TorchSpatial: A Location Encoding Framework and Benchmark for Spatial Representation Learning

NeurIPS 2024

Nemin Wu, Qian Cao*, Zhangyu Wang, Zeping Liu, Yanlin Qi, Jielu Zhang, Joshua Ni, X. Angela Yao, Hongxu Ma, Lan Mu, Stefano Ermon, Tanuja Ganu, Akshay Nambi, Ni Lao, Gengchen Mai†*

⁎ Both authors contributed equally to this research
† Corresponding author

UNIVERSITY OF GEORGIA 1785

SPATIALLY EXPLICIT AI

TEXAS The University of Texas at Austin

Motivation

Spatial representation learning (SRL) is a set of technique learning neural spatial representations from spatial data in their native formats while avoiding manual feature engineering or data conversion.

1. **Lack of a community-shared framework** for SRL model development
2. **Lack of benchmarks** to systematically evaluate the location encoders' impact on model performance
3. **Lack of evaluations on the geographic bias** of geo-aware AI approaches

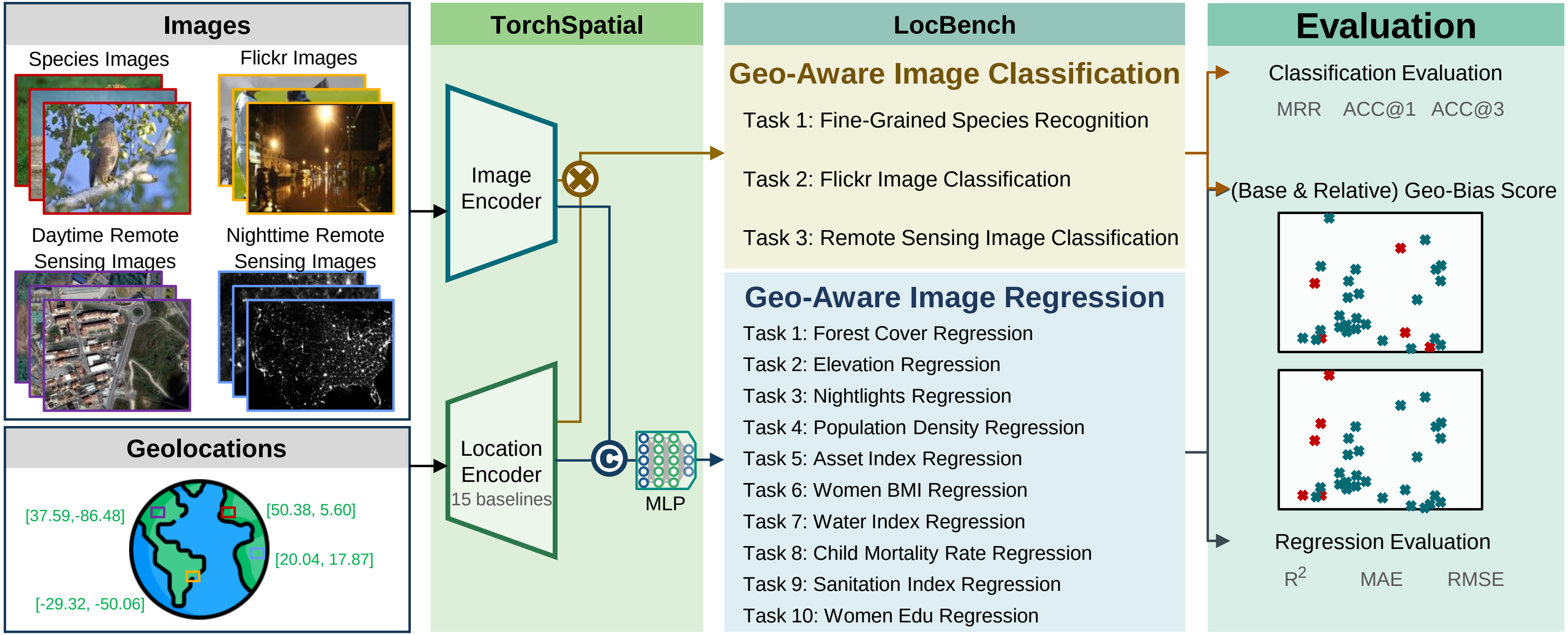
TorchSpatial

- **A model framework** that consolidates **15 location encoders** and necessary model building blocks for future location encoders.

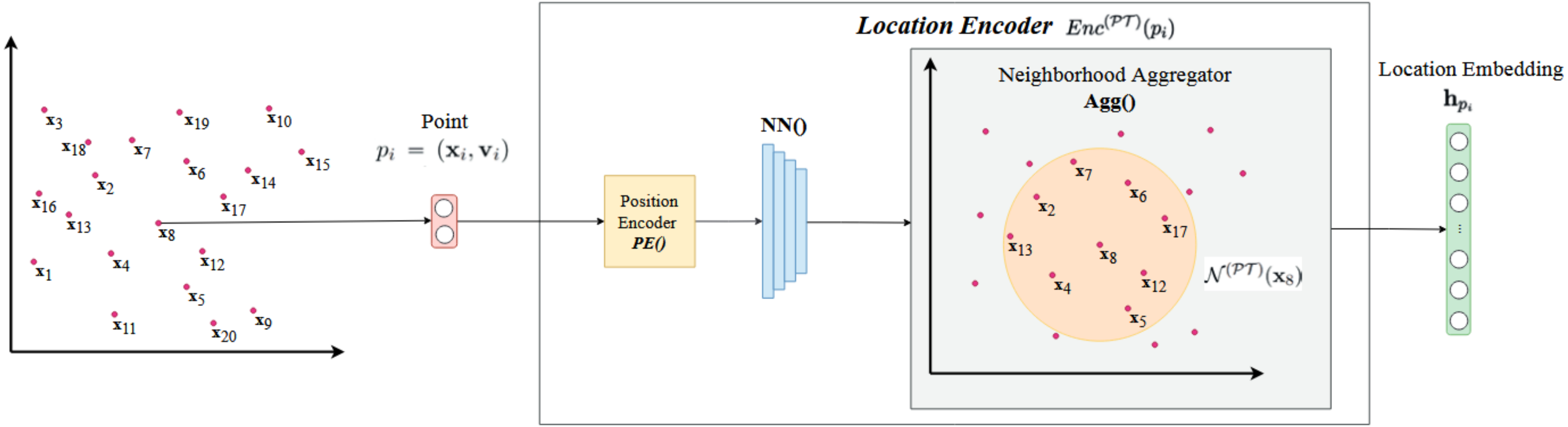
- **A benchmark** which contains 7 geo-aware image classification and 10 image regression datasets.

- **A set of evaluation metrics** to quantify location encoders' overall model performance and their **geographic bias**.

TorchSpatial Framework



TorchSpatial Framework – Location Encoder



$$\boxed{Enc(\mathbf{x}) = \mathbf{NN}(PE(\mathbf{x}))}$$

$\mathbf{x} \in \mathbb{R}^L \; (L = 2, 3)$ *: input location*

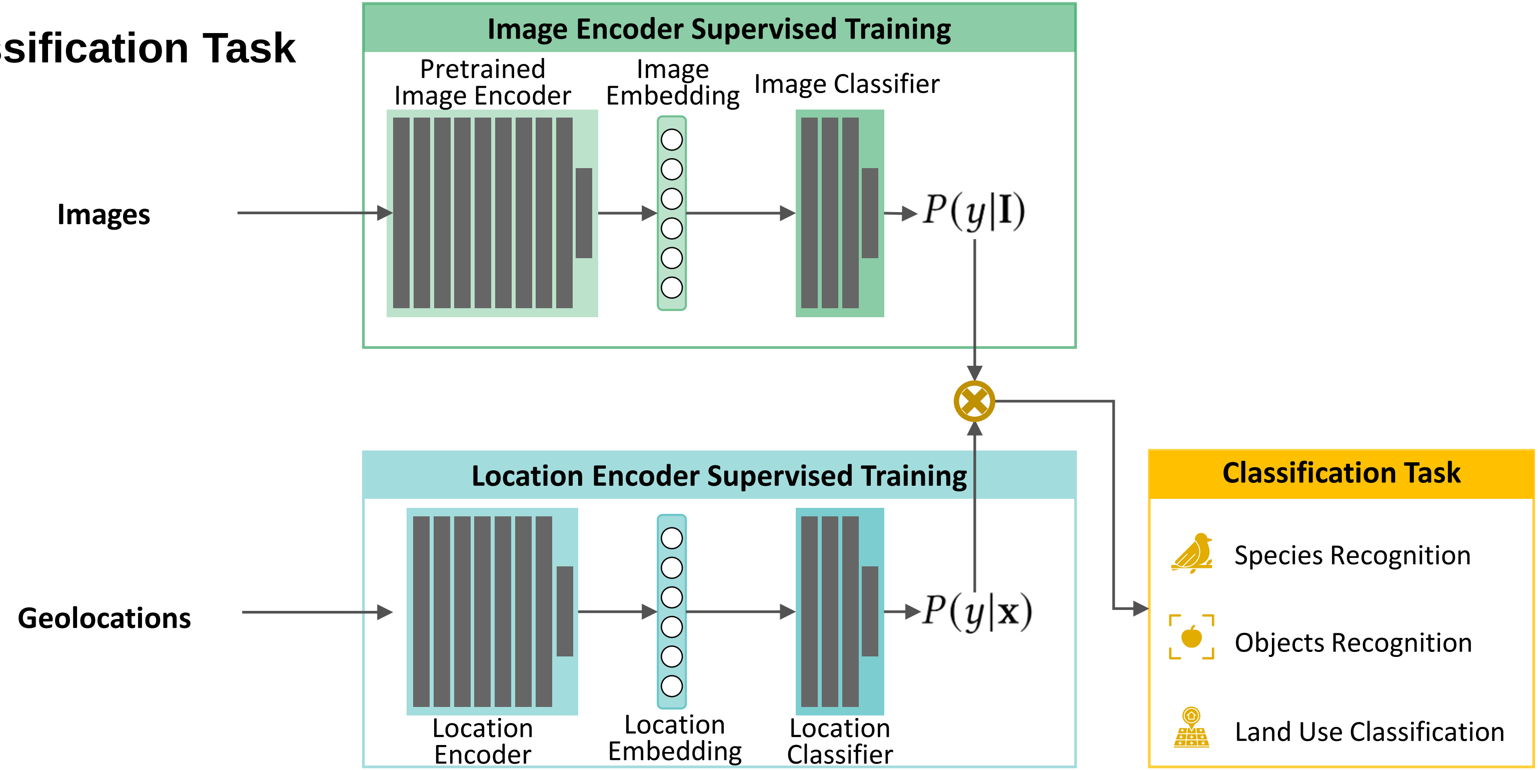
$PE(\mathbf{x}) \in \mathbb{R}^W$ *: position encoder*

$\mathbf{NN}(\cdot) : \mathbb{R}^W \rightarrow \mathbb{R}^d$ *: learnable neural nets*

Mai, Gengchen, et al. "A review of location encoding for GeoAI: methods and applications." *International Journal of Geographical Information Science*

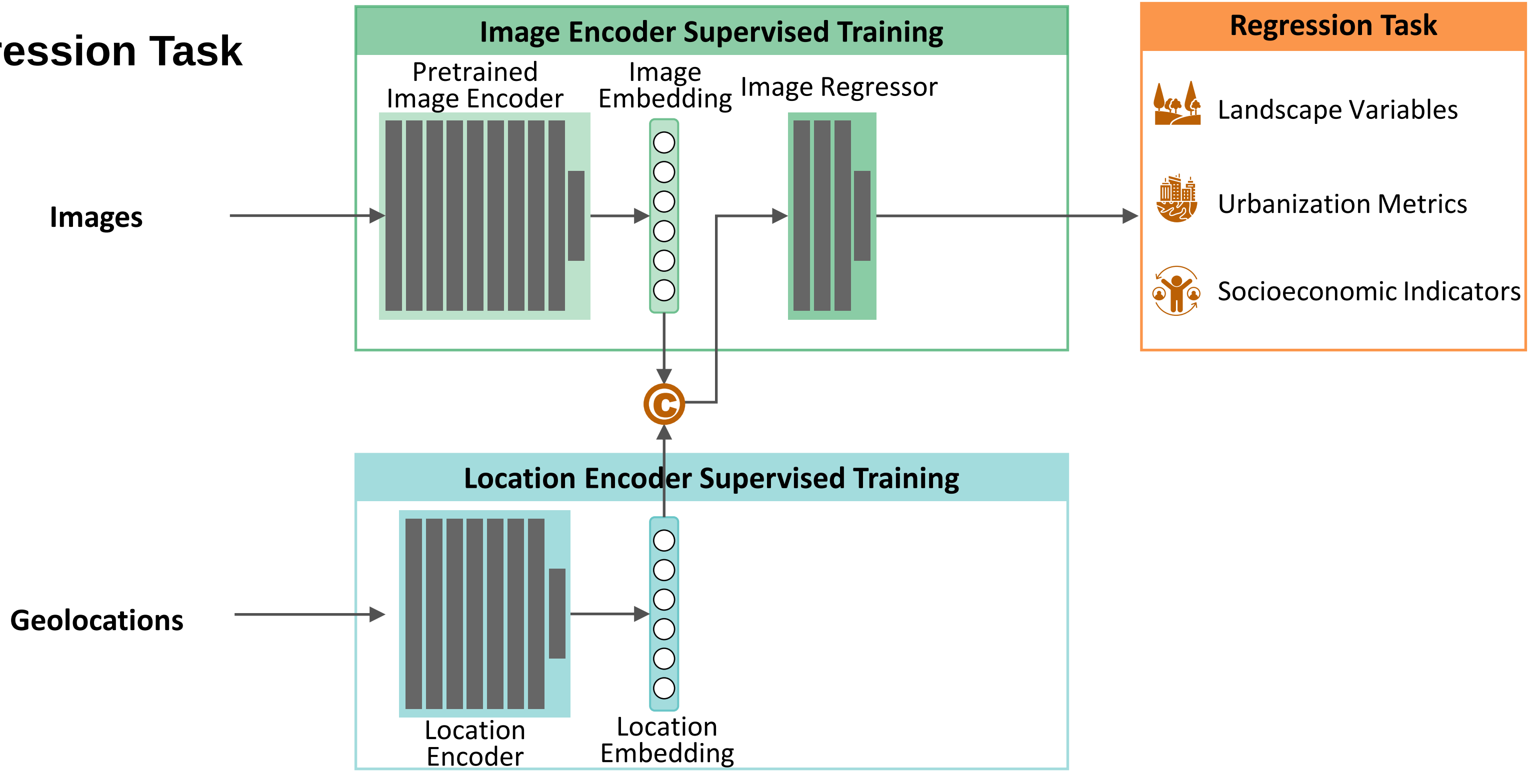
TorchSpatial Framework – Inference Architecture

- **Classification Task**



TorchSpatial Framework – Inference Architecture

- **Regression Task**



LocBench

Geo-Aware Image Classification

Classify a given image into its correct category based on the image itself and location metadata.

Task 1: Fine-Grained Species Recognition

BirdSnap	500 categories	19,576 examples
BirdSnap†	500 categories	43,470 examples
NABirds†	555 categories	23,699 examples
iNat2017	5,089 categories	675,170 examples
iNat2018	8,142 categories	461,939 examples

Task 2: Flickr Image Classification

YFCC	100 categories	36,146 examples

Task 3: Remote Sensing Image Classification

fMoW	63 categories	416,612 examples

Geo-Aware Image Regression

Given an image and its location information, predict the continuous target label for the image.

Cat. 1: Landscape Variables Regression

Task 1: Forest Cover	MOSAIKS	498,106 examples
Task 2: Elevation	MOSAIKS	498,115 examples

Cat. 2: Urbanization Metrics Regression

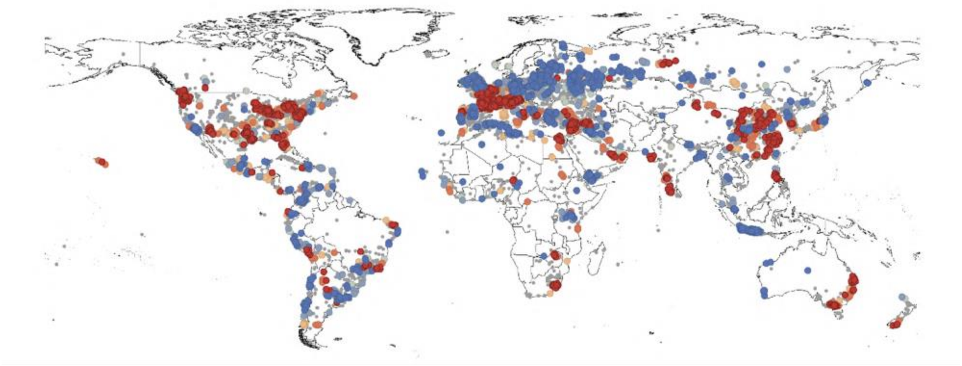
Task 3: Nightlights	MOSAIKS	492,226 examples
Task 4: Population Density	MOSAIKS	425,637 examples

Cat. 3: SDGs Indicators Regression

Task 5: Asset Index	SustainBench	89,936 examples
Task 6: Women BMI	SustainBench	94,866 examples
Task 7: Water Index	SustainBench	86,938 examples
Task 8: Child Mortality Rate	SustainBench	105,582 examples
Task 9: Sanitation Index	SustainBench	89,271 examples
Task 10: Women Edu	SustainBench	117,062 examples

Geo-Bias Score

Geographic bias: a phenomenon in which an AI model **performs differently across geographic regions** and its **predictions are biased** toward some predominated regions.

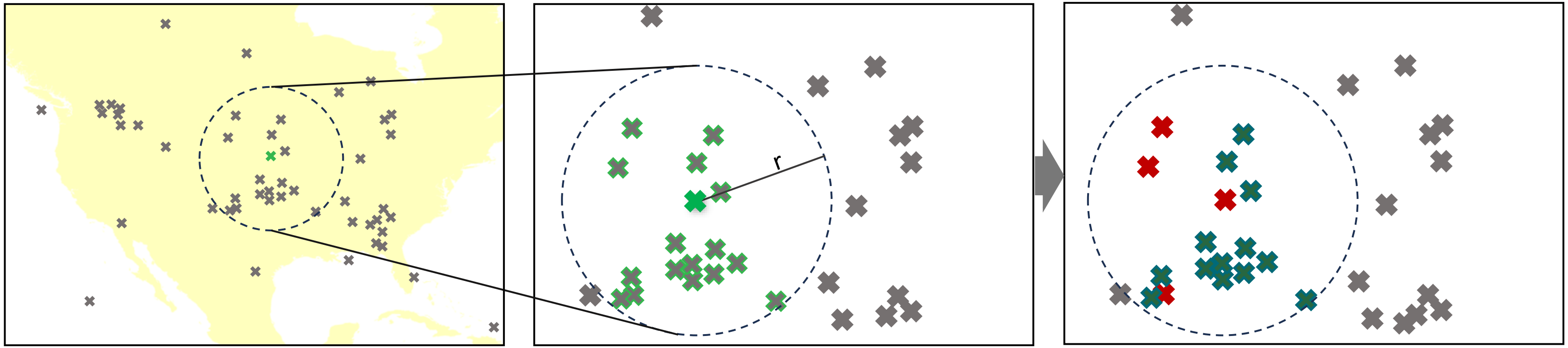


Hot spot analysis of HIT@1 of *space2vec-theory* on fMoW

- We employ the **spatial self-information (SSI),** using a Gaussian distribution to approximate the probability of observing certain types of spatial patterns.

The lower the SSI, the more likely the current spatial patterns arise randomly, namely less geo-baised.

Geo-Bias Score



sample point

Extract a low-performance observation's neighborhood by radius r.

- **Base Geo-Bias Score**

The SSI of low-performance observations **if they are completely randomly distributed**

- **Relative Geo-Bias Score**

Difference between the SSI of the low-performance observations (red crosses) and the SSI of completely random low-performance observations (cyan-blue crosses).

Zhangyu Wang, Krzysztof Janowicz, **Gengchen Mai**, Ivan Majic. Probing the Information Theoretical Roots of Spatial Dependence Measures, In: *Proceedings of the 16th Conference on Spatial Information Theory (COSIT 2024)*, Sep 17 - 20, Québec City, Canada. [ArXiv]

Experiments - Image Classification - Top1Acc

	Task	Species Recognition						Flickr	RS
	Image Classification Dataset	BirdSnap	BirdSnap†	NABirds†	iNat2017	iNat2018	avg	YFCC	fMOW
	$P(y\|x)$ - Prior Type	Test	Test	Test	Val	Val	-	Test	Val
A	No Prior (i.e. image model)	70.07	70.07	76.08	63.27	60.20	67.94	50.15	69.83
B	$tile$ (Tang et al., 2015)	70.20	70.56	75.78	62.54	56.30	67.08	50.01	69.86
	$wrap$ (Mac Aodha et al., 2019)	**72.06**	79.35	**81.78**	68.16	73.11	74.89	51.03	70.34
	$wrap + ffn$ (Mai et al., 2020)	71.93	79.05	81.40	**69.52**	72.29	74.84	50.71	70.11
	rbf (Mai et al., 2020)	71.79	79.58	81.74	68.24	70.03	74.28	51.22	**70.68**
	rff(Rahimi et al., 2007)	71.84	78.91	81.61	68.86	72.32	74.71	50.81	70.24
	Space2Vec-$grid$ (Mai et al., 2020)	71.75	**80.24**	81.70	68.23	73.06	75.00	**51.25**	70.67
	Space2Vec-$theory$ (Mai et al., 2020)	71.79	80.11	81.65	68.30	**73.52**	**75.07**	51.24	70.49
C	xyz (Mai et al., 2023)	71.88	78.96	81.15	68.65	71.44	74.42	50.87	70.16
	$NeRF$ (Mildenhall et al., 2020)	72.10	79.93	81.62	68.74	72.91	75.06	51.27	70.60
	Sphere2Vec-$sphereC$ (Mai et al., 2023)	72.10	79.97	81.91	69.34	72.93	75.25	**51.35**	70.85
	Sphere2Vec-$sphereC+$ (Mai et al., 2023)	**72.15**	**80.90**	**82.13**	68.29	**73.45**	**75.38**	51.31	**70.93**
	Sphere2Vec-$sphereM$ (Mai et al., 2023)	71.88	79.93	81.86	68.51	72.94	75.02	51.18	**70.93**
	Sphere2Vec-$sphereM+$ (Mai et al., 2023)	72.06	79.09	81.67	69.18	72.06	74.81	51.27	70.19
	Sphere2Vec-dfs (Mai et al., 2023)	71.79	78.69	81.44	**69.42**	72.16	74.70	50.65	70.27
	Siren(SH) (Russwurm et al., 2024)	71.88	78.82	81.72	67.68	71.33	74.29	50.57	70.20
D	GPT-4V	55.02	48.89	73.00*	28.00*	18.00*	44.00*	34.00*	17.00*

Experiments - Image Regression - R2

	Image Regression Dataset	Population Density	Forest Cover	Nightlight Luminosity	Elevation
A	No Prior (i.e. image model)	0.38	0.52	0.33	0.27
B	$tile$ (Tang et al., 2015)	0.04	0.46	0.18	0.76
	$wrap$ (Mac Aodha et al., 2019)	0.57	0.72	0.31	**0.79**
	$wrap$ + ffn (Mai et al., 2020)	0.47	0.67	0.28	0.73
	rbf (Mai et al., 2020)	0.25	0.54	**0.32**	0.39
	rff(Rahimi et al., 2007)	0.57	**0.73**	0.23	0.77
	Space2Vec-$grid$ (Mai et al., 2020)	**0.65**	0.69	0.22	0.76
	Space2Vec-$theory$ (Mai et al., 2020)	0.57	**0.73**	0.21	0.78
C	xyz (Mai et al., 2023)	0.49	0.58	0.28	0.72
	$NeRF$ (Mildenhall et al., 2020)	0.60	0.68	0.23	0.76
	Sphere2Vec-$sphereC$ (Mai et al., 2023)	0.63	0.73	0.28	**0.82**
	Sphere2Vec-$sphereC$+ (Mai et al., 2023)	**0.64**	**0.75**	0.27	**0.82**
	Sphere2Vec-$sphereM$ (Mai et al., 2023)	0.62	0.71	0.23	**0.82**
	Sphere2Vec-$sphereM$+ (Mai et al., 2023)	0.53	0.67	0.32	0.74
	Sphere2Vec-dfs (Mai et al., 2023)	0.52	0.62	**0.35**	0.66
	Siren(SH) (Russwurm et al., 2024)	0.62	0.72	0.34	0.80

* (A) No Prior indicates image-only models; (B) geo-aware models with 2D location encoders; (C) geo-aware models with 3D location encoders.

Experiments - Image Regression - R2

	Image Regression Dataset	Asset Index	Women BMI	Water Index	Child Mortality Rate	Sanitation Index	Women Edu
A	No Prior (i.e. image model)	0.40	0.27	0.26	0.02	0.33	0.22
B	$tile$ (Tang et al., 2015)	0.00	0.00	0.00	0.00	0.00	0.00
	$wrap$ (Mac Aodha et al., 2019)	0.47	0.64	0.31	0.33	0.42	0.50
	$wrap$ + ffn (Mai et al., 2020)	0.45	0.63	0.29	0.32	0.39	0.49
	rbf (Mai et al., 2020)	0.56	**0.66**	0.40	**0.36**	0.56	0.51
	rff(Rahimi et al., 2007)	0.50	0.64	0.33	0.34	0.46	0.53
	Space2Vec-$grid$ (Mai et al., 2020)	0.66	**0.66**	0.49	0.32	0.59	0.64
	Space2Vec-$theory$ (Mai et al., 2020)	**0.70**	0.65	**0.52**	0.33	**0.61**	**0.66**
C	xyz (Mai et al., 2023)	0.44	0.62	0.28	0.31	0.38	0.48
	$NeRF$ (Mildenhall et al., 2020)	0.65	0.68	0.50	0.34	0.60	0.64
	Sphere2Vec-$sphereC$ (Mai et al., 2023)	**0.69**	**0.69**	0.52	**0.37**	0.62	**0.66**
	Sphere2Vec-$sphereC$+ (Mai et al., 2023)	**0.69**	0.68	**0.53**	**0.37**	**0.64**	**0.66**
	Sphere2Vec-$sphereM$ (Mai et al., 2023)	0.67	0.68	0.52	**0.37**	0.63	**0.66**
	Sphere2Vec-$sphereM$+ (Mai et al., 2023)	0.45	0.62	0.29	0.31	0.39	0.48
	Sphere2Vec-dfs (Mai et al., 2023)	0.45	0.63	0.30	0.32	0.40	0.49
	Siren(SH) (Russwurm et al., 2024)	0.52	0.65	0.35	0.35	0.47	0.54

Experiments - Image Classification - Geo-Bias

	Task	Species Recognition										Flickr		RS	
	Image Classification Dataset	BirdSnap		BirdSnap†		NABirds†		iNat2017		iNat2018		YFCC		fMOW	
	$P(y\|x)$ - Prior Type	Test		Test		Test		Val		Val		Test		Val	
	Geo-Bias Score	base	rel	base	rel	base	rel	base	rel	base	rel	base	rel	base	rel
A	No Prior (i.e. image model)	28.22	33.11	8.22	7.06	39.71	31.33	26.60	20.37	18.20	13.38	8.05	4.45	375.73	319.66
B	$tile$ (Tang et al., 2015)	27.65	32.10	8.53	7.37	38.43	30.26	26.08	19.91	16.80	12.22	8.41	4.77	375.43	319.77
	$wrap$ (Mac Aodha et al., 2019)	27.76	32.98	**17.17**	**16.60**	**57.37**	**41.99**	**34.83**	**27.50**	**30.78**	**24.31**	7.99	4.41	380.20	323.67
	$wrap + ffn$ (Mai et al., 2020)	29.50	34.99	8.25	7.07	**57.03**	**42.43**	**35.73**	**28.20**	**27.68**	**21.57**	7.77	4.21	377.41	321.20
	rbf (Mai et al., 2020)	17.24	19.75	9.37	8.52	**58.05**	**43.05**	**34.05**	**26.80**	20.48	15.28	7.37	3.86	380.64	324.46
	rff(Rahimi et al., 2007)	28.03	33.61	**13.70**	**12.80**	**57.71**	**42.63**	**34.45**	**27.21**	**28.63**	**22.45**	7.87	4.29	377.94	317.65
	Space2Vec-$grid$ (Mai et al., 2020)	22.26	25.10	**16.27**	**15.42**	**58.96**	**43.38**	**34.10**	**26.87**	**31.12**	**24.71**	7.99	4.43	380.23	323.17
	Space2Vec-$theory$ (Mai et al., 2020)	**36.78***	**42.98***	**15.27**	**14.36**	**59.62**	**44.38***	**34.12**	**26.87**	**31.68***	**24.92**	7.99	4.41	382.49	324.52
C	xyz (Mai et al., 2023)	29.64	35.02	**14.22**	**13.38**	**220.96***	34.09	**34.89**	**27.53**	**26.33**	**20.44**	7.79	4.24	379.84	323.12
	$NeRF$ (Mildenhall et al., 2020)	29.66	35.16	**16.13**	**15.53**	**57.86**	**42.61**	**34.93**	**27.62**	**30.46**	**23.90**	7.81	4.26	375.81	320.30
	Sphere2Vec-$sphereC$ (Mai et al., 2023)	28.84	34.02	**14.78**	**13.94**	**59.26**	**43.68**	**35.77***	**28.21***	**31.61**	**24.96***	7.67	4.16	377.07	320.78
	Sphere2Vec-$sphereC$+ (Mai et al., 2023)	30.43	36.48	**19.99***	**19.24***	**59.13**	**43.47**	**33.14**	**26.02**	**31.55**	**24.85**	8.22	4.66	379.92	323.04
	Sphere2Vec-$sphereM$ (Mai et al., 2023)	31.49	37.02	**16.75**	**16.70**	**58.68**	**43.10**	**33.97**	**26.75**	**31.66**	**24.95**	8.06	4.51	377.26	321.56
	Sphere2Vec-$sphereM$+ (Mai et al., 2023)	27.55	33.04	**14.35**	**13.46**	**53.71**	**40.03**	**35.44**	**27.97**	**26.88**	**20.83**	8.13	4.56	376.64	321.21
	Sphere2Vec-dfs (Mai et al., 2023)	26.39	30.93	**13.57**	**12.50**	**55.43**	**40.75**	**35.52**	**28.05**	**26.00**	**20.13**	7.87	4.30	380.82	323.78
	Siren(SH) (Russwurm et al., 2024)	27.67	32.91	**14.87**	**14.50**	**57.57**	**42.60**	**35.47**	**28.07**	**26.24**	**20.26**	7.68	4.15	377.23	321.15
D	GPT-4V	28.58	34.01	7.06	6.21	-	-	-	-	-	-	-	-	-	-

Experiments - Image Regression - Geo-Bias

	Image Regression Dataset	Population Density		Forest Cover		Nightlight Luminosity		Elevation	
	Geo-Bias Score	base	rel	base	rel	base	rel	base	rel
A	No Prior (i.e. image model)	5.93*	2.53	6.71	2.73	7.47	0.71	6.71	2.92
B	$tile$ (Tang et al., 2015)	5.40	2.34	5.73	2.38	7.60	0.54	6.62	2.87
	$wrap$ (Mac Aodha et al., 2019)	4.86	2.01	5.17	2.27	7.36	**1.08**	5.71	2.74
	$wrap + ffn$ (Mai et al., 2020)	5.04	1.90	5.55	2.40	7.61	0.27	6.12	2.90
	rbf (Mai et al., 2020)	5.39	2.02	5.83	2.28	7.69	0.42	**7.28**	**3.37**
	rff(Rahimi et al., 2007)	5.09	2.27	5.13	2.04	7.55	0.50	5.51	2.61
	Space2Vec-$grid$ (Mai et al., 2020)	5.48	2.42	5.25	2.11	7.64	0.80	6.22	2.65
	Space2Vec-$theory$ (Mai et al., 2020)	5.00	1.97	5.46	2.33	7.55	0.75	5.07	2.24
C	xyz (Mai et al., 2023)	5.64	2.38	5.65	2.37	7.55	0.54	5.96	2.97
	$NeRF$ (Mildenhall et al., 2020)	5.90	2.80*	5.64	2.60	7.70	0.54	4.94	2.03
	Sphere2Vec-$sphereC$ (Mai et al., 2023)	5.69	2.35	6.72*	2.24	7.73*	0.36	**8.83***	**4.16***
	Sphere2Vec-$sphereC$+ (Mai et al., 2023)	5.34	2.37	5.08	2.18	7.50	0.74	5.49	2.51
	Sphere2Vec-$sphereM$ (Mai et al., 2023)	5.21	2.39	5.12	2.31	7.58	0.78	5.01	2.21
	Sphere2Vec-$sphereM$+ (Mai et al., 2023)	5.20	2.51	5.45	2.31	7.54	0.67	6.55	3.13
	Sphere2Vec-dfs (Mai et al., 2023)	5.45	2.59	6.08	2.94*	7.53	**1.10**	6.03	2.58
	Siren(SH) (Russwurm et al., 2024)	5.10	2.29	5.82	2.46	7.48	**1.21***	5.39	2.35

Bold numbers indicate that the scores that are significantly larger (>30%) than the No Prior model; * indicates the scores that are the largest among all models for this dataset.

Thank You

Contact:

Nemin Wu
Nemin.Wu@uga.edu
Qian Cao
Qian.Cao1@uga.edu