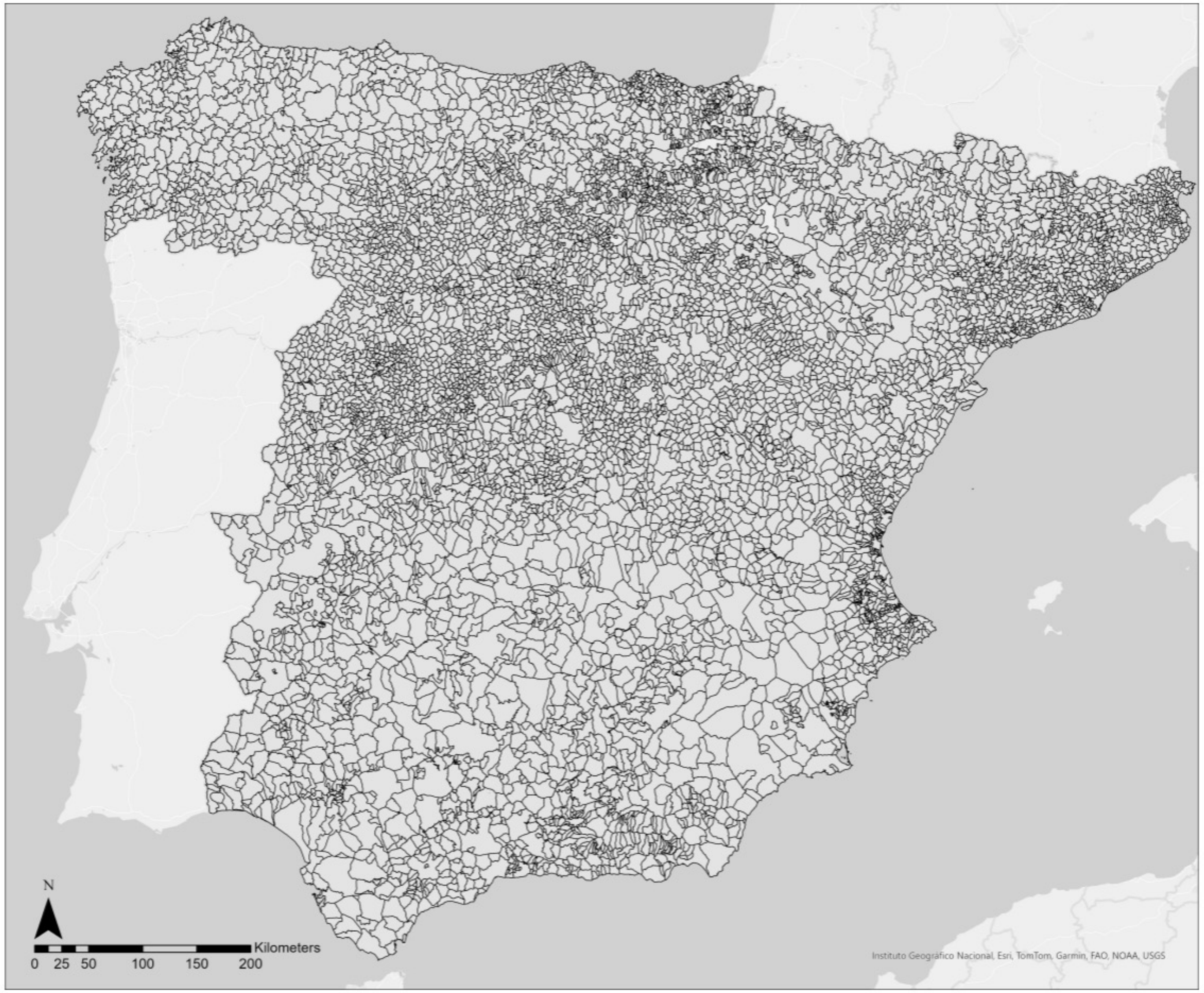
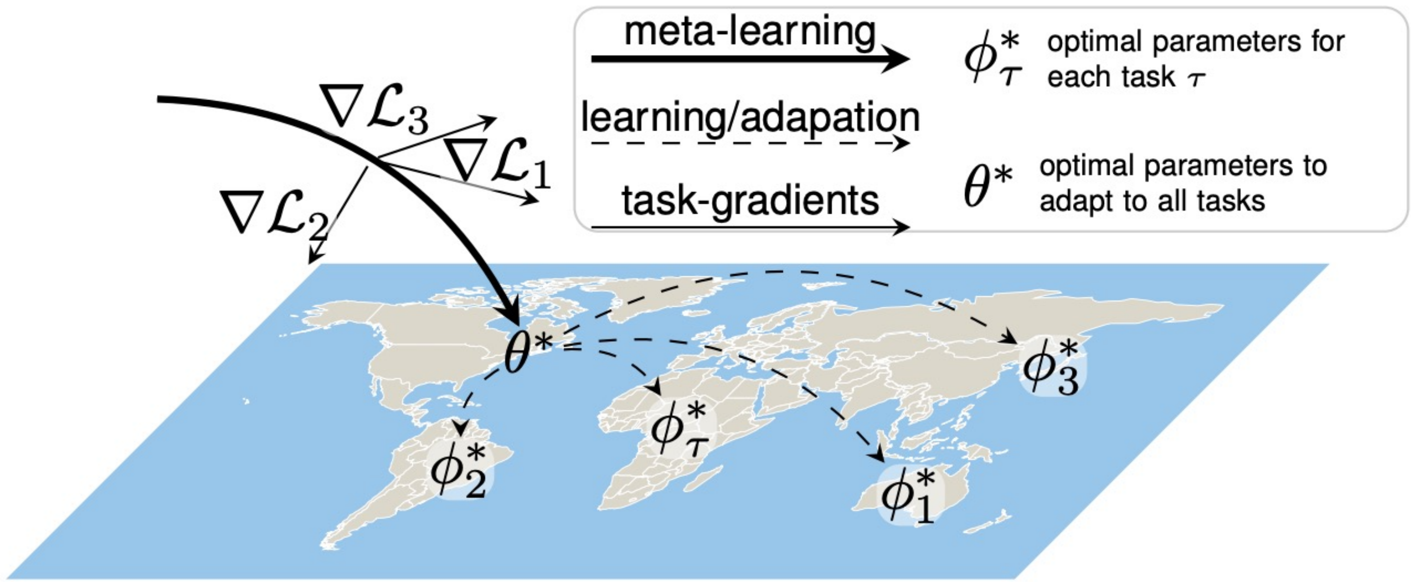
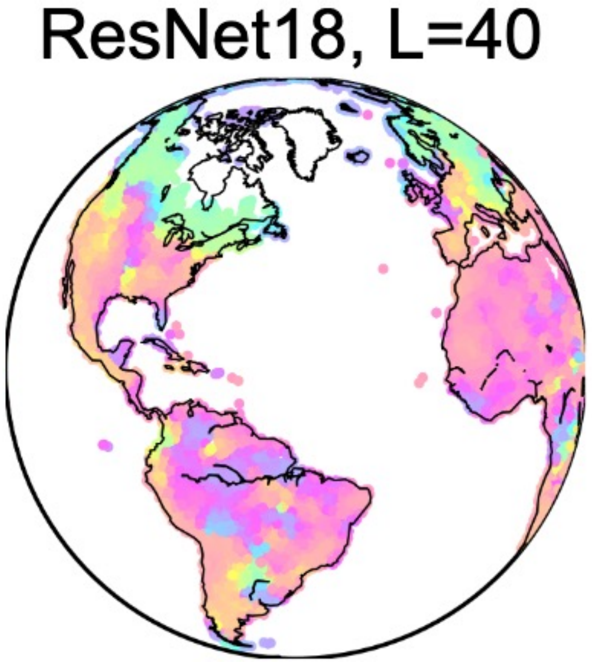


Figure 1. The Spanish municipalities[27] used for this research.

Location-specific calibration of classification MOdels





Rußwurm, M., Wang, S., Korner, M., & Lobell, D. (2020). Meta-learning for few-shot land cover classification. In Proceedings of the ieee/cvf conference on computer vision and pattern recognition workshops (pp. 200-201).

Rußwurm, M., Wang, S., Kellenberger, B., Roscher, R., & Tuia, D. (2024). Meta-learning to address diverse Earth observation problems across resolutions. Communications Earth & Environment, 5(1), 37.

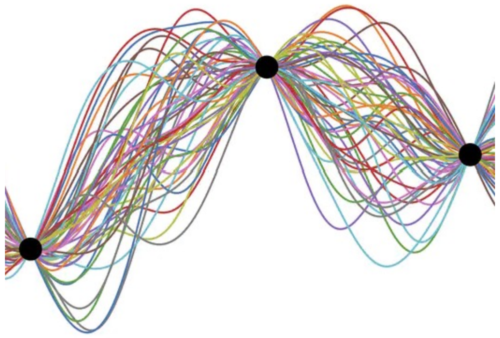
Recap & Takeaways - 1

1 Spatial Modeling meets Implicit Neural Representations

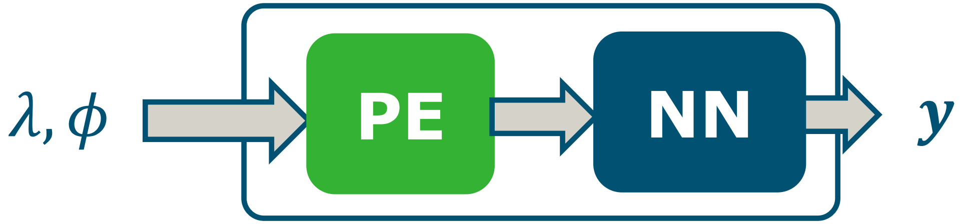
We **learn a continuous function over space**:

1. with classic methods like interpolation and/or Kriging
2. by fitting/learning a neural network to reproduce the data

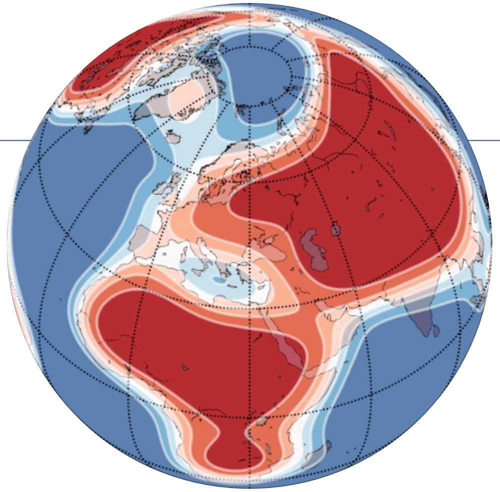
Gaussian Process/Kriging



Location Encoder Network



Recap & Takeaways - 2



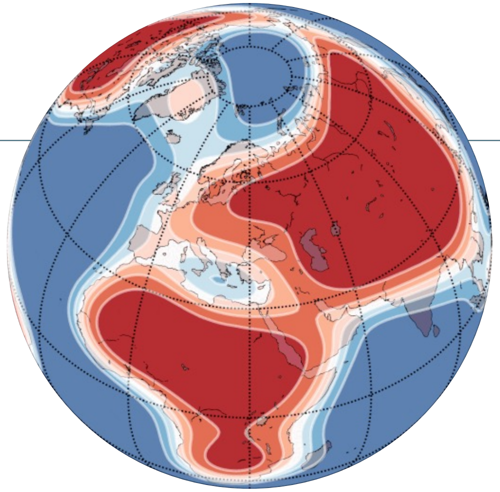
1 Spatial Modeling meets Implicit Neural Representations

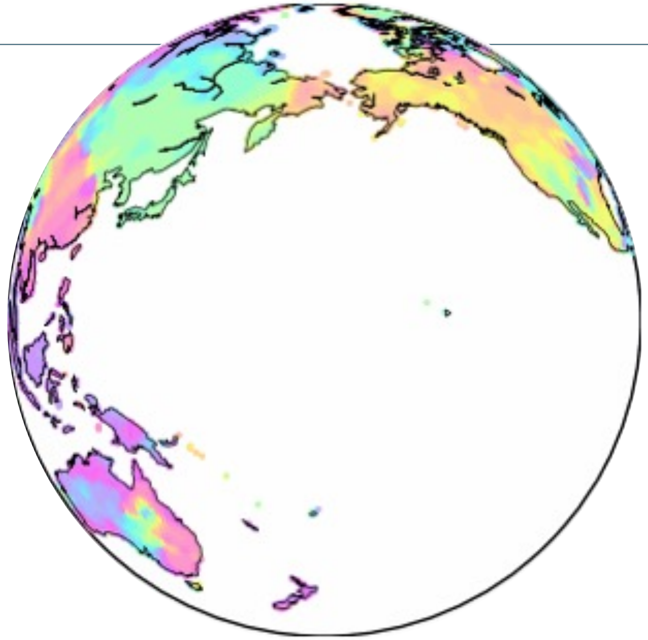
2 Siren(SH) Location Encoder
Rußwurm et al., ICLR 2024

As location encoder, we 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 matter

Recap & Takeaways - 3





1 Spatial Modeling meets Implicit Neural Representations

2 Siren(SH) Location Encoder
Rußwurm et al., ICLR 2024

3 SatCLIP Encoder
Klemmer et al., 2024 –
In submission

SatCLIP: learning a

- **continuous spatial representation** of **visual similarity**
- by training a location encoder on geolocation (i.e., playing Geoguesser)

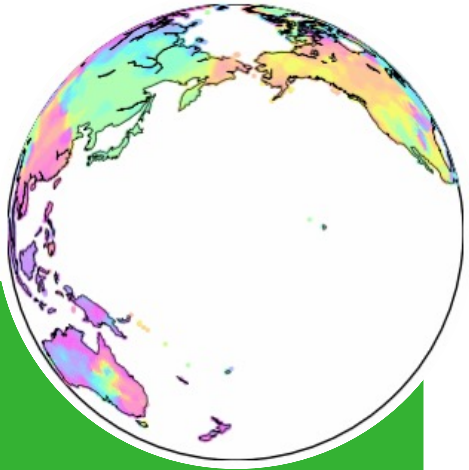
Thank you & happy to take questions!

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

SatCLIP:
Klemmer et al., 2024

https://github.com/microsoft/satclip
https://arxiv.org/pdf/2311.17179.pdf

Watch 14 | Fork 22 | Star 227



Marc Rußwurm,
marc.russwurm@wur.nl

Assistant Professor
Machine Learning & Remote Sensing

Wageningen University, Netherlands

Collaborators:
Konstantin Klemmer, Esther Rolf, Robin Zbinden, Devis Tuia, Caleb Robinson, Lester Mackey

WAGENINGEN UNIVERSITY & RESEARCH | HARVARD UNIVERSITY | EPFL | Microsoft