The National Academies of
SCIENCES • ENGINEERING • MEDICINE

SPATIALLY EXPLICIT AI

TEXAS
The University of Texas at Austin

June 2025

Spatial Representation Learning: What, How, and Why

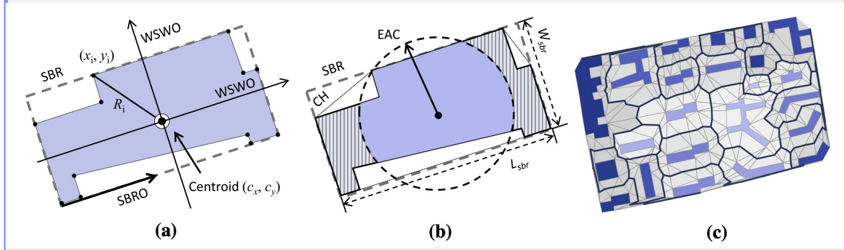
GENGCHEN MAI
Assistant Professor, Department of Geography and the Environment, The University of Texas at Austin

Presentation at the **NASEM GGSC Meeting - Evolving Geodigital Data: Opportunities, Challenges & Disruptions**

Acknowledgement: Google, Microsoft, Microsoft Research, OpenAI, GEORGIA, TEXAS The University of Texas at Austin, UCSB

Comprised Approaches Due to the Lack of SRL

Feature Engineering



Extract features from building polygons (Yan et al, 2022)

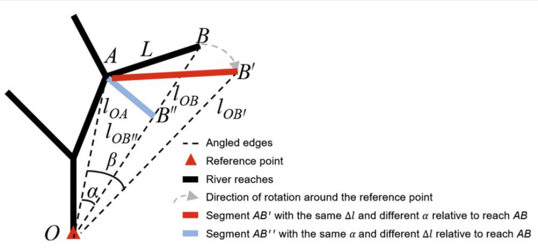
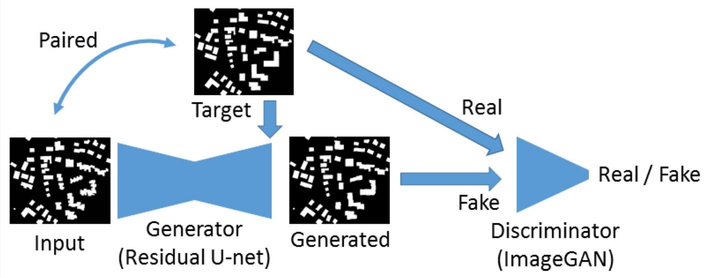


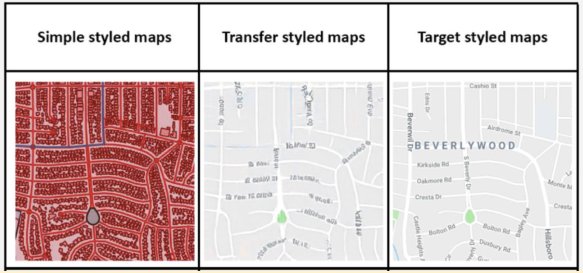
Fig. 4. Schematic diagram of the calculation of the distance difference and the angle of the reference point.

Extract features from drainage patterns (Yu et al, 2022)

Data Conversion



Building polygons to raster images (Feng et al, 2019)



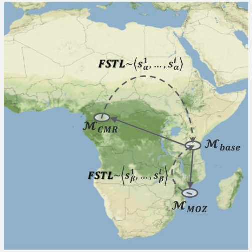
Map vector files to raster images (Kang et al, 2019)

Comprised Approaches Due to the Lack of SRL

Feature Engineering

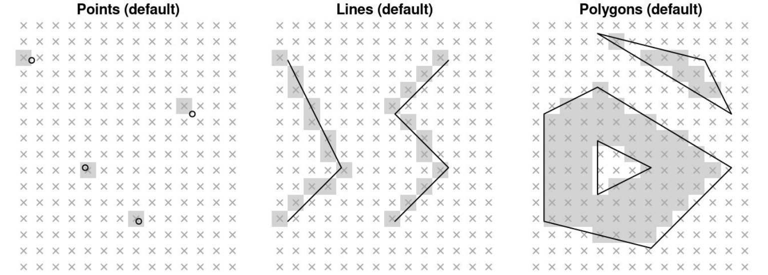
- Heavily relies on domain knowledge

- Hard to generalize to new regions and tasks

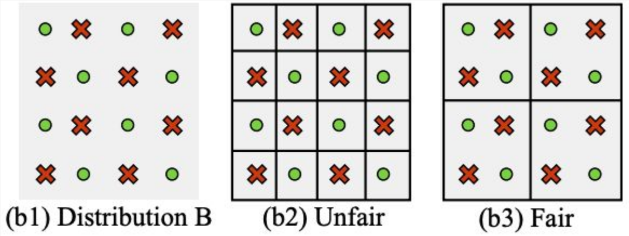


Data Conversion

- Reduced data precision and increase data storage requirement

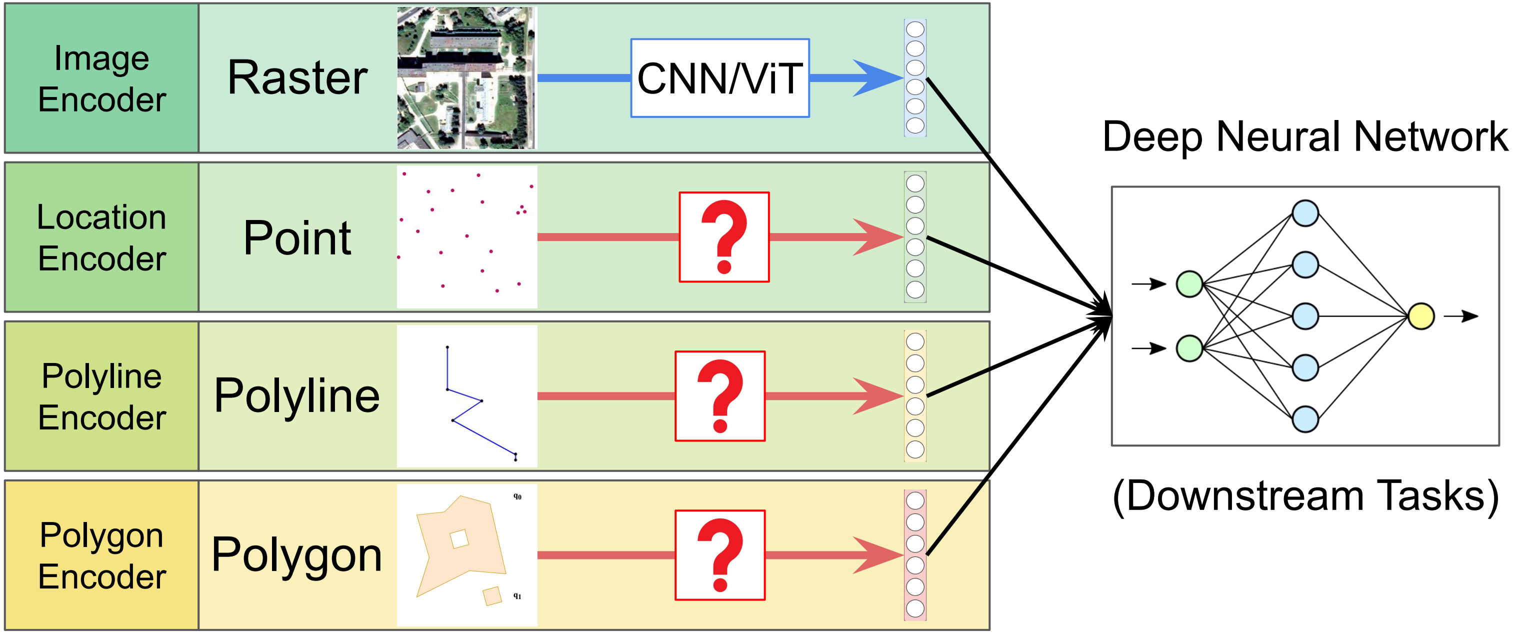


- Modifiable areal unit problem (MAUP)



Spatial Representation Learning (SRL)

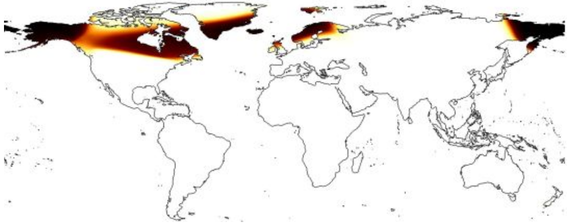
Directly learning neural spatial representations of various types of spatial data in their **native data format** without the need for feature engineering or data conversion step



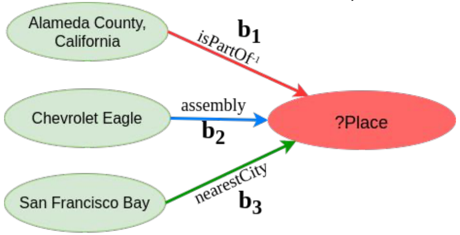
- **Gengchen Mai,** et al. Towards General-Purpose Representation Learning of Polygonal Geometries. *GeoInformatica* 2023.
- **Gengchen Mai**, et al. SRL: Towards a General-Purpose Framework for Spatial Representation Learning (Vision Paper), In: *ACM SIGSPATIAL 2024*.
- **Gengchen Mai**, et al. Towards the Next Generation of Geospatial Artificial Intelligence, *International Journal of Applied Earth Observation and Geoinformation*, 2025.

Various Geospatial Tasks

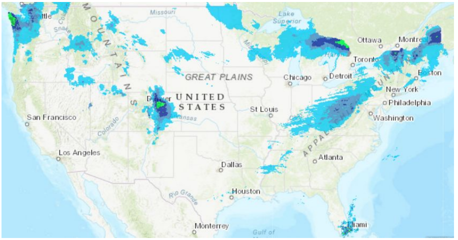
Ecology:
Species Distribution Modeling
(Mac Aoda et al, ICCV 2019; Mai et al., ICLR 2020; Mai et al., ISPRS PHOTO 2023; Mai et al. ICML 2023)



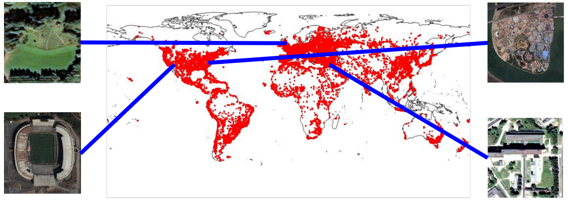
Geospatial Semantics:
Geographic Question Answering
(Mai et al., TGIS 2020; Mai et al., GeoInformatica 2023)



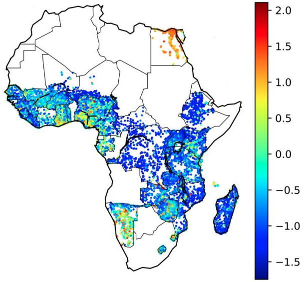
Earth System Science:
Weather Forecasting
(Nguyen et al., ICML 2023)



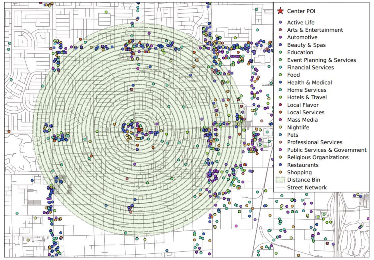
Remote Sensing:
RS Image Classification
(Mai et al., ISPRS PHOTO 2023; Mai et al. ICML 2023; Li et al., SIGSPATIAL 2023)



Sustainability:
Wealth Index Prediction
(Sheehan et al., KDD 2019; Manvi et al., ICLR 2024)



Urban Data Science:
POI Type Prediction
(Mai et al., ICLR 2020)



TorchSpatial

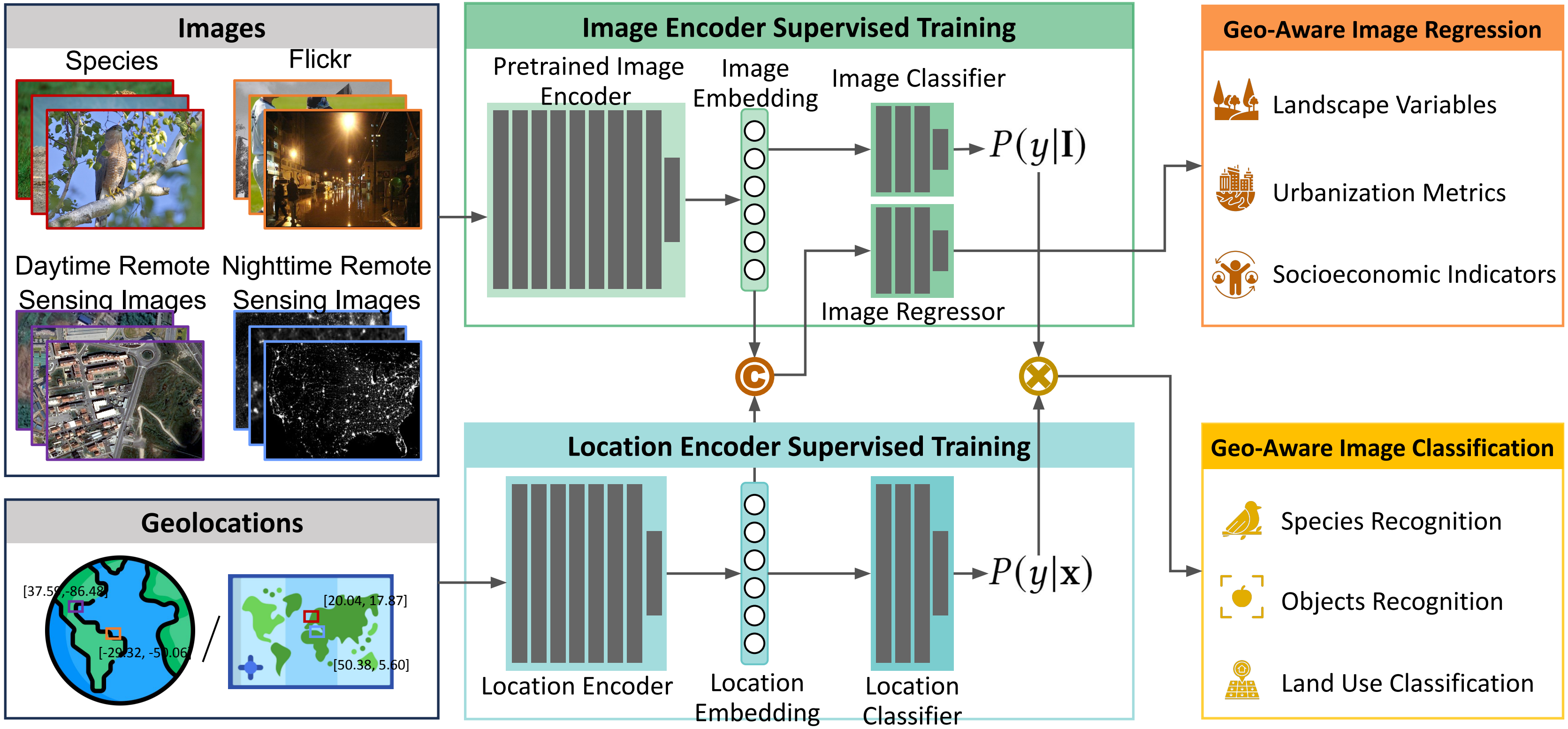
- **A location encoding framework** that consolidates **15 location encoders** and necessary model building blocks for future location encoders.

- **A LocBench benchmark** that encompassing 7 geo-aware image classification datasets and 10 geo-aware image regression datasets

- **An evaluation framework** to quantify geo-aware models' overall performance and their geographic bias, with a novel **Geo-Bias Score** metric

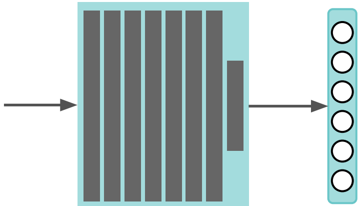
Wu, Nemin, Qian Cao, Zhangyu Wang, Zeping Liu, Yanlin Qi, Jielu Zhang, Joshua Ni, Xiaobai Yao, Hongxu Ma, Lan Mu, Stefano Ermon, Tanuja Ganu, Akshay Nambi, Ni Lao*, **Gengchen Mai***. "TorchSpatial: A Location Encoding Framework and Benchmark for Spatial Representation Learning." **NeurIPS 2024 Data and Benchmark Track**.
***Corresponding Author**

TorchSpatial Framework



Location Encoders

The process of representing a location as a **high dimensional vector** (**location embedding**) such that it can be used for downstream tasks.



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

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

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

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

- **Gengchen Mai**, et al. Multi-Scale Representation Learning for Spatial Feature Distributions using Grid Cells. In ICLR 2020.
- **Gengchen Mai**, et al. A review of location encoding for GeoAI: methods and applications. International Journal of Geographical Information Science 36, no. 4 (2022): 639-673.
- **Gengchen Mai**, et al. Sphere2Vec: A general-purpose location representation learning over a spherical surface for large-scale geospatial predictions. ISPRS Journal of Photogrammetry and Remote Sensing 202 (2023): 439-462.
- **Gengchen Mai**, et al. CSP: Self-supervised contrastive spatial pre-training for geospatial-visual representations. In ICML 2023.

TorchSpatial: Location Encoders

SPATIALLY EXPLICIT AI
TEXAS
The University of Texas at Austin

Category	Location Encoder	Description	Year
2D	*tile*	A discretization-based location encoder	2014
	wrap	A sinusoidal location encoder with $\mathrm{NN}^{wrap}()$	2019
	wrap + *ffn*	A sinusoidal location encoder with $\mathrm{NN}^{ffn}()$	2023
	rbf	A kernel-based location encoder	2020
	rff	A kernel-based location encoder	2020
	Space2Vec-grid	A set of sinusoidal multi-scale location encoders	2020
	Space2Vec-theory		2020
3D	*xyz*	Transform latitude-longitude into 3D Cartesian coordinates	2023
	NeRF	A multiscale version of *xyz*	2021
	Sphere2Vec-sphereC	A set of multi-scale location encoders for spherical surface based on Double Fourier Sphere (DFS) and Space2Vec.	2023
	Sphere2Vec-sphereC+		2023
	Sphere2Vec-sphereM		2023
	Sphere2Vec-sphereM+		2023
	Sphere2Vec-dfs		2023
	Siren(SH)	A learned Double Fourier Sphere location encoder	2024

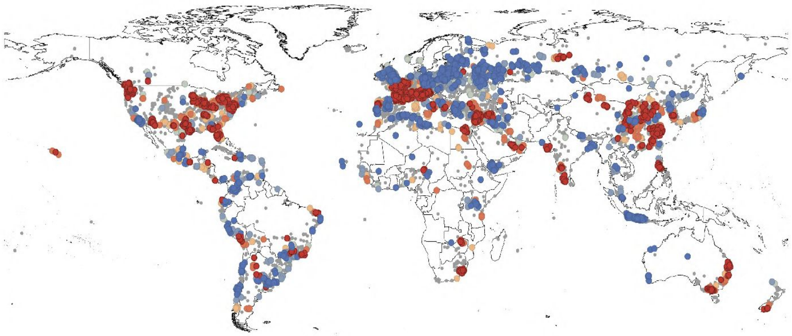
TorchSpatial: LocBench Datasets

Dataset	Category	Label	Task	Instances
BirdSnap	fine-grained species images	species	classification	19,576
BirdSnap+				43,470
NABirds+				23,699
iNat2017				675,170
iNat2018				461,939
YFCC	Flickr images	object categories		36,146
fMoW	remote sensing images	land use types		1,047,691
Population Density		population density	regression	425,637
Forest Cover		forest cover ratio		498,106
Nightlight Luminosity		nightlight luminosity		492,226
Elevation		elevation		498,115
Asset Index	daytime remote sensing imagery & nightlights images	asset wealth index		2,079,036
Women BMI		women BMI		1,781,403
Water Index		water quality index		2,105,026
Child Mortality		child mortality rate		1,936,904
Sanitation Index		sanitation index		2,143,329
Women Edu		women educational attainment		2,910,286

TorchSpatial: Geo-Bias Score Metrics

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

Lower geographic bias means the possibility of encountering a wrong prediction is more uniform across the region of interest



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

Can we use the classic spatial statistic measures (e.g., Moran's I) to quantify the geographic bias?

- Classic spatial autocorrelation statistics metrics can't measure geographic bias because these statistics are **not numerically comparable across different spatial patterns.**
- Moran's I values only tell whether a spatial sample is significantly autocorrelated or not. It is **non-comparable across scenes**.

Spatial Self-Information (SSI)

- **Spatial Self-Information (SSI)**: using a Gaussian distribution to **approximate the probability of observing certain types of spatial patterns**.

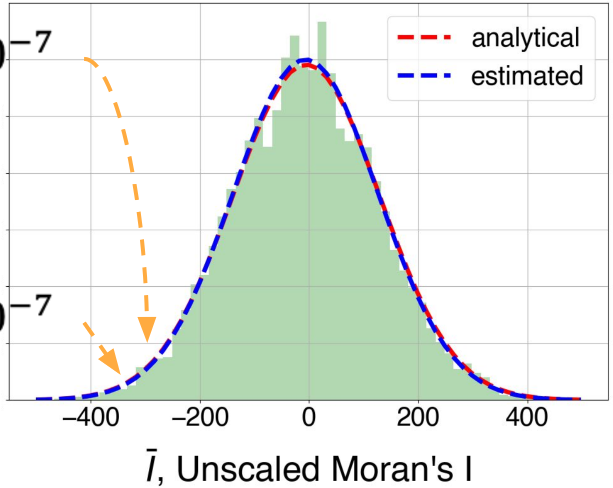
$\mathbb{P}(\;\cdot\;) = 8.32\times10^{-7}$

$J(\;\cdot\;) = 14.66$

$\mathbb{P}(\;\cdot\;) = 3.06\times10^{-7}$

$J(\;\cdot\;) = 15.22$

$$\mathbb{P}(I = i|Rnd)$$

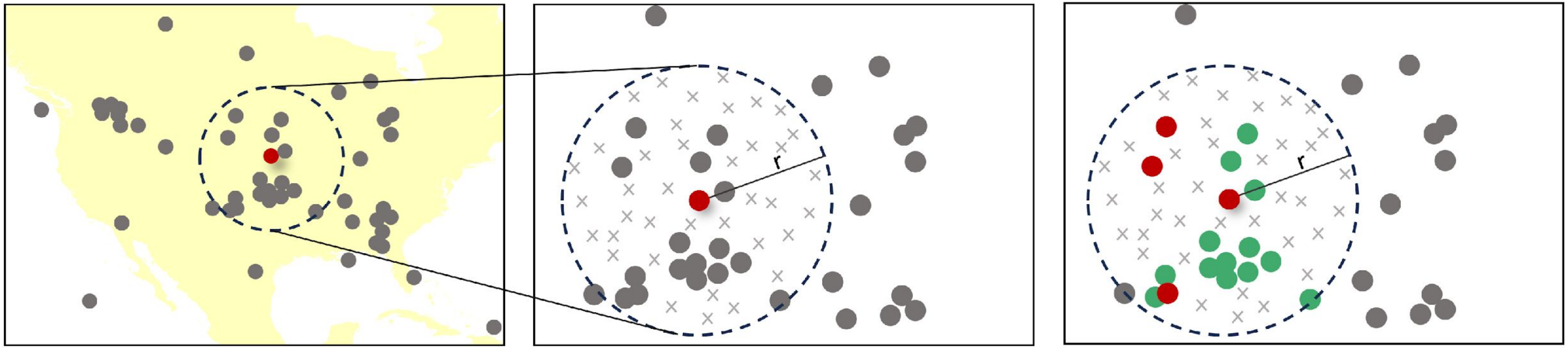


The lower the probability, the less likely the current spatial patterns arise randomly, and consequently the stronger the spatial autocorrelation.

Zhangyu Wang, Krzysztof Janowicz, **Gengchen Mai**, Ivan Majic. Probing the Information Theoretical Roots of Spatial Dependence Measures, In: *COSIT 2024*, Sep 17 - 20, Québec City, Canada. [ArXiv] *** Best Paper in COSIT 2024**

TorchSpatial: Geo-Bias Score metric

Using SSI for Geo-Bias quantification: higher SSI scores indicates stronger the spatial autocorrelation, thus higher geographic bias



sample point

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

Unmarked SSI

The intrinsic **sampling geo-bias** regardless of the markers (model performances)

Marked SSI

The **geo-bias of model performance**, considering both where the data are observed and how the model performs at these locations

GitHub **Website** **Datasets**

Geo-Bias of Various Foundation Models

SPATIALLY EXPLICIT AI

TEXAS The University of Texas at Austin

Table 4: Accuracy and Global Geo-Bias Scores of remote sensing image classification. All geo-bias scores use an ROI radius of 0.01 radian. **Bold** numbers indicate the best performance or the lowest geo-bias. **Bold** numbers indicate the best performance or the lowest geo-bias scores.

	Model	Acc ↑	U-SSI ↓	M-SSI ↓	SG-SRE ↓	DL-SRE ↓	DS-SRE ↓	SPAD ↓
fMoW-sentinel	**Hyperparam**	-	-	-	0.01	0.01	8	-
	GPT-4o	5.72	**516.80**	**63.96**	3.81	1.32	0.76	18.25
	CROMA ft	52.67	560.80	447.89	61.47	16.79	19.94	39.19
	CROMA lp	31.46	560.11	466.31	152.72	38.69	42.17	36.37
	SatMAE ft	**64.77**	560.96	16.29	**2.16**	**0.57**	**0.74**	12.84
	SatMAE lp	62.76	561.29	14.06	2.47	0.61	0.88	**12.36**
WorldStrat -LCCS	**Hyperparam**	-	-	-	0.05	0.005	8	-
	GPT-4o	41.33	399.27	276.18	7.87	54.87	62.21	66.93
	CROMA ft	**60.78**	**354.01**	275.65	12.91	18.89	23.46	63.23
	CROMA lp	58.73	369.52	305.35	3.98	10.59	21.51	66.56
	SatMAE ft	52.37	418.63	**6.00**	**0.06**	**0.11**	**0.14**	16.51
	SatMAE lp	44.29	416.44	6.95	**0.06**	0.12	0.16	**15.29**
WorldStrat -IPCC	**Hyperparam**	-	-	-	0.05	0.005	8	-
	GPT-4o	51.92	404.86	200.12	3.82	12.06	12.40	56.40
	CROMA ft	**69.61**	**359.10**	251.67	7.64	13.21	14.67	52.27
	CROMA lp	65.79	379.79	271.37	4.64	8.28	9.09	56.12
	SatMAE ft	66.56	410.33	19.23	**0.07**	0.18	**0.15**	**15.81**
	SatMAE lp	45.36	416.16	**7.06**	0.08	**0.14**	0.16	16.07
EuroSAT	**Hyperparam**	-	-	-	0.01	0.005	12	-
	GPT-4o	44.89	119.43	79.59	2.62	1.29	0.64	53.52
	CROMA ft	**97.43**	115.72	96.58	0.25	0.67	0.48	**8.67**
	CROMA lp	92.87	**100.00**	60.35	0.56	0.44	0.37	19.23
	SatMAE ft	74.30	115.93	13.02	0.03	**0.07**	**0.05**	15.65
	SatMAE lp	56.54	113.19	**6.43**	**0.02**	**0.07**	0.06	34.91

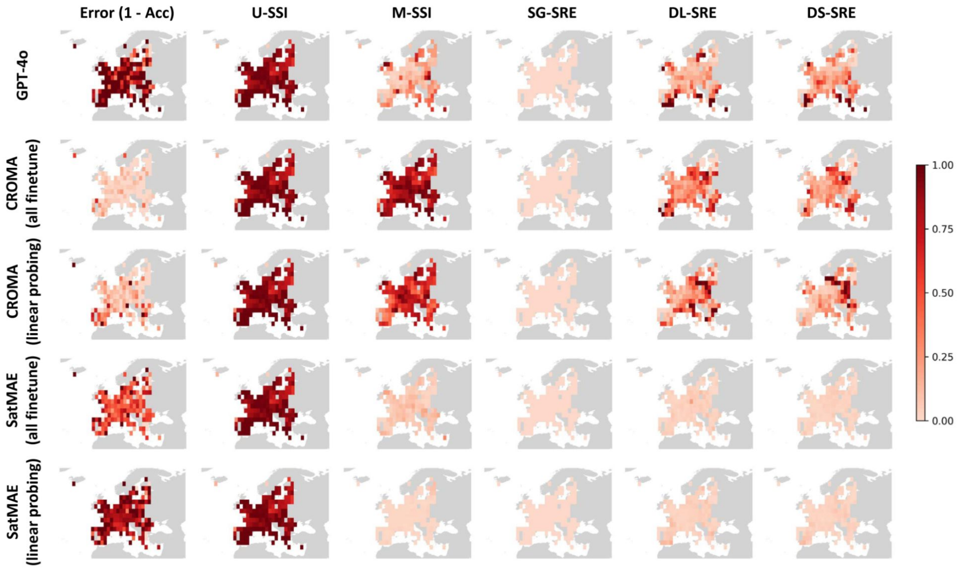
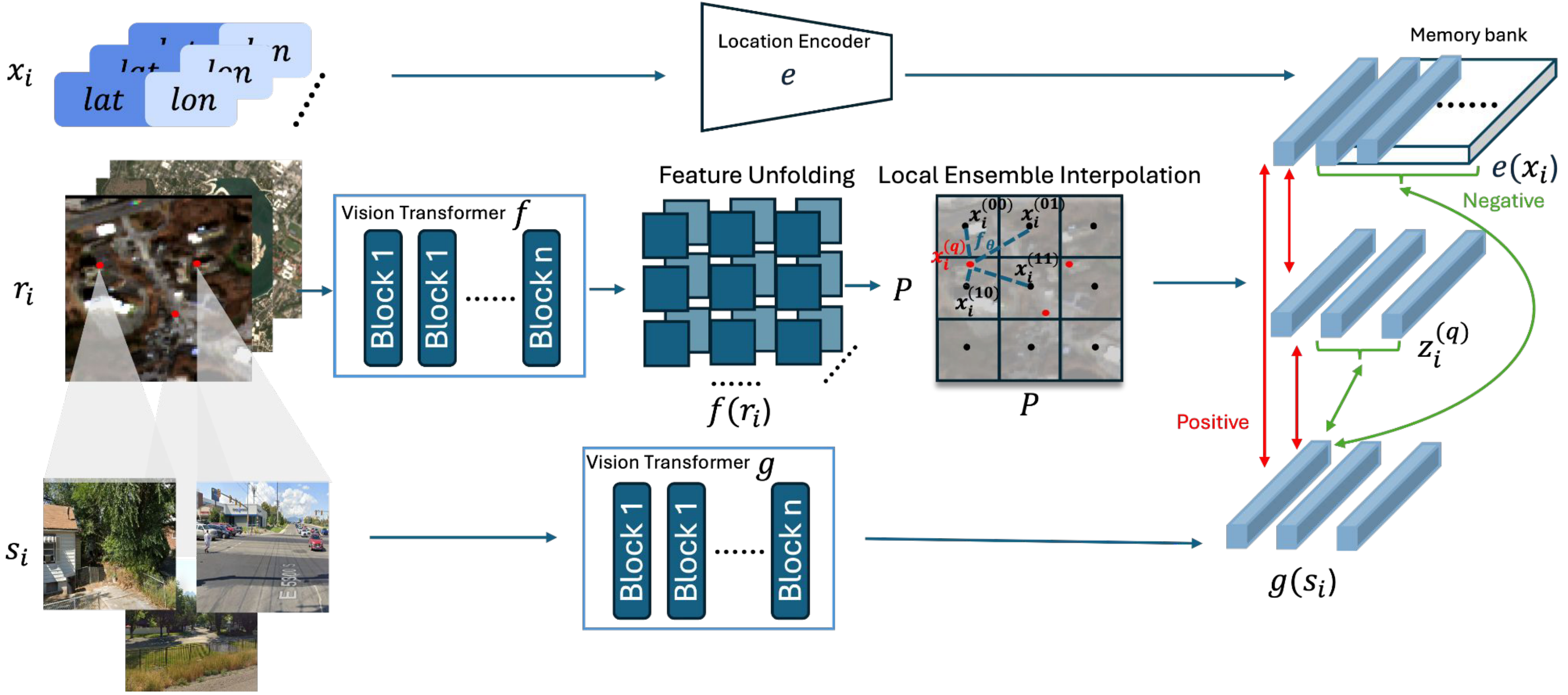


Figure 5: Geographical distributions of error rate and local geo-bias scores of different remote sensing foundation models on EuroSAT. The spatial distributions of U-SSI across models are the same because local U-SSI scores are unmarked and only dependent on data instead of models.

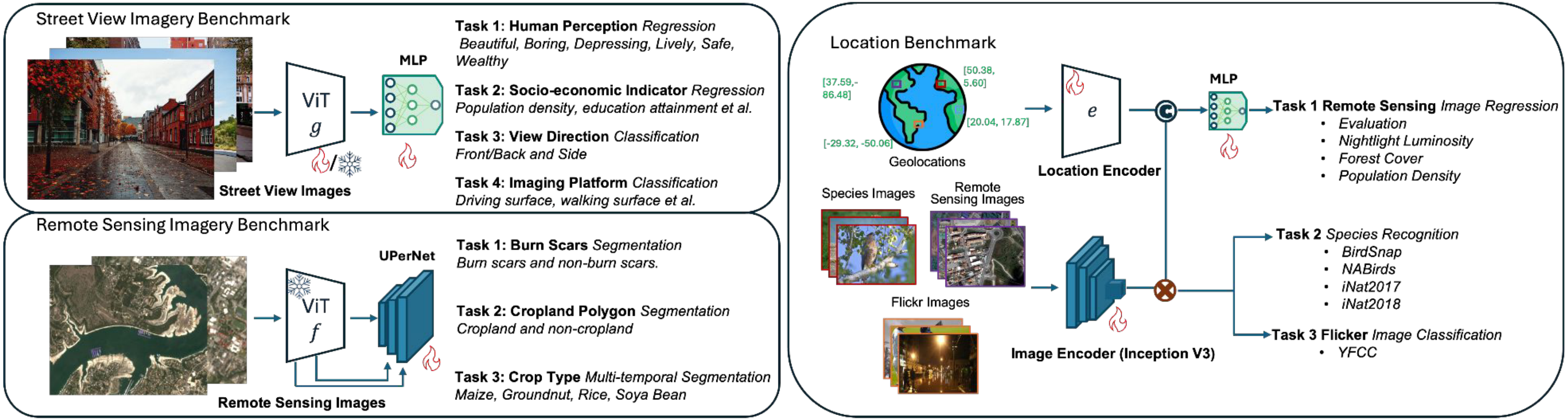
GAIR: Framework



GAIR: Tasks

Evaluation GAIR on 10 downstream tasks (22 datasets):

- Street View Image Benchmark Tasks
- Remote Sensing Imagery Benchmark Tasks
- Location Benchmark Task



GAIR: Evaluations

Table 2. Performance comparison (mIoU ↑) of GAIR and other Geo-Foundation Models (GeoFMs) across four remote sensing benchmark datasets. SS refers to semantic segmentation tasks, while Linear and L-TAE [22] represent two different multi-temporal augmentation strategies

Task	Single Temporal SS		Multi Temporal SS	
	Burn Scars	Cropland Polygon	Crop Type	
Model			Linear	L-TAE
CROMA [20]	81.95	25.65	47.02	49.38
DOFA [76]	78.96	27.07	49.81	51.33
GFM-Swin [52]	76.17	27.19	39.72	46.98
Prithvi [34]	82.67	26.86	39.92	43.07
RemoteCLIP [39]	75.55	25.12	46.50	52.05
SatlasNet [6]	79.69	25.13	46.97	46.97
Scale-MAE [58]	76.71	21.47	21.39	25.42
SpectralGPT [28]	80.47	26.75	53.50	46.95
S12-Data2Vec [61]	81.14	24.23	54.01	54.03
S12-DINO [61]	81.44	25.62	46.56	48.66
S12-MAE [61]	80.86	24.69	46.28	45.80
S12-MoCo [61]	80.76	25.38	44.22	48.58
GAIR-MAE	74.15	22.77	34.18	40.44
GAIR w/o Loc	82.94	43.28	55.41	**54.32**
GAIR	**83.26**	**43.35**	**55.53**	54.01

Table 3. Evaluation results on LocBench [74], including three tasks: image regression (4 datasets), species recognition (4 datasets), and Flickr image classification (1 dataset). For image regression tasks, we report the R^2 score (↑), while for species recognition and Flickr classification, we report the Top-1 accuracy (↑).

Initialization	Task	Image Regression			
	Model	Population Density	Forest Cover	Nightlight Luminosity	Elevation
Random Init.	No Prior	0.38	0.52	0.33	0.27
	RFF	0.57	0.84	0.35	0.76
GeoCLIP Init.	RFF	0.61	0.84	0.37	0.78
GAIR Init.	RFF	**0.67**	**0.86**	**0.40**	**0.82**

Initialization	Task	Species Recognition				Flicker Classification
	Model	BirdSnap	NABirds	iNat2017	iNat2018	YFCC
Random Init.	No Prior	70.07	76.08	63.27	60.20	50.15
	RFF	70.07	81.63	67.73	71.66	**51.13**
GeoCLIP Init.	RFF	70.56	81.65	67.78	71.93	51.01
GAIR Init.	RFF	**72.07**	**81.76**	**67.84**	**72.48**	**51.13**

Paper

Zeping Liu, Fan Zhang, Junfeng Jiao, Ni Lao*, **Gengchen Mai***. "GAIR: Improving Multimodal Geo-Foundation Model with Geo-Aligned Implicit Representations." **Under Review 2025**. ***Corresponding Author**

Conclusion

Takeaways

- **Spatial Representation Learning (SRL)** is a key component for Geospatial AI model development
- **Geo-Bias** is a unique and important issue for geospatial data and model requiring attention
- **Geo-Foundation Model** is one major research direction for GeoAI where SRL will play a major role

Further Thoughts

- Other GeoEthics concerns: **geo-privacy, spatiotemporal replicability,** and **explainability**

- Wu, Nemin, Qian Cao, Zhangyu Wang, Zeping Liu, Yanlin Qi, Jielu Zhang, Joshua Ni, Xiaobai Yao, Hongxu Ma, Lan Mu, Stefano Ermon, Tanuja Ganu, Akshay Nambi, Ni Lao*, **Gengchen Mai***. TorchSpatial: A Location Encoding Framework and Benchmark for Spatial Representation Learning. **NeurIPS 2024 Data and Benchmark Track**. ***Corresponding Author**
- Zeping Liu, Fan Zhang, Junfeng Jiao, Ni Lao*, **Gengchen Mai***. "GAIR: Improving Multimodal Geo-Foundation Model with Geo-Aligned Implicit Representations." **Under Review 2025**. ***Corresponding Author**

TorchSpatial

GAIR

GitHub **Website** **Datasets** **Paper**