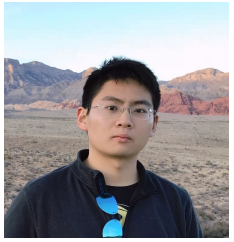
CSP: Self-Supervised Contrastive Spatial Pre-Training for Geospatial-Visual Representations
– Towards a Multimodal Foundation Model for GeoAI

Dr. Gengchen Mai

[2022 - Now] Assistant Professor at *Department of Geography, University of Georgia*

https://gengchenmai.github.io/



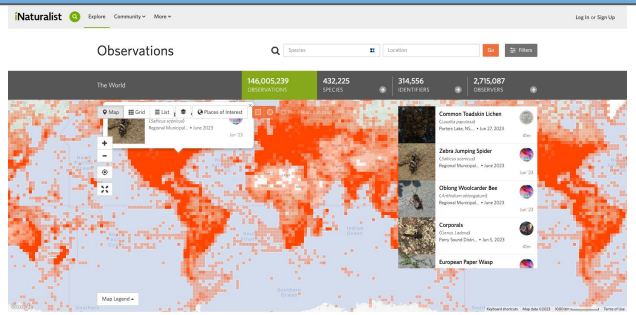
Acknowledgement:

Massive Unlabeled Geo-tagged Image Datasets

Unlabeled RS Images

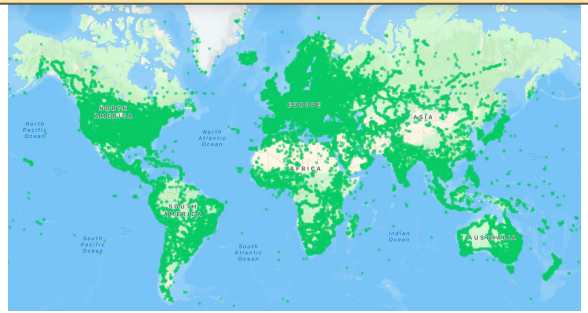
Billions of unlabeled satellite images are collected from various sensors everyday (Figure from NASA Website)

Unlabeled iNaturalist Images



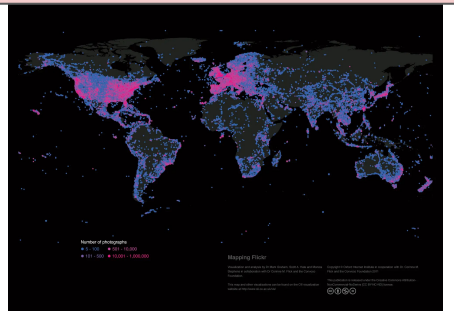
Millions of unlabeled geo-tagged species images are collected everyday (Figure from iNaturalist Website)

Unlabeled StreetView Images



Billions of unlabeled Mapillary StreetView images are uploaded everyday (Figure from Mapiliary Website)

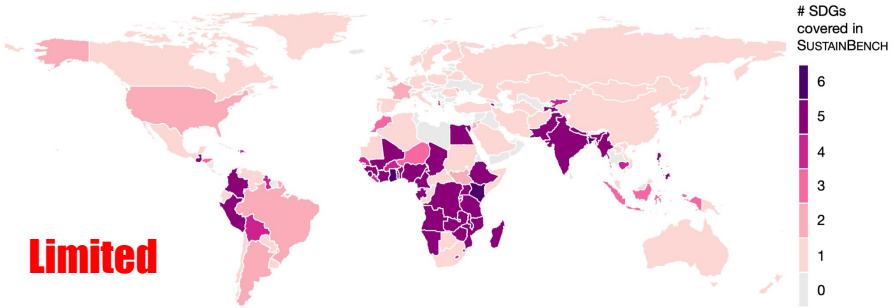
Unlabeled Flickr Images



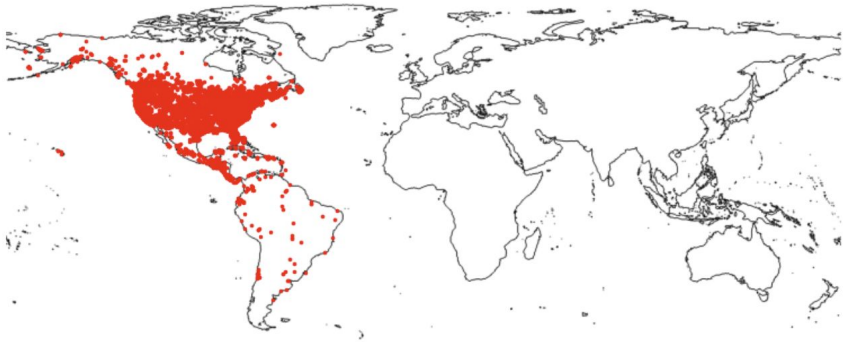
Billions of unlabeled Flickr images are uploaded everyday (Figure from Oxford Internet Institute)

Unlabeled v.s. Labeled Geospatial Image Datasets

Well-curated geospatial dataset, in contrast, have **limited sizes**, **imbalanced geographic coverage**, and **potentially oversimplified label distributions**



Geographic coverages of labeled satellite/streetview image datasets of a collections of 15 benchmark tasks in the SustainBench dataset (Yeh et al., 2022)

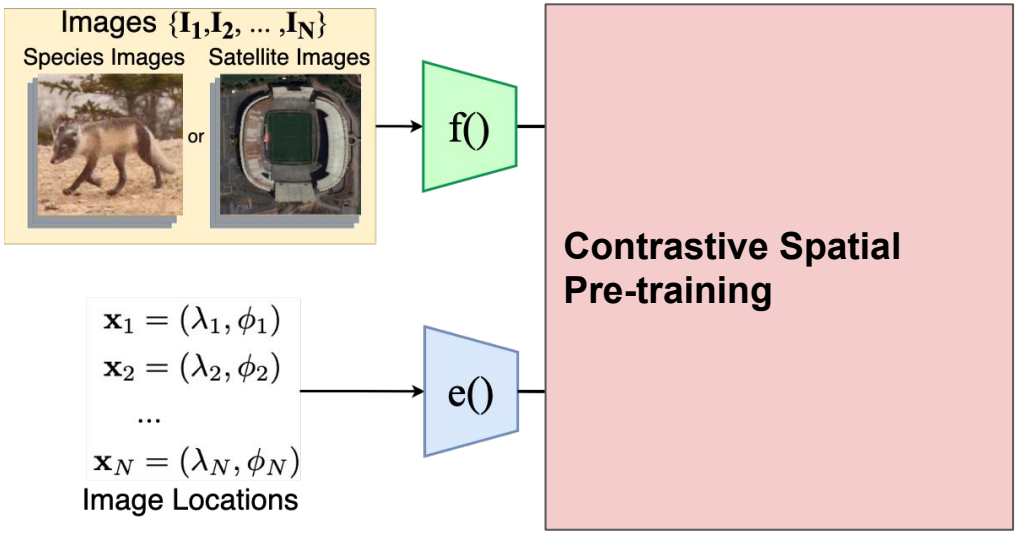


Geographic coverage of labeled species fine-grained recognition dataset – NABird (Mai et al., 2023)

Solution: instead of only supervised training on labeled geospatial images, we build a **multi-modal SSL framework** between **geo-locations** and **images** on the **massive unlabeled geo-tagged images**.

A Multimodal Pre-training Objective for GeoAI

Build a **contrastive pre-training** objective between **geospatial** and **visual** signals



Gengchen Mai, et al. CSP: Self-Supervised Contrastive Spatial Pre-Training for Geospatial-Visual Representations, In: *ICML 2023*.

Geo-Aware Image Classification

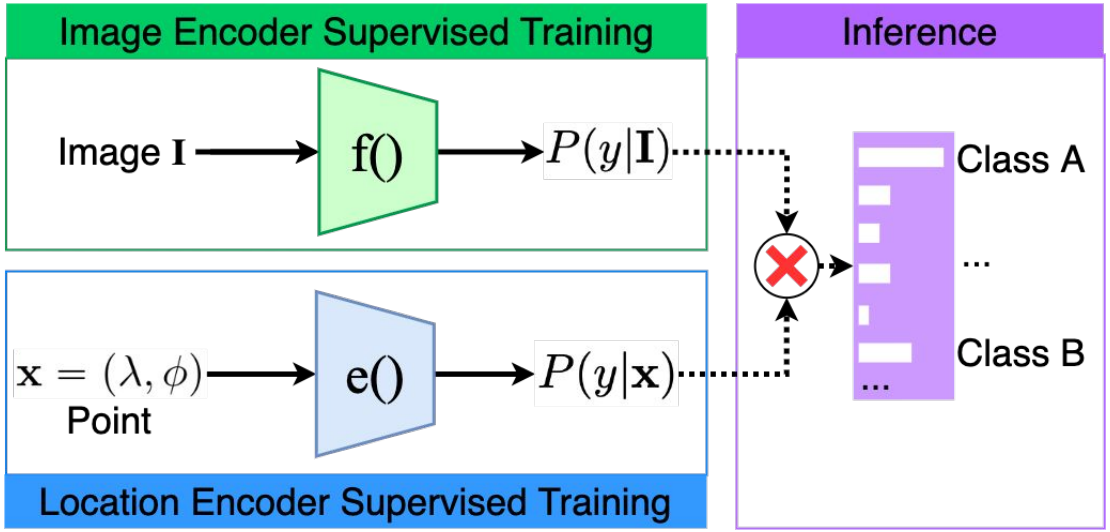


Figure 2(a) Sup. Only: Geo-aware Supervised Learning (Mac Aodha et al., 2019; Mai et al., 2020b; Mai et al., 2023)

Geo-Aware Image Classification

- **ImageNet Pretraining** (Deng et al., 2009): pre-training f() on ImageNet dataset;
- **Tile2Vec** (Jean et al., 2019): pretraining f() with an unsupervised geo-aware triplet loss;
- **Geo-SSL** (Ayush et al., 2021) and **SeCo** (Manas et al., 2021): pretraining f() with a geo-aware contrastive loss;
- **GeoKR** (Li et al., 2021a): pretraining f() in a teacher-student network by minimizing the KL loss between the image representations and a spatially aligned land cover maps M.

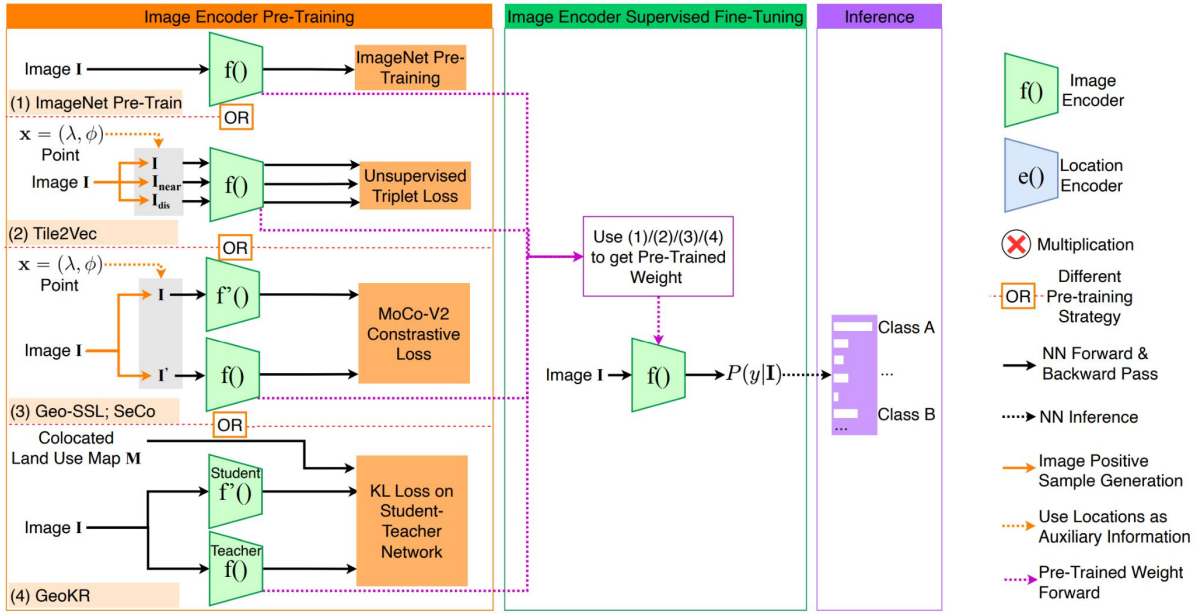


Figure 2(b) Img. Only: Image Encoder Pre-Training with Geographic Knowledge

Contrastive Spatial Pre-Training (CSP)

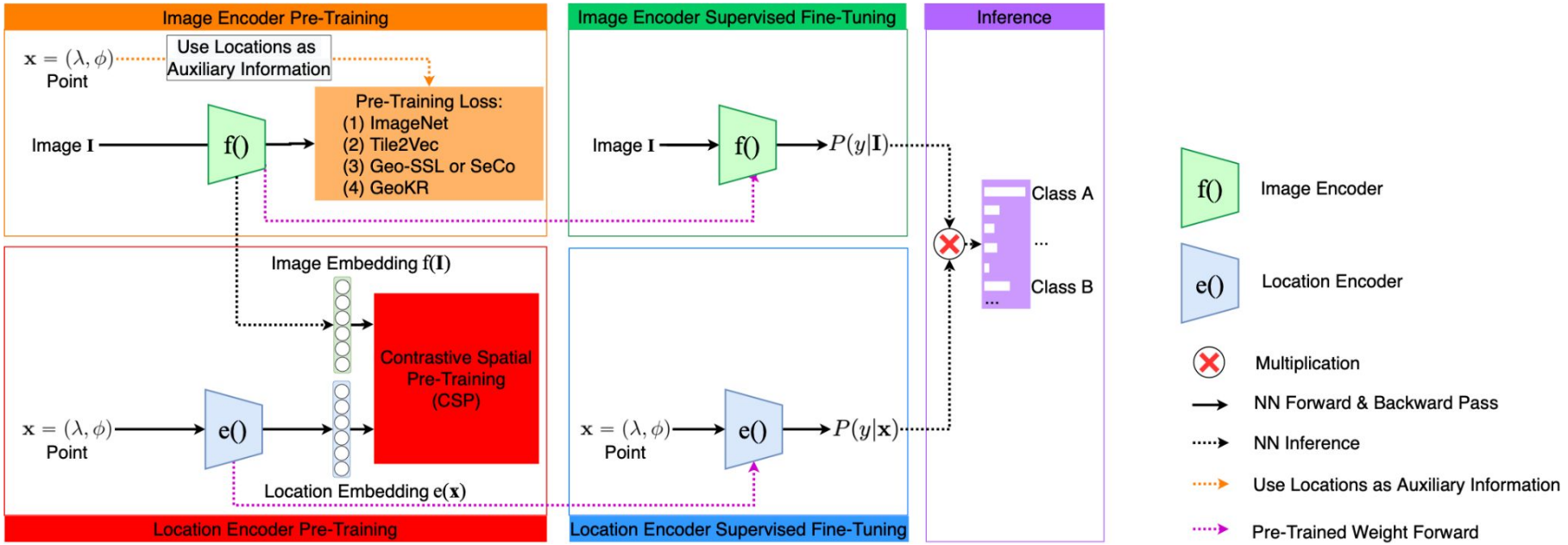
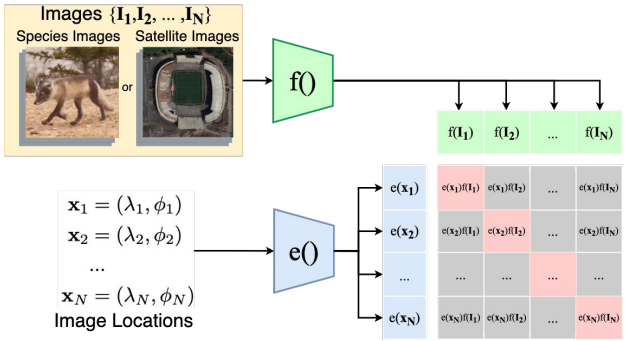


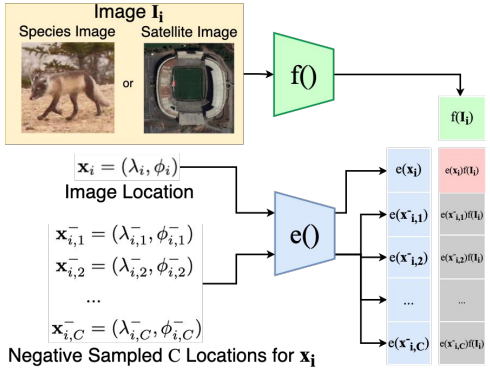
Figure 2(c) Contrastive Spatial Pre-Training (CSP)

Contrastive Spatial Pre-Training (CSP)

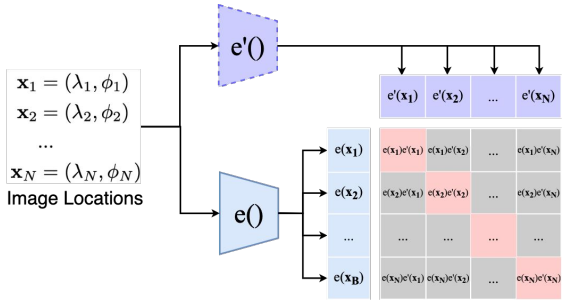
Contrast the representations between **geo-locations** and **images** in a self-supervised learning manner in three ways:



(a) In-batch negative sampling



(b) Random negative location sampling



(c) SimCSE sampling

Geo-Aware Image Classification

- CSP can improve model performance on both **iNat2018** and **fMoW** dataset on both **few-shot** and **fully supervised** learning setting with **various labeled training data sampling ratios.**
- On iNat2018, CSP significantly boosts the model performance with **10-34%** relative improvement with v**arious labeled training data sampling ratios**.

Fine-grained species recognition on iNat2018 dataset

Table 1: The Top1 accuracy of different models and training strategies on the iNat2018 validation dataset for the species fine-grain recognition task with different training data ratios, where $\lambda\% = 100\%$ indicates the fully supervised setting. We run each model 5 times and report the standard deviation in "()".

Ratio $\lambda\%$	5%	10%	20%	100%
Img. Only (ImageNet) (Szegedy et al., 2016)	5.28 (-)	12.44 (-)	25.33 (-)	60.2 (-)
Sup. Only (wrap) (Mac Aodha et al., 2019)	7.12 (0.02)	12.50 (0.02)	25.36 (0.03)	72.41 (-)
Sup. Only (grid) (Mai et al., 2020b)	8.16 (0.01)	14.65 (0.03)	25.40 (0.05)	72.98 (0.04)
MSE	8.15 (0.02)	17.80 (0.05)	27.56 (0.02)	73.27 (0.02)
CSP-NCE-BLD	8.65 (0.02)	18.75 (0.12)	28.15 (0.07)	73.33 (0.01)
CSP-MC-BLD	**9.01 (0.02)**	**19.68 (0.05)**	**29.61 (0.03)**	**73.79 (0.02)**

Satellite image scene classification on fMoW dataset

Table 5: The Top1 accuracy of different models and training strategies on the fMoW val dataset for the satellite image classification task with different training data ratios, where $\lambda\% = 100\%$ indicates fully supervised setting. We report the standard errors (SE) over 5 different runs.

Ratio $\lambda\%$	5%	10%	20%	100%
Img. Only (Tile2Vec) (Jean et al., 2019)	59.41 (0.23)	61.91 (0.31)	62.96 (0.51)	64.45 (0.37)
Img. Only (Geo-SSL) (Ayush et al., 2021)	65.22 (-)	66.46 (-)	67.66 (-)	69.83 (-)
Sup. Only (wrap) (Mac Aodha et al., 2019)	66.67 (0.03)	68.22 (0.01)	69.45 (0.01)	70.30 (0.02)
Sup. Only (grid) (Mai et al., 2020b)	67.01 (0.02)	68.91 (0.04)	70.20 (0.03)	70.77 (0.03)
MSE	67.06 (0.04)	68.90 (0.05)	70.16 (0.02)	70.45 (0.01)
CSP-NCE-BLD	67.29 (0.03)	69.20 (0.03)	70.65 (0.02)	70.89 (0.04)
CSP-MC-BLD	**67.47 (0.02)**	**69.23 (0.03)**	**70.66 (0.03)**	**71.00 (0.02)**

Ablation Study

Ablation study 1: The effect of **different SSL pre-training objectives**

Table 2: Ablation studies on different CSP-MC-* pretraining objectives on the iNat2018 validation dataset with different $\lambda\%$. Here, CSP-MC-BLD indicates the CSP training on the MC loss with all three components. CSP-MC-BL deletes the SimCSE $l^D_{\mathrm{MC}}(\mathbb{X})$ component in Equation 4. The rest models follow similar logic.

Ratio $\lambda\%$	5%	10%	20%	100%
CSP-MC-BLD	**9.01**	**19.68**	**29.61**	**73.79**
CSP-MC-BD	8.63	19.60	29.52	73.15
CSP-MC-BL	8.40	17.17	26.63	73.36
CSP-MC-B	8.16	16.58	25.89	73.10

Ablation study 2: The effect of **location embedding dimensions**

Table 3: Ablation studies on different location embedding dimensions d on the iNat2018 validation dataset with different $\lambda\%$.

	d	5%	10%	20%	100%
CSP-MC-BLD	64	7.64	16.57	25.31	71.76
CSP-MC-BLD	128	8.5	19.35	29.11	72.89
CSP-MC-BLD	256	**9.01**	**19.68**	**29.61**	73.62
CSP-MC-BLD	512	8.97	18.8	27.96	73.67
CSP-MC-BLD	1024	8.78	17.94	26.65	**73.79**

Ablation study 3: The effect of d**ifferent image encoders**

Table 4: Ablation studies on different image neural network $\mathbb{F}()$ (InceptionV3 (Szegedy et al., 2016) and ViT (Dosovitskiy et al., 2021)) on the iNat2018 validation dataset with $\lambda\% = 5\%$.

$\mathbb{F}()$	Inception V3	ViT
Img. Only (ImageNet) (Szegedy et al., 2016)	5.28	12.46
Sup. Only (wrap) (Mac Aodha et al., 2019)	7.12	18.66
Sup. Only (grid) (Mai et al., 2020b)	8.16	18.68
MSE	8.15	20.02
CSP-NCE-BLD	8.65	20.16
CSP-MC-BLD	**9.01**	**20.78**

Website: https://gengchenmai.github.io/csp-website/

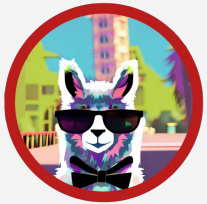
ArXiv: https://arxiv.org/abs/2305.01118

Code: https://github.com/gengchenmai/csp

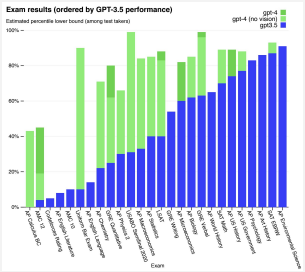
Foundation Models (FMs) in Different Domains

Natural Language Processing

Stanford Alpaca



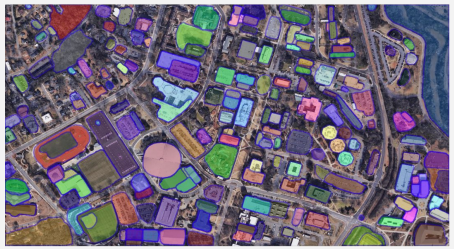
Stanford Alpaca



ChatGPT/GPT-4 (OpenAI. 2023)

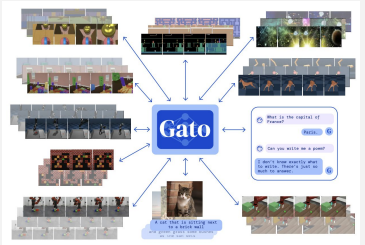
Computer Vision

Imagen (Saharia et al. 2022)



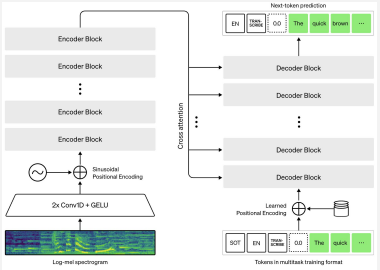
Segment Anying (Kirillov et al, 2023)

Reinforcement Learning



Gato (Reed et al. 2022)

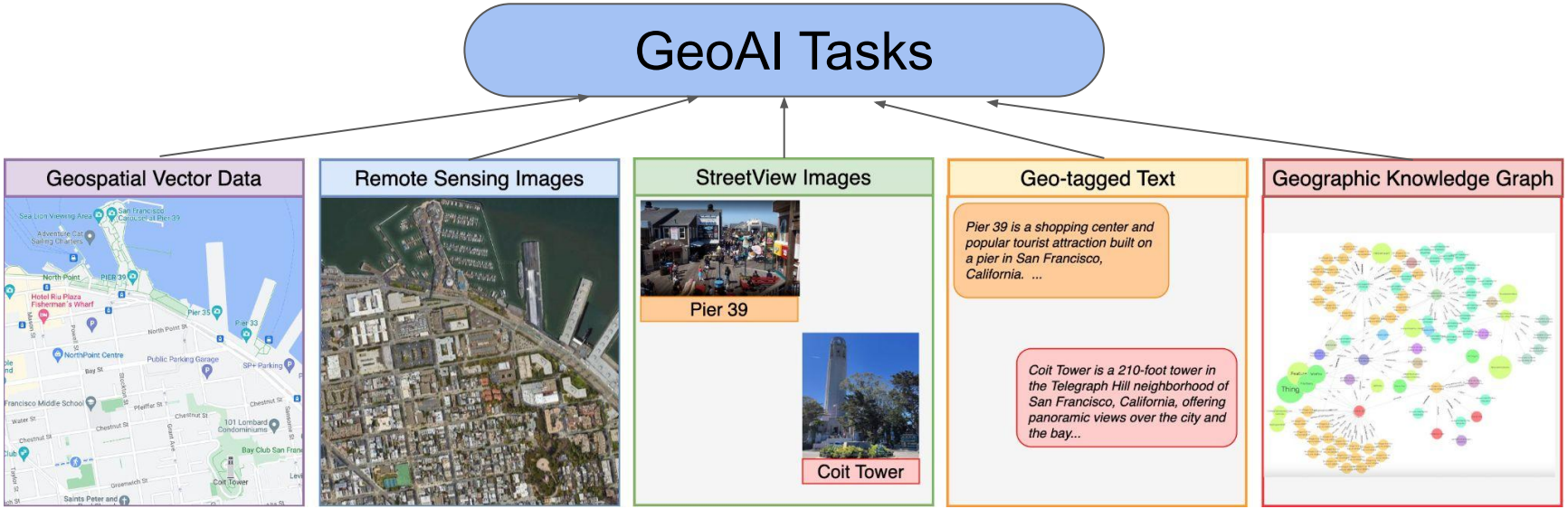
Signal Processing



Whisper (Radford et al. 2022)

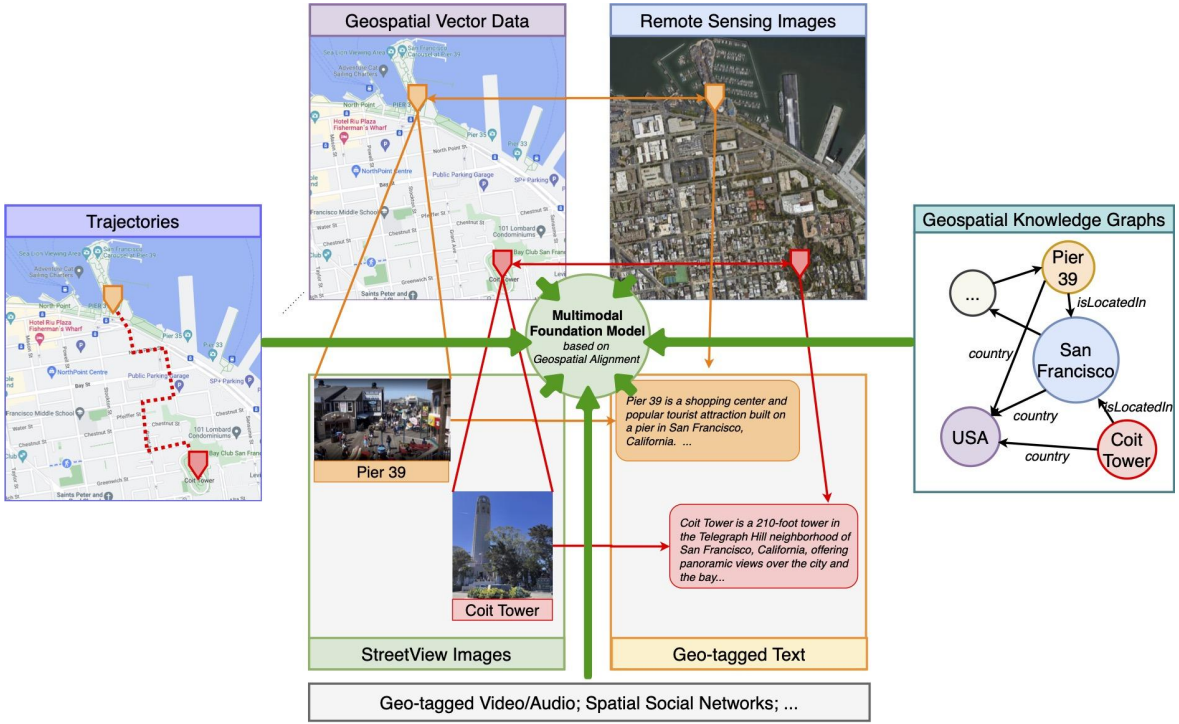
Unique Challenges of GeoAI for FMs

- **Uniqueness of GeoAI Tasks: many data modalities** which calls for **multimodal approaches**



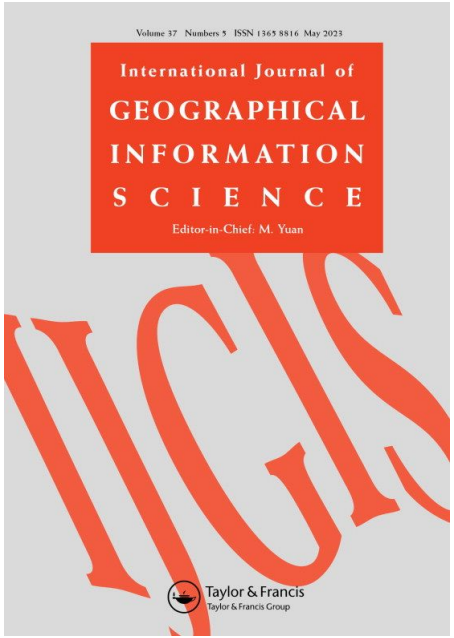
A Multimodal FM for GeoAI

Vision: a multimodal FM for GeoAI that use their **geospatial relationships as alignments** among **different data modalities.**



IJGIS Special Issue on Geo-Foundation Models

GeoFM: Foundation Models for Geospatial Artificial Intelligence



Relevant Topics Include

- Benchmark the effectiveness of foundation models on different geospatial applications
- Novel prompt engineering methods for geo-foundation models
- Zero-shot and few-shot learning with geo-foundation models
- Fine-tuning foundation models on various geospatial tasks
- Development of (multimodal) foundation models for GeoAI applications
- Societal impacts, risks, and biases of foundation models for geospatial problems
- Endeavors in gathering and curating large-scale geospatial datasets for training/finetuning/evaluating foundation models.
- …

Submission Procedure

Interested authors should first submit a short abstract (250 words max) to Krzysztof Janowicz (krzysztof.janowicz@univie.ac.at) and Gengchen Mai (gengchen.mai25@uga.edu) before September 23th, 2023.

Important Dates

- Abstracts (no more than 250 words) Due: **Sep. 23, 2023**
- Decisions on abstracts: **Sep. 30, 2023**
- Full manuscripts Due: **Nov. 30, 2023**

Special Issue Guest Editors

Krzysztof Janowicz, University of Vienna & UC Santa Barbara (krzysztof.janowicz@univie.ac.at)
Gengchen Mai, University of Georgia, Athens, Georgia, USA (gengchen.mai25@uga.edu)
Rui Zhu, University of Bristol, Bristol, UK (rui.zhu@bristol.ac.uk)
Weiming Huang, Nanyang Technological University, Singapore (weiming.huang@ntu.edu.sg)
Ni Lao, Google, Mountain View, CA, USA (nlao@google.com)
Ling Cai, IBM Research, San Jose, CA, USA (lingcai@ibm.com)

JAG Special Issue on Spatially Explicit AI & ML

Spatially Explicit Machine Learning and Artificial Intelligence

Relevant Topics Include

- Spatially Explicit AI for Geospatial Semantics
- Spatially Explicit AI for Remote Sensing
- Spatially Explicit AI for Urban Computing
- Spatially Explicit AI for Earth System Science
- Spatially Explicit AI for Computational Sustainability
- Spatially Explicit AI for Health
- …

Important Dates

- Submission deadline: March 15, 2024

Special Issue Guest Editors

Prof. Gengchen Mai, University of Georgia, USA (gengchen.mai25@uga.edu)

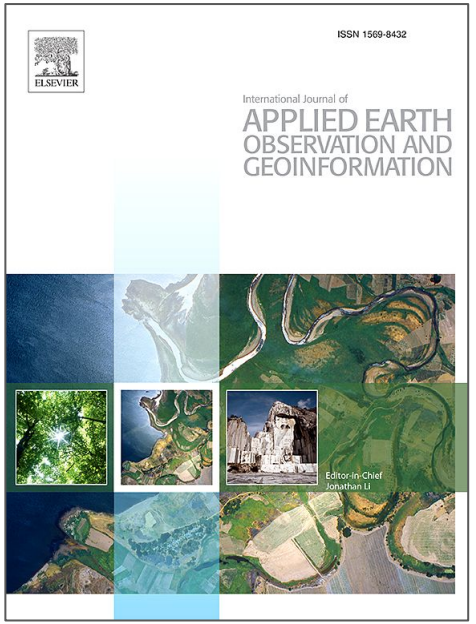
Prof. Xiaobai Angela Yao, University of Georgia, USA (xyao@uga.edu)

Prof. Yao-Yi Chiang, University of Minnesota-Twin Cities, USA, (yaoyi@umn.edu)

Dr. Weiming Huang, Nanyang Technological University, Singapore (weiming.huang@ntu.edu.sg)

Prof. Yiqun Xie, University of Maryland, College Park (xie@umd.edu)

Prof. Rui Zhu, University of Bristol, UK (rui.zhu@bristol.ac.uk)



Reference

1) **Gengchen Mai**, Weiming Huang, Jin Sun, Suhang Song, Deepak Mishra, Ninghao Liu, Song Gao, Tianming Liu, Gao Cong, Yingjie Hu, Chris Cundy, Ziyuan Li, Rui Zhu, Ni Lao. On the Opportunities and Challenges of Foundation Models for Geospatial Artificial Intelligence. arXiv preprint arXiv:2304.06798 (2023).
2) **Gengchen Mai,** Krzysztof Janowicz, Bo Yan, Rui Zhu, Ling Cai, Ni Lao. Multi-Scale Representation Learning for Spatial Feature Distributions using Grid Cells, In: *Proceedings of ICLR 2020*.
3) **Gengchen Mai**, Ni Lao, Yutong He, Jiaming Song, Stefano Ermon. Self-Supervised Contrastive Spatial Pre-Training for Geospatial-Visual Representations, In: *Proceedings of ICML 2023*.
4) **Gengchen Mai**, Yao Xuan, Wenyun Zuo, Yutong He, Jiaming Song, Stefano Ermon, Krzysztof Janowicz, Ni Lao. 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.

UNIVERSITY OF GEORGIA
Single Line Name

Contact

Prof. **Gengchen Mai**

Email: gengchen.mai25@uga.edu

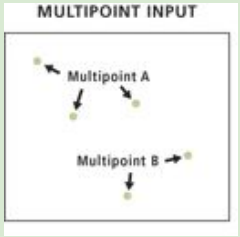
Website: https://gengchenmai.github.io/

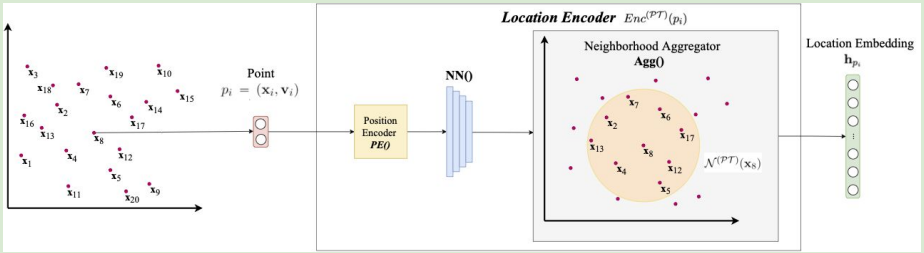
Acknowledgement: NSF, National Institutes of Health, aws, Alfred P. Sloan Foundation, Georgia Department of Geography, Stanford Artificial Intelligence

Spatial Representation Learning

Represent Spatial Data into the Embedding Space (Mai et al., 2023, Handbook of GeoAI)

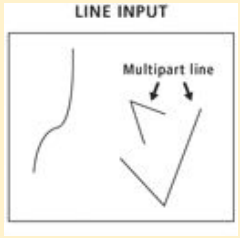
Location Encoding

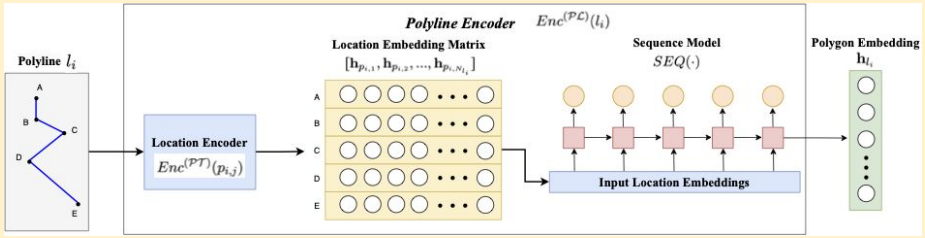




- Qi et al., CVPR 2017
- Mac Aodha et al. CVPR 2019
- Mai et al., ICLR 2020
- Mai et al., IJGIS 2022
- Mai et al., ICML 2023
- Mai et al., ISPRS PHOTO 2023
- …

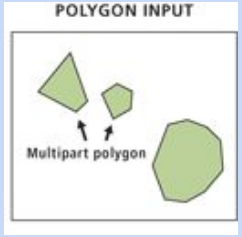
Polyline Encoding

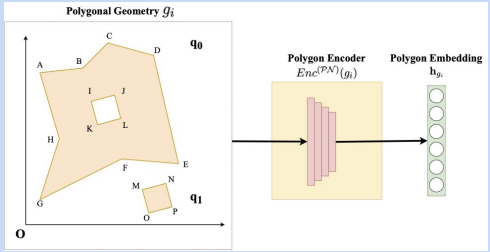




- Alahi et al., CVPR 2016
- Rao et al., GIScience 2020
- Yu et al., IJGIS 2022
- Rempe et al., CVPR 2023
- …

Polygon Encoding





- Jiang et al', ICLR 2019
- Yan et al., ISPRS PHOTO 2021
- Jiang et al., ICCV 2021
- Mai et al., Geoinformatica 2023
- …

Foundation Models

UNIVERSITY OF GEORGIA

Task-Specific Models

Training **specific** models for **specific** tasks

- Question Answering Models
- Machine Translation Models
- Common Sensing Reasoning Models
- Reading Comprehension Models
- Natural Language Inference Models
- Image Classification Models
- Text-to-image Generation Models
- Image Editing Models

Paradigm shift

Foundation Models (FMs)

A large task-agnostic pre-trained model which can be adapted via **fine-tuning** or **few-shot/zero-shot** learning on **a wide range of domains**. (Bommasani et al, 2021)

GPT-3 (Brown et al., 2020)

Few-shot Adaptation

Various NLP Tasks

- Closed Book Question Answering
- Machine Translation
- Common Sense Reasoning
- Reading Comprehension
- Natural Language Inference
- …

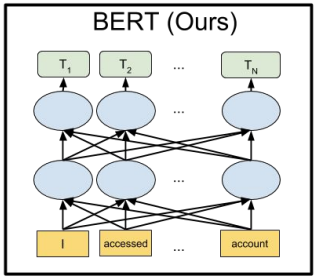
DALL·E 2 (Ramesh et al., 2022)

Zero-shot Transfer

Various CV Tasks

- Text-to-image generation
- Image Completion
- Image Editing
- Style Transfer
- …

Large Language Model



UNIVERSITY OF GEORGIA

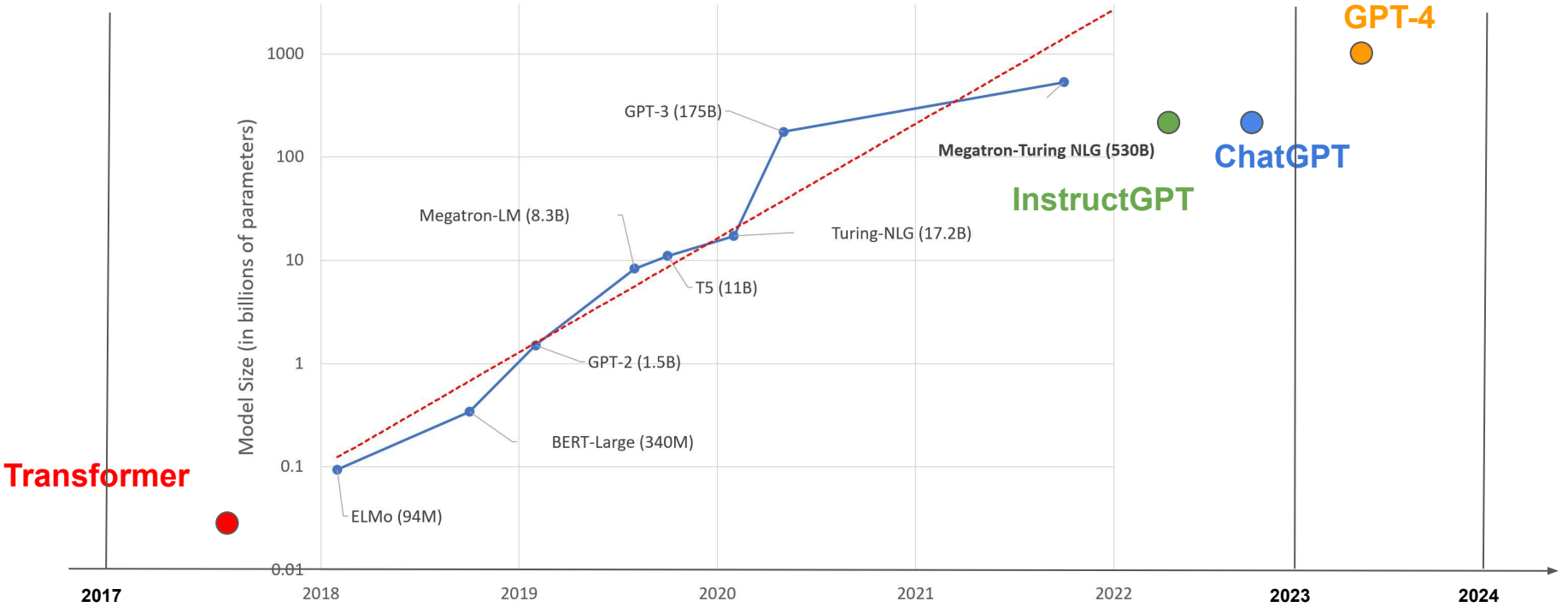
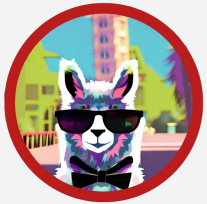


Image from Huggingface (https://huggingface.co/blog/large-language-models)

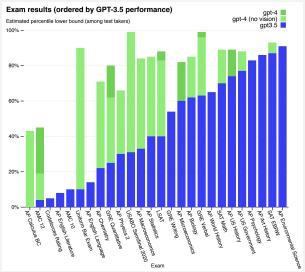
Foundation Models (FMs) in Different Domains

Natural Language Processing

Stanford Alpaca



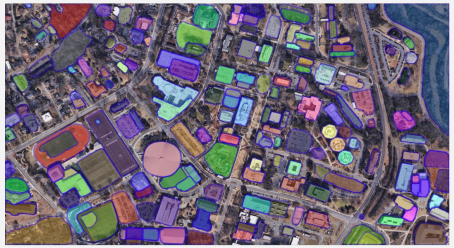
Stanford Alpaca



ChatGPT/GPT-4 (OpenAI. 2023)

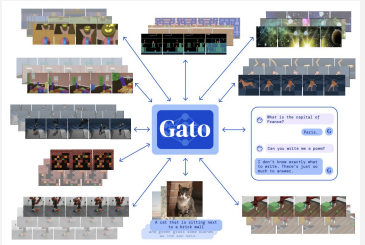
Computer Vision

Imagen (Saharia et al. 2022)



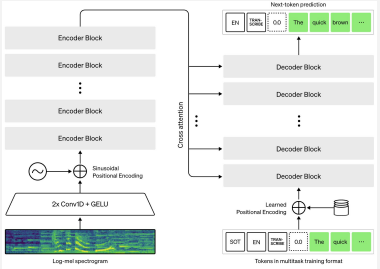
Segment Anying (Kirillov et al, 2023)

Reinforcement Learning



Gato (Reed et al. 2022)

Signal Processing

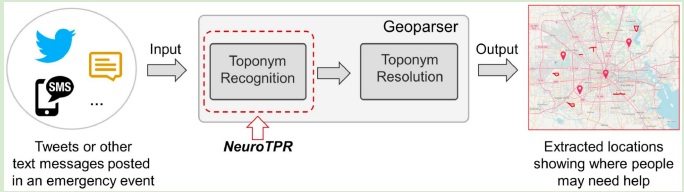


Whisper (Radford et al. 2022)

Applicability of FMs on Geospatial Problems

How do the existing cutting-edge foundation models perform when compared with the state-of-the-art fully supervised task-specific models on various geospatial tasks?

Geospatial Semantics – Topo.Recg.



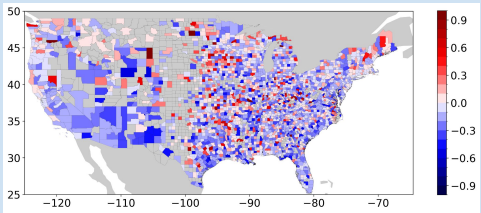
Urban Geography

Urban Function

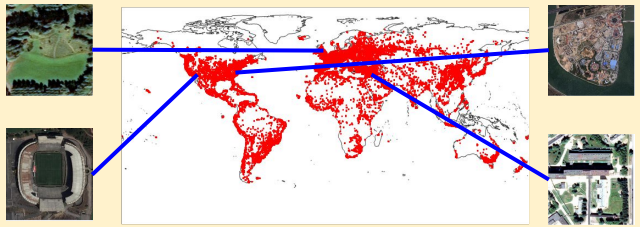
Urban Perception

(a) Very Quiet (0.100)

Health Geography – Dementia Forecast



Remote Sensing – RS Image Clas.



Gengchen Mai, et al. On the Opportunities and Challenges of Foundation Models for Geospatial Artificial Intelligence. arXiv preprint arXiv:2304.06798 (2023). ACM TSAS (Under Review)

Applicability of FMs on Geospatial Semantics

Toponym Recognition: extract place names (e.g., cities, counties, states, countries) from text

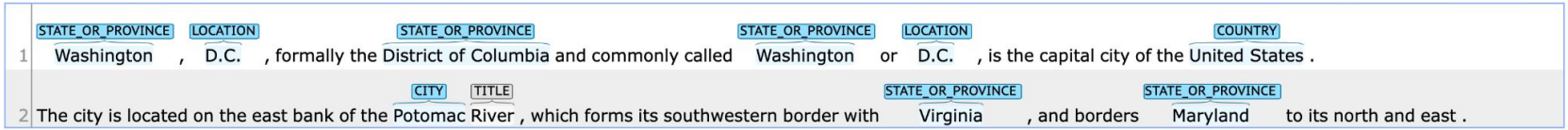
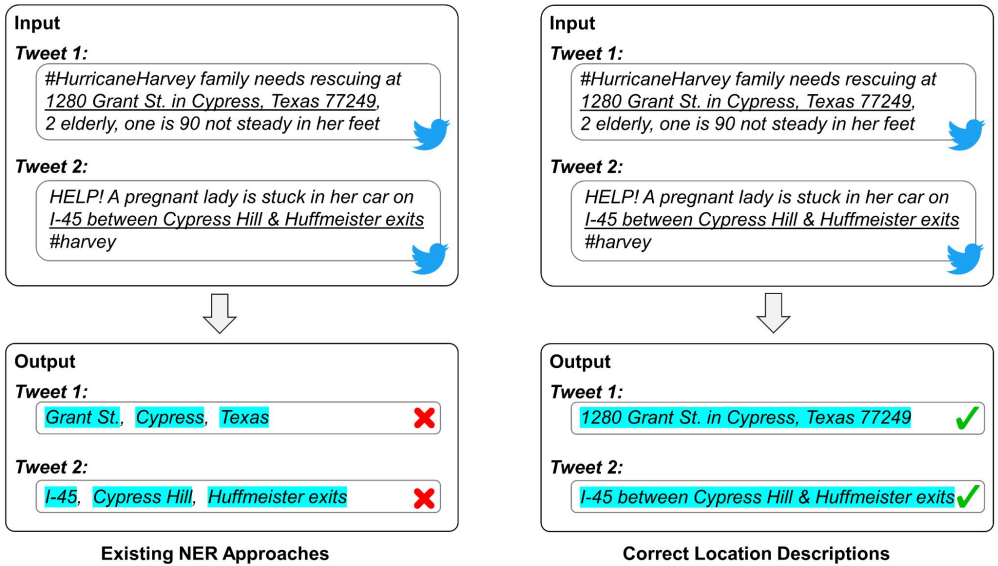


*figure from CoreNLP

Location Description Recognition: extract multi-entity location descriptions from text



(Mai et al., 2022; Mai et al 2023)

Applicability of FMs on Geospatial Semantics

- Investigate the performance of **GPT-3 & ChatGPT** on some well established **geospatial semantic tasks**:

Toponym Recognition

```
[Instruction] ...
Paragraph: Alabama State Troopers say a Greenville man has died of his injuries
    ↪ after being hit by a pickup truck on Interstate 65 in Lowndes County.
Q: Which words in this paragraph represent named places?
A: Alabama; Greenville; Lowndes
...
--
Paragraph: The Town of Washington is to what Williamsburg is to Virginia.
Q: Which words in this paragraph represent named places?
A: Washington; Williamsburg; Virginia
```

*toponyms: proper names of places, also known as place names and geographic names.

Location Description Recognition

```
[Instruction] ...
Paragraph: Papa stranded in home. Water rising above waist. HELP 8111 Woodlyn Rd
    ↪ , 77028 #houstonflood
Q: Which words in this paragraph represent location descriptions?
A: 8111 Woodlyn Rd, 77028
...
--
Paragraph: HurricaneHarvey Help Need AT 7506 Jackrabbit Rd, Houston, TX 77095.
Q: Which words in this paragraph represent location descriptions?
A: 7506 Jackrabbit Rd, Houston, TX 77095
```

Applicability of FMs on Geospatial Semantics

Task 1 & 2: Toponym Recognition & Location Description Recognition

- Toponym recognition: FMs (e.g., GPT-2/3) consistently outperform the **fully-supervised** baselines with only **8 few-shot** examples
- Location Description Recognition: GPT-3 achieves the best Recall score across all methods

			Toponym Recognition		Location Description Recognition		
	Model	#Param	Hu2014	Ju2016	HaveyTweet2017		
			Accuracy ↓	Accuracy ↓	Precision ↓	Recall ↓	F-Score ↓
(A)	Stanford NER (nar. loc.) [30]	-	0.787	0.010	**0.828**	0.399	0.539
	Stanford NER (bro. loc.) [30]	-	-	0.012	0.729	0.44	0.548
	Retrained Stanford NER [30]	-	-	0.078	0.604	0.410	0.489
	Caseless Stanford NER (nar. loc.) [30]	-	-	0.460	0.803	0.320	0.458
	Caseless Stanford NER (bro. loc.) [30]	-	-	0.514	0.721	0.336	0.460
	spaCy NER (nar. loc.) [44]	-	0.681	0.000	0.575	0.024	0.046
	spaCy NER (bro. loc.) [44]	-	-	0.006	0.461	0.304	0.366
	DBpedia Spotlight[99]	-	0.688	0.447	-	-	-
(B)	Edinburgh [7]	-	0.656	0.000	-	-	-
	CLAVIN [134]	-	0.650	0.000	-	-	-
	TopoCluster [23]	-	0.794	0.158	-	-	-
(C)	CamCoder [33]	-	0.637	0.004	-	-	-
	Basic BiLSTM+CRF [77]	-	-	0.595	0.703	0.600	0.649
	DM NLP (top. rec.) [139]	-	-	0.723	0.729	0.680	0.703
	NeuroTPR [135]	-	0.675†	0.821	0.787	0.678	**0.728**
(D)	GPT2 [115]	117M	0.556	0.650	0.540	0.413	0.468
	GPT2-Medium [115]	345M	0.806	0.802	0.529	0.503	0.515
	GPT2-Large [115]	774M	0.813	0.779	0.598	0.458	0.518
	GPT2-XL [115]	1558M	0.869	**0.846**	0.492	0.470	0.481
	GPT-3 [15]	175B	**0.881**	0.811*	0.603	**0.724**	0.658
	InstructGPT [106]	175B	0.863	0.817*	0.567	0.688	0.622
	ChatGPT (Raw.) [104]	176B	0.800	0.696*	0.516	0.654	0.577
	ChatGPT (Con.) [104]	176B	0.806	0.656*	0.548	0.665	0.601

Applicability of FMs on Health Geography

UNIVERSITY OF GEORGIA

Task 3: US County-Level Dementia Time Series Forecasting

```
[Instruction] This is a set of time series forecasting problems.
The `Paragraph` is a time series of the numbers of deaths from
    ↪ alzheimer's disease for one of US counties from 1999 to 2019.
The goal is to predict the number of deaths from alzheimer's disease at
    ↪  this county in 2020. Please give a single number as the
    ↪ prediction.
--
--
Paragraph:  At Santa Barbara County, CA, from 1999 to 2019, the numbers
    ↪ of deaths from alzheimer's disease are
    ↪  126 in 1999, 114 in 2000, 124 in 2001, 127 in 2002, 156 in 2003,
    ↪   154 in 2004, 175 in2005, 172 in 2006, 171 in 2007, 248 in 2008, 204
    ↪  in 2009, 241 in 2010, 260 in 2011, 297 in 2012, 283 in 2013, 308 in
    ↪  2014, 358 in 2015, 365 in 2016, 334 in 2017, 363 in 2018,
    ↪  and 328 in 2019.
Q:  Please forecast the number in 2020 at Santa Barbara County, CA?
A:  345
```

Listing 4. US county-level Alzimier time series forecasting with LLMs by zero-shot learning. Yellow block: the historical time series data of one US county. Orange box: the outputs of InstructGPT. Here, we use Santa Barbara County, CA as an example and the correct answer is 373.

Table 3. Evaluation results of various GPT models and baselines on the US county-level dementia time series forecasting task. We use same model set and evaluation metrics as Table 2.

	Model	#Param	MSE ↓	MAE ↓	MAPE ↓	R^2 ↑
(A) Simple	Persistence [103, 107]	-	1,648	16.9	0.189	0.979
(B) Supervised ML	ARIMA [58]	-	1,133	15.1	0.193	0.986
(C) Zero shot LLMs	GPT2 [115]	117M	77,529	92.0	0.587	-0.018
	GPT2-Medium [115]	345M	226,259	108.1	0.611	-2.824
	GPT2-Large [115]	774M	211,881	94.3	0.581	-1.706
	GPT2-XL [115]	1558M	162,778	99.8	0.627	-1.082
	GPT-3 [15]	175B	1,105	14.5	0.180	0.986
	InstructGPT [106]	175B	**831**	**13.3**	**0.179**	**0.989**
	ChatGPT (Raw.) [104]	176B	4,115	23.2	0.217	0.955
	ChatGPT (Con.) [104]	176B	3,402	20.7	0.231	0.944

Applicability of FMs on Health Geography

UNIVERSITY OF GEORGIA 1785

US County-Level Dementia Time Series Forecasting: Prediction Error Maps

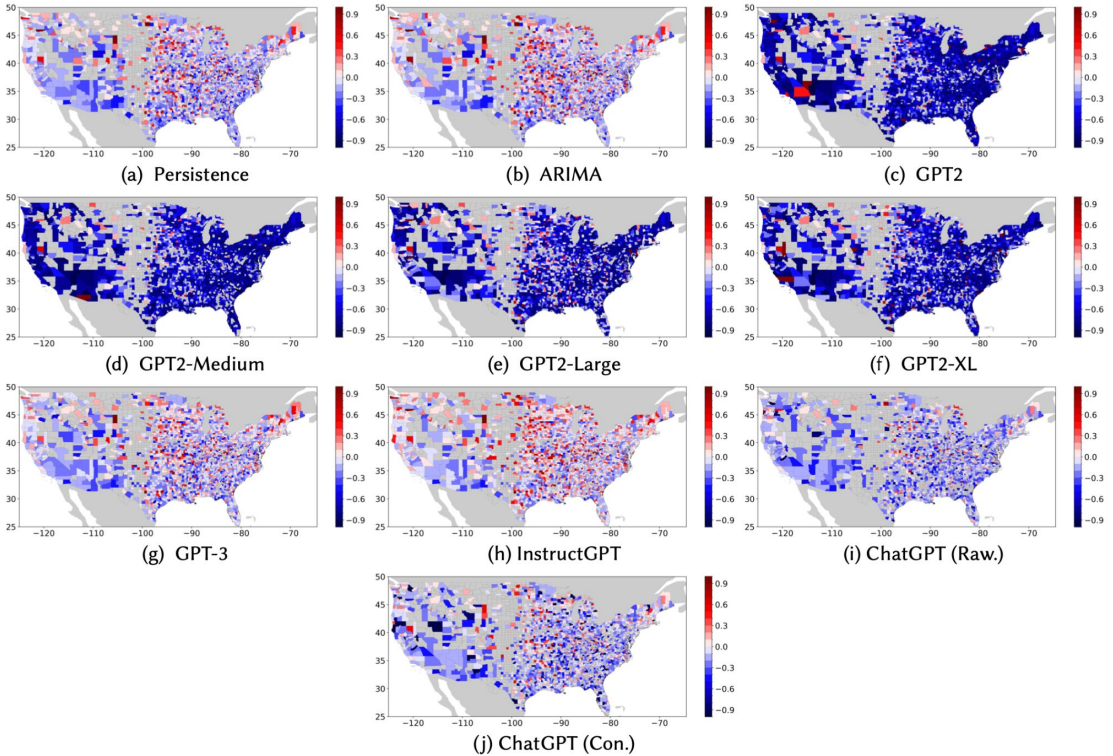


Fig. 1. Prediction error maps of each baseline and GPT model on US county-level dementia death count time series forecasting task. The color on each US count indicates the percentage error $PE = (Prediction - Label)/Label$ of each model prediction on this county. Those counties in gray color indicate their dementia data during 1999 and 2020 are not available.

Applicability of FMs on Urban Geography

UNIVERSITY OF GEORGIA 1785

Task 4: POI-Based Urban Function Classification

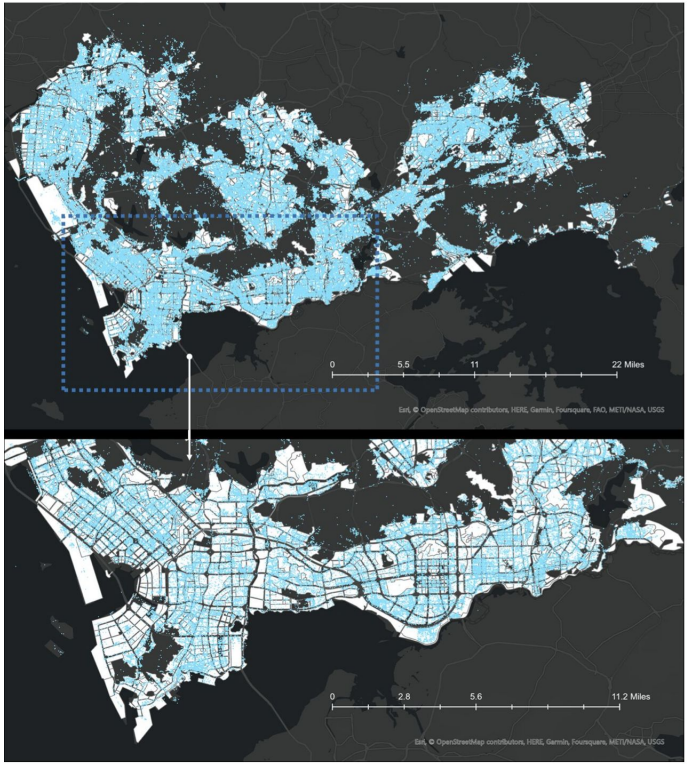


Fig. 2. The spatial distributions of POI data in the *UrbanPOI5K* dataset.

```
[Instruction] There are six land use types: (1) residential, (2) commercial, (3) industrial, (4) education, health care, civic, governmental
    ↪ and cultural, (5) transportation facilities, and (6) outdoors and natural.
Paragraph: In this urban region, there are 128 points of interest, including 2 Chinese restaurant, 1 food restaurant, 2 hotel, 2 apartment
    ↪ hotel, 1 daily life service, 1 mobile communication shop, 24 company, 1 logistics company, 1 real estate agency, 1 lottery retailer,
    ↪ 3 beauty shop, 1 manicure, 2 barber shop, 4 Internet cafe, 3 bath massage, 2 stadium, 4 training institutions, 1 pharmacy, 4
    ↪ automative sale, 6 car service, 2 car repair, 1 Car rental, 1 Automobile parts, 3 shopping, 5 shop, 5 parking lot, 5 Parking lot
    ↪ entrance, 2 transportation facility, 1 port harbor, 1 road intersection, 1 atm machine, 2 office building, 2 residential area, 7
    ↪ building, 1 real estate, 1 park, 1 factory, 7 administrative agency, 1 entrance and exit, 3 gate door, 6 convenience store, 4 home
    ↪ building materials.
Q: What is the primary land use category of this urban region?
A: outdoors and natural

Paragraph: In this urban region, there are 17 points of interest, including 1 food restaurant, 3 public toilet, 3 funeral service, 2 road station for walking and cycling,
    ↪ 1 beach, 2 parking lot, 2 road intersection, 1 corporate company enterprise, 2 administrative agency.
Q: What is the primary land use category of this urban region?
A: outdoors and natural
```

Listing 5. POI-based urban function classification with LLMs, e.g., ChatGPT (Raw.). Yellow block: the POI statistic of a new urban neighborhood to be classified. Orange box: ChatGPT (Raw.) outputs.

Table 5. Evaluation results of various GPT models and supervised baseline on the *UrbanPOI5K* dataset for the POI-based urban function classification task. We divide the models into three groups: (A) supervised learning-based neural network models; (B) Zero-shot learning with LLMs. (C) One-shot learning with LLMs. We use accuracy, weighted precision, and weighted recall as evaluation metrics. We do not include weighted F1 scores since it is the same as the accuracy score. The best model of each group is highlighted.

	Model	Accuracy	Precision	Recall
(A) Supervised NN	Place2Vec [145, 152]	0.540	0.512	0.516
	HGI [52]	**0.584**	**0.568**	**0.563**
(B) Zero-shot LLMs	GPT2 [115]	**0.318**	0.105	**0.158**
	GPT2-Medium [115]	0.025	0.102	0.040
	GPT2-Large [115]	0.005	0.001	0.002
	GPT2-XL [115]	0.001	0.108	0.002
	GPT-3 [15]	0.144	**0.448**	0.141
	ChatGPT (Raw.) [104]	0.075	0.376	0.106
	ChatGPT (Con.) [104]	0.051	0.232	0.046
(C) One-shot LLMs	GPT2 [115]	0.149	0.079	0.085
	GPT2-Medium [115]	0.317	0.104	0.156
	GPT2-Large [115]	0.057	0.083	0.021
	GPT2-XL [115]	**0.324**	0.105	0.159
	GPT-3 [15]	0.176	0.486	0.190
	ChatGPT (Raw.) [104]	0.195	**0.524**	**0.245**
	ChatGPT (Con.) [104]	0.093	0.451	0.085

Applicability of FMs on Urban Geography

UNIVERSITY OF GEORGIA

Task 5: Street View Image-Based Urban Noise Intensity Classification

(a) Very Quiet (0.100) (b) Quiet (0.350) (c) Noisy (0.600) (d) Very Noisy (0.965)

Fig. 6. Some street view image examples in *SingaporeSVI579* dataset. The image caption indicates the noise intensity class this image belongs to and the numbers in parenthesis indicate the original noise intensity scores from Zhao et al. [162].

Table 6. Evaluation results of various vision-language foundation models and baselines on the urban street view image-based noise intensity classification dataset, SingaporeSVI579 [162]. We classify models into two groups: (A) Supervised finetuned convolutional neural networks (CNNs); (B) Zero-shot learning with visual-language foundation models (VLFMs). We use accuracy and weighted F1 scores as evaluation metrics. The best scores for each group are highlighted.

	Model	#Param	Accuracy	F1
(A) Supervised Finetuned CNNs	AlexNet [74]	58M	0.452	0.405
	ResNet18 [37]	11M	0.493	**0.442**
	ResNet50 [37]	24M	**0.500**	0.436
	DenseNet161 [48]	27M	0.486	0.382
(B) Zero-shot FMs	OpenCLIP-L [54, 113, 127]	427M	0.128	0.089
	OpenCLIP-B [54, 113, 127]	2.5B	0.169	0.178
	BLIP [81, 82]	3.9B	**0.452**	**0.405**
	OpenFlamingo-9B [11]	8.3B	0.262	0.127

Applicability of FMs on Remote Sensing

UNIVERSITY OF GEORGIA 1785

Task 6: Remote Sensing Image Scene Classification

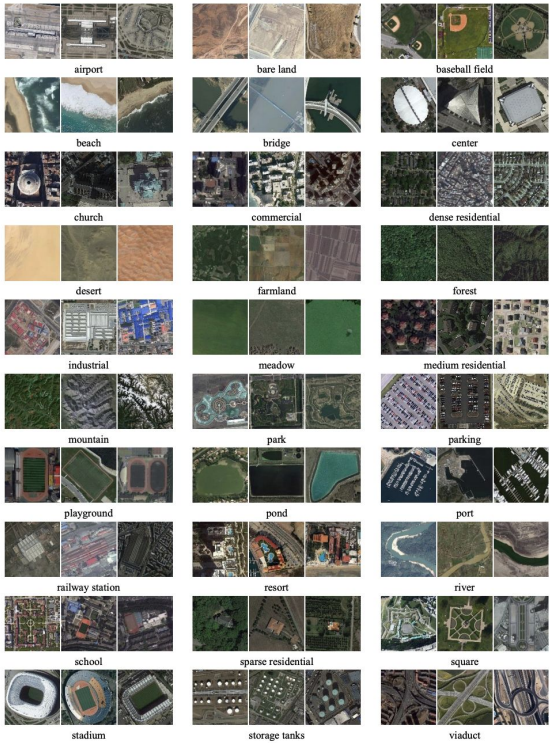


Figure 1: Samples of AID: three examples of each semantic scene class are shown. There are 10000 images within 30 classes.

Figure from Xia et al. (2016)

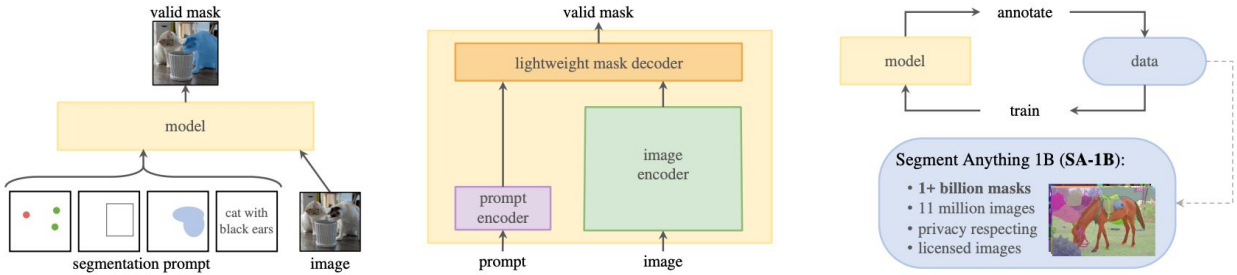
Table 7. Evaluation results of various vision-language foundation models and baselines on the remote sensing image scene classification dataset, *AID* [144]. We use the same model set as Table 6. "(*Origin*)" denotes we use the original remote sensing image scene class name from *AID* to populate the prompt while "(*Updated*)"indicates we update some class names to improve its semantic interpretation for FMs. We use accuracy and F1 score as evaluation metrics.

	Model	#Param	Accuracy	F1
Supervised Finetuned CNNs	AlexNet [74]	58M	**0.831**	**0.827**
	ResNet18 [37]	11M	0.752	0.730
	ResNet50 [37]	24M	0.757	0.738
	DenseNet161 [48]	27M	0.818	0.807
Zero-shot FMs	OpenCLIP-L (*Origin*) [54, 113, 127]	427M	0.708	0.688
	OpenCLIP-L (*Updated*) [54, 113, 127]	427M	**0.710**	**0.698**
	OpenCLIP-B (*Origin*) [54, 113, 127]	2.5B	0.699	0.668
	OpenCLIP-B (*Updated*) [54, 113, 127]	2.5B	0.705	0.686
	BLIP (*Origin*) [82]	2.5B	0.500	0.473
	BLIP (*Updated*) [82]	2.5B	0.520	0.494
	OpenFlamingo-9B [11]	8.3B	0.206	0.154

Applicability of FMs on Remote Sensing

Task 7: Remote Sensing Semantic Segmentation

- See the original SAM demo video here: https://segment-anything.com/assets/section-1.1b.mp4
- SAM (Segment Anything Model) has following limitations:
  - SAM works best for points and box prompts. Currently, it does not have a robust way to handle text prompt inputs.
  - A naive usage of SAM will produce a set of polygon masks for everything.
  - SAM only produces masks but not label for each mask which is required for semantic segmentation



(a) **Task**: promptable segmentation (b) **Model**: Segment Anything Model (**SAM**) (c) **Data**: data engine (top) & dataset (bottom)

Figure 1: We aim to build a foundation model for segmentation by introducing three interconnected components: a promptable segmentation *task*, a segmentation *model* (SAM) that powers data annotation and enables zero-shot transfer to a range of tasks via prompt engineering, and a *data* engine for collecting SA-1B, our dataset of over 1 billion masks.

Kirillov, Alexander, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao et al. "Segment anything." arXiv preprint arXiv:2304.02643 (2023).

Applicability of FMs on Remote Sensing

Task 7: Remote Sensing Semantic Segmentation

A pipeline that **leverages multiple FMs** to facilitate remote sensing image semantic segmentation tasks

Figure 1: The overall structure of our pipeline consists of three methods for guiding the SAM model. First, a text prompt is used as input for Grounding DINO, which generates bounding boxes. These bounding boxes are then input into SAM to produce a segmentation map. Second, a text prompt is input into CLIP Surgery, yielding a heatmap. This heatmap is sampled to create point prompts for SAM, which then generates segmentation masks. Lastly, we first utilize SAM to generate all available segmentation maps, and then employ CLIP to compare their semantic similarity with the text prompt.

Jielu Zhang, Zhongliang Zhou, **Gengchen Mai**, Lan Mu, Mengxuan Hu, Sheng Li. Text2Seg: Remote Sensing Image Semantic Segmentation via Text-Guided Visual Foundation Models. arXiv preprint arXiv:2304.10597.

Applicability of FMs on Remote Sensing

Task 7: Remote Sensing Semantic Segmentation

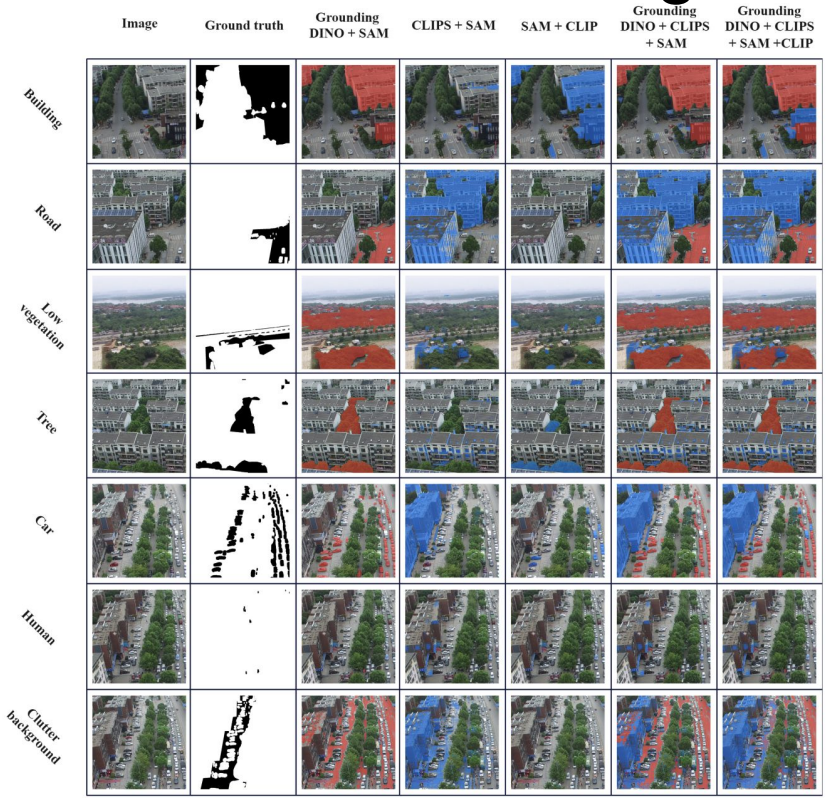
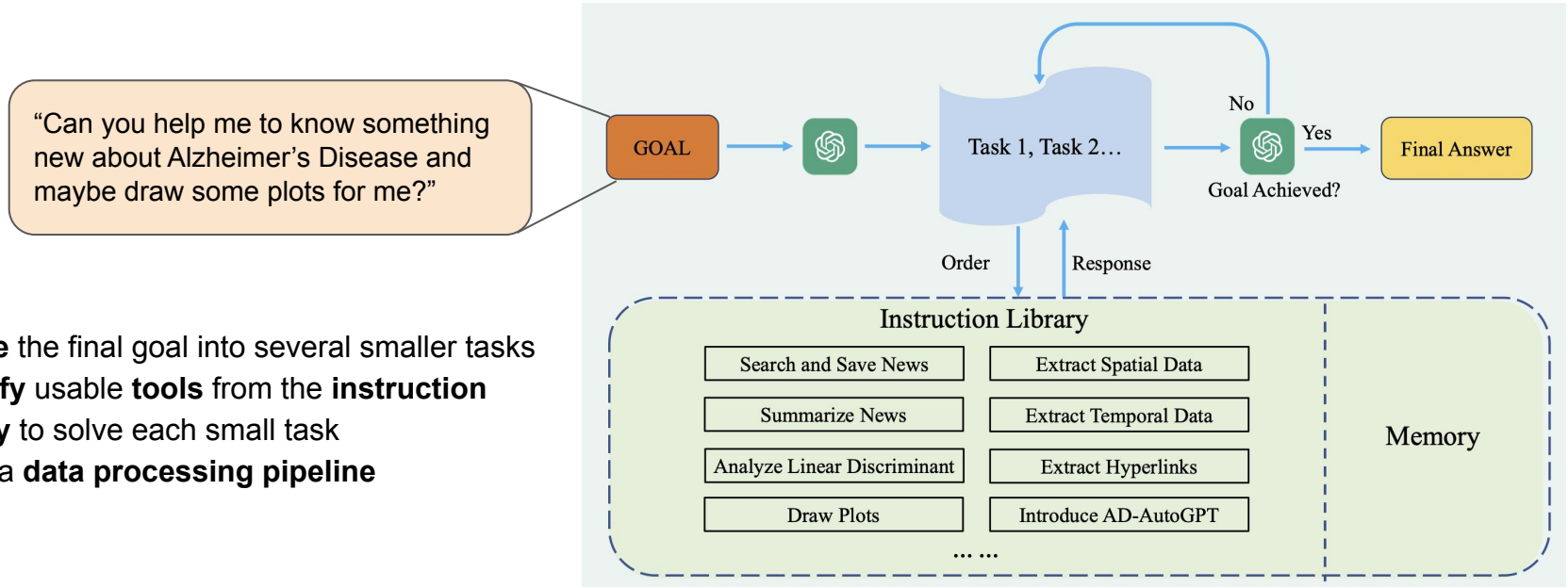


Figure 3: Result for UAVid dataset. From left to right, we show the original images, ground truth label with the object highlighted by black, Grounding DINO with SAM(Grounding DINO+SAM, results are marked as red), CLIP Surgery with SAM(CLPS+SAM, results are marked as blue), SAM with CLIP filter(SAM+CLIP, results are marked as blue), Grounding DINO with CLIP Surgery and SAM(Grounding DINO+CLIPS+SAM, results are marked as red and blue), Grounding DINO with CLIP Surgery, SAM and CLIP filter(Grounding DINO+CLIPS+SAM+CLIP, results are marked as red and blue). The original masks are semantically labeled.

Autonomous GPT for Alzheimer's Disease Infodemiology

AD-AutoGPT: A **GPT-4 based AI Assistant tool** which can conduct data collection, processing, and analysis about complex health narratives of **Alzheimer's Disease** in an **autonomous manner** via users' textual prompts.

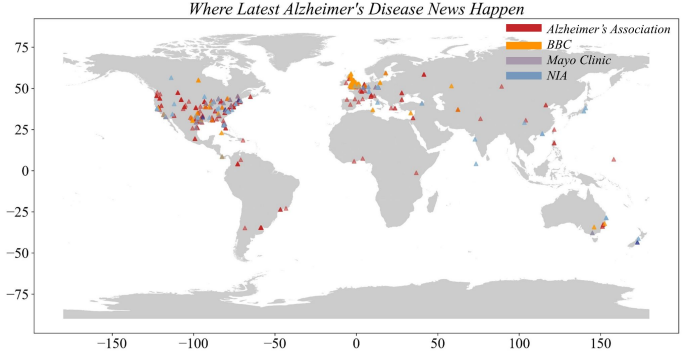


- **Divide** the final goal into several smaller tasks
- **Identify** usable **tools** from the **instruction library** to solve each small task
- Form a **data processing pipeline**

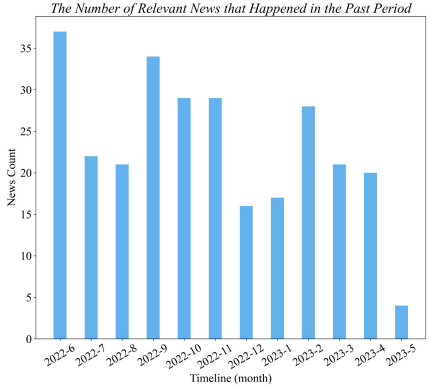
Haixing Dai, et al. AD-AutoGPT: Harnessing AutoGPT for Alzheimer's Disease Infodemiology, (Under review)

Autonomous GPT for Alzheimer's Disease Infodemiology

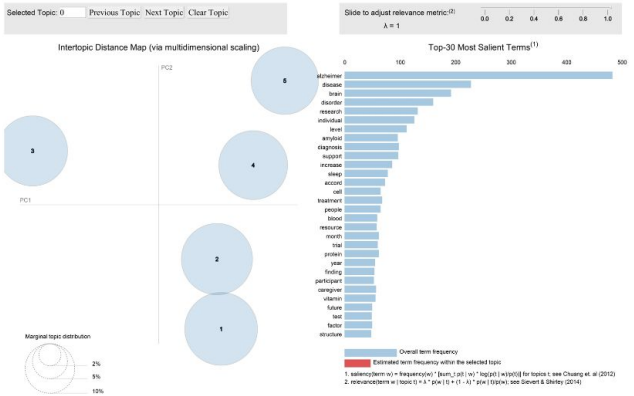
AD-AutoGPT is able to automatically: 1) search in Google; 2) save news articles; 3) extract spatiotemporal info; 4) do LDA topic modeling; 5) visualize analysis results.



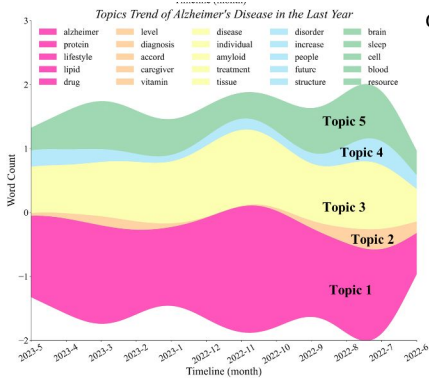
(a) The spatial distribution of extract places



(b) The news count per months



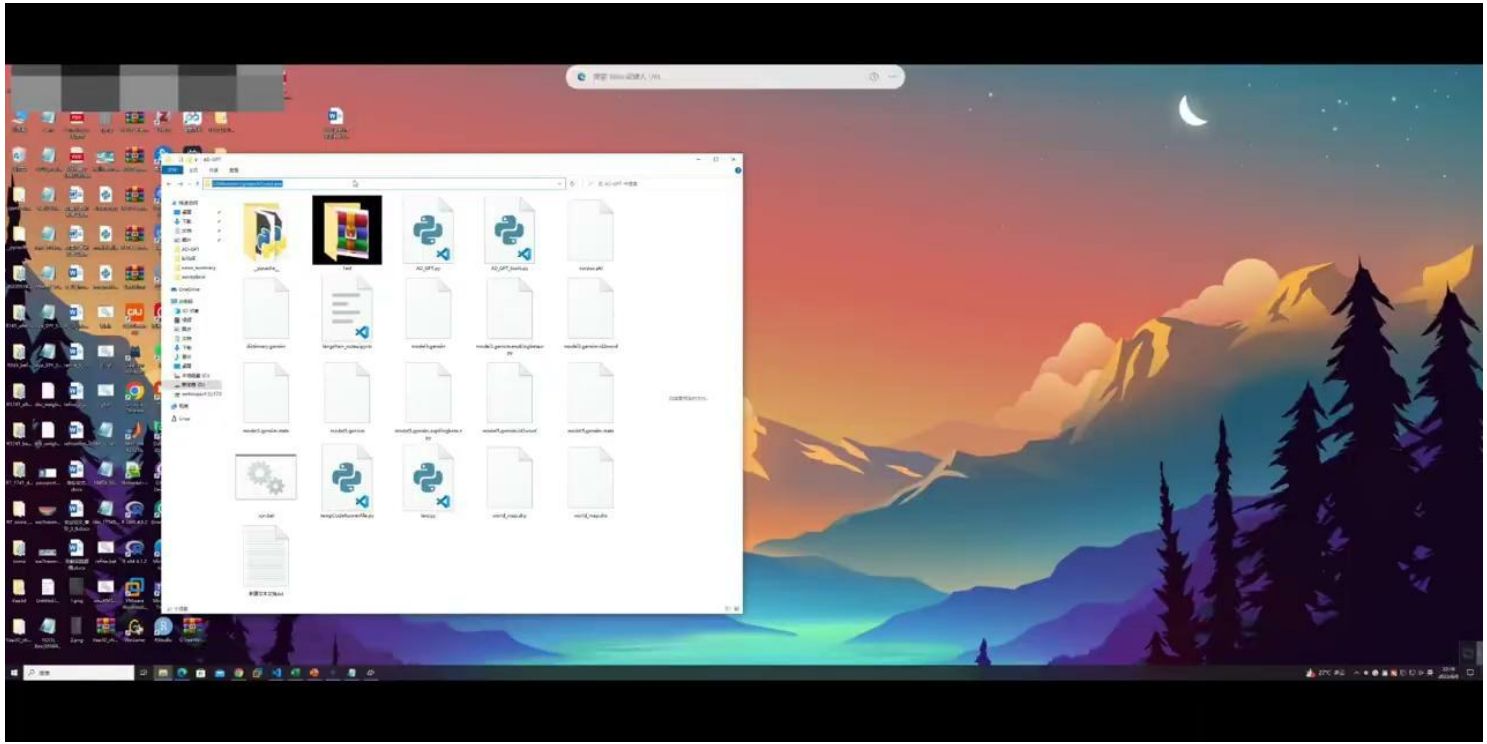
(c) LDA topic modeling on all news



(d) A streamgraph of the topic trend over time

Autonomous GPT for Alzheimer's Disease Infodemiology

AD-AutoGPT demo



Unique Challenges of GeoAI for FMs

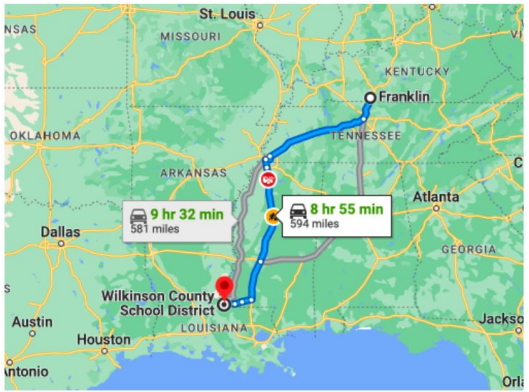
- **Shortcoming of language FMs**: by design they are unable to handle **geo-coordinates**.
- Thus they cannot perform **implicit spatial reasoning** in a way that is grounded in the real world
  - The predicted coordinates are not accurate

Geoparsing

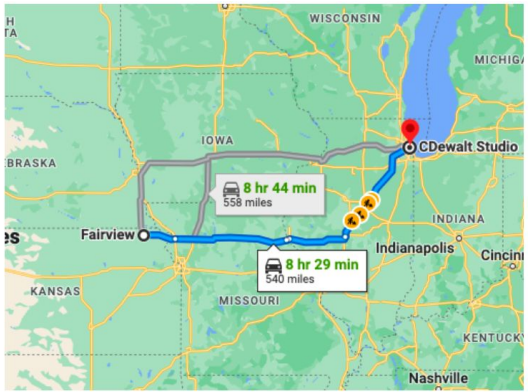
[Instruction] ...
Paragraph: San Jose was founded in 1803 when allotments of land were made ...
Q: Which words in this paragraph represent named places?
A: San Jose; New Mexico

Q: What is the location of San Jose?
A: 35.39728, -105.47501
...
--
Paragraph: the city of fairview had a population of 260 as of july 1, 2015. ...
Q: Which words in this paragraph represent named places?
A: Fairview

Q: What is the location of Fairview?
A: 41.85003, -87.65005



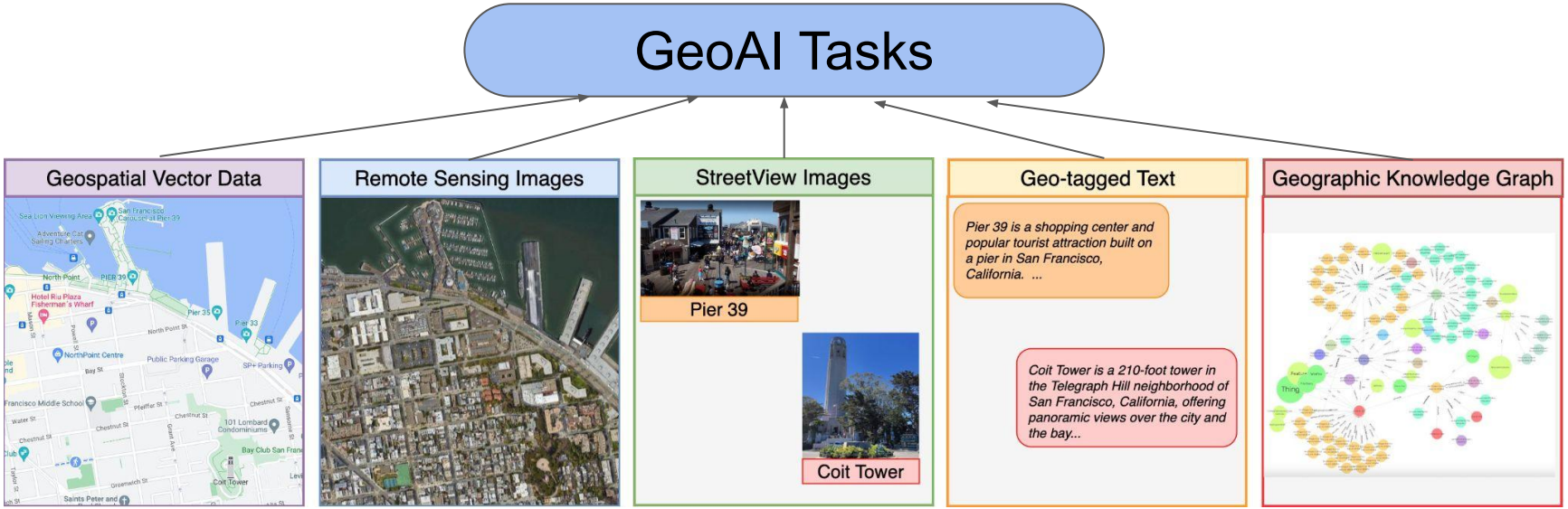
(a) [TEXT]: Franklin is a city in and the county seat of simpson county, ...



(b) [TEXT]: the city of Fairview had a population of 260 as of july 1, 2015. ...

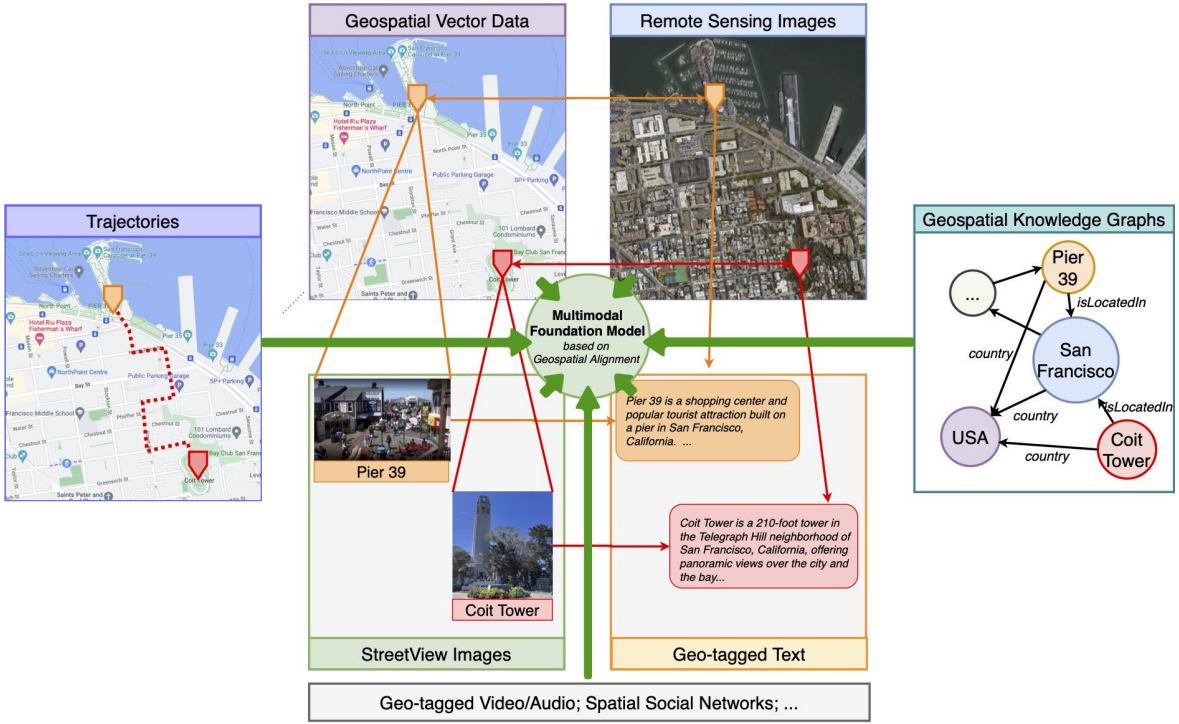
Unique Challenges of GeoAI for FMs

- **Uniqueness of GeoAI Tasks: many data modalities** which calls for **multimodal approaches**



A Multimodal FM for GeoAI

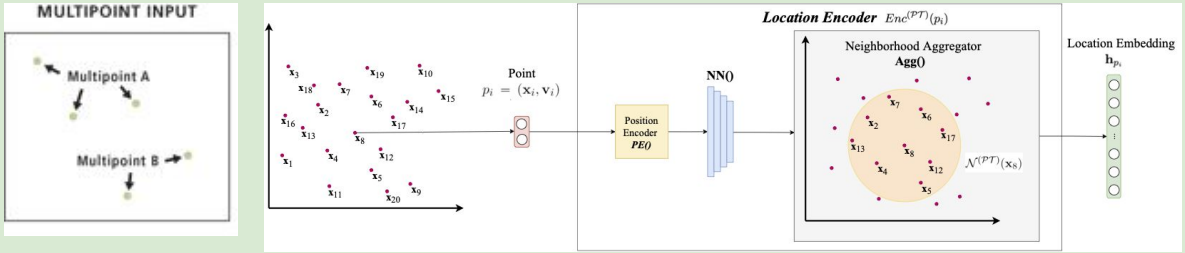
Vision: a multimodal FM for GeoAI that use their **geospatial relationships as alignments** among **different data modalities.**



Spatial Representation Learning

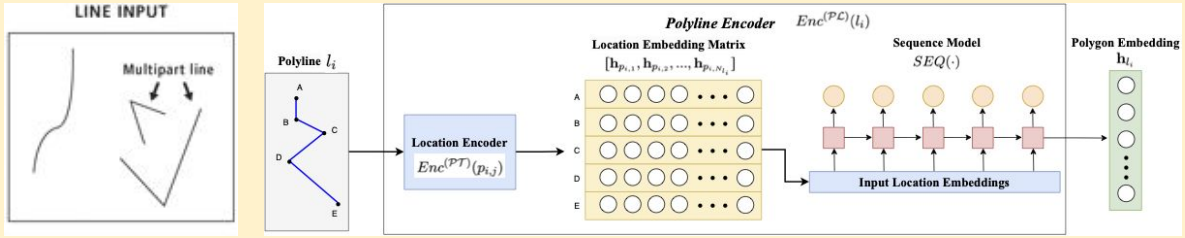
Represent Spatial Data into the Embedding Space (Mai et al., 2023, Handbook of GeoAI)

Location Encoding



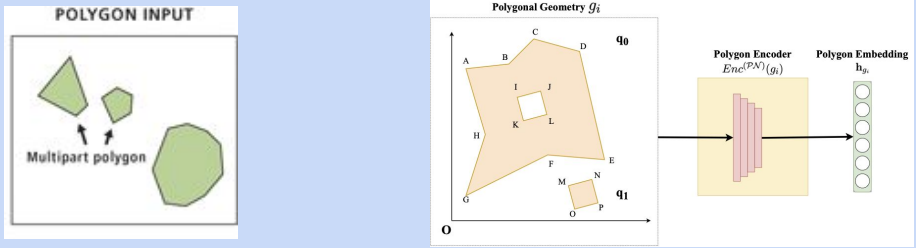
- Qi et al., CVPR 2017
- Mac Aodha et al. CVPR 2019
- Mai et al., ICLR 2020
- Mai et al., IJGIS 2022
- Mai et al., ICML 2023
- Mai et al., ISPRS PHOTO 2023
- …

Polyline Encoding



- Alahi et al., CVPR 2016
- Rao et al., GIScience 2020
- Yu et al., IJGIS 2022
- Rempe et al., CVPR 2023
- …

Polygon Encoding



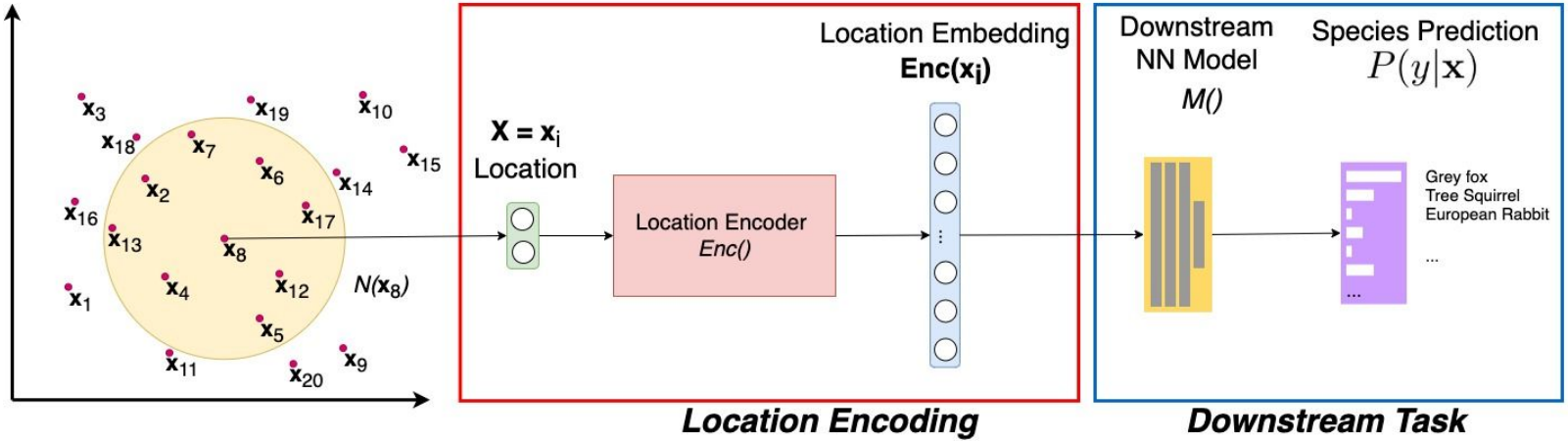
- Jiang et al', ICLR 2019
- Yan et al., ISPRS PHOTO 2021
- Jiang et al., ICCV 2021
- Mai et al., Geoinformatica 2023
- …

Location Encoding

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

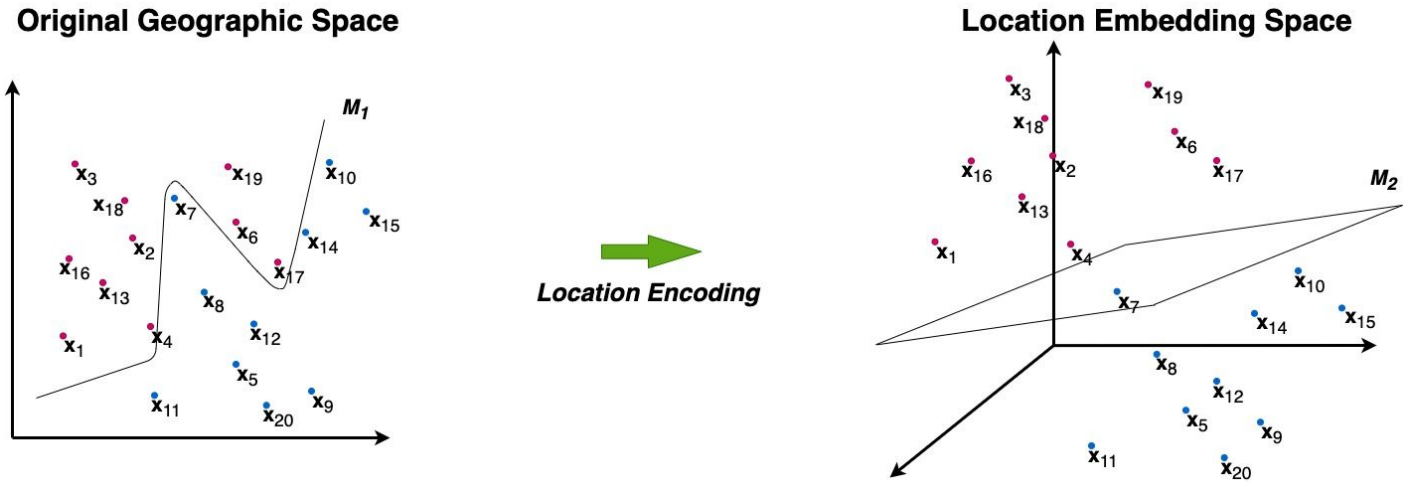
Expected Properties:

- **Preserve** different spatial information (e.g., **distance**, **direction**)
- **Learning-friendly** for downstream models (e.g., deep neural network)



Why we need Location Encoding?

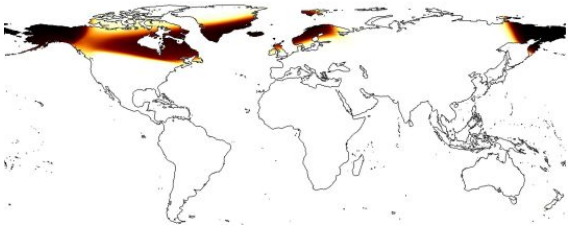
Simple explanation: do **feature decomposition** to produce **learning friendly representations** of geographic locations



Mai, G., et al, 2022. **A Review of Location Encoding for GeoAI: Methods and Applications**. In International Journal of Geographical Information Science, pp.1-35

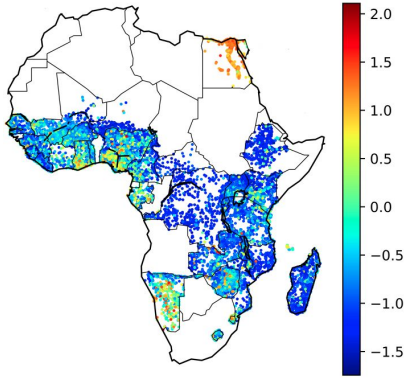
Various Geospatial Tasks

Ecology:
Species Distribution Modeling
(Mac Aodha et al., 2019; Mai et al., 2020)

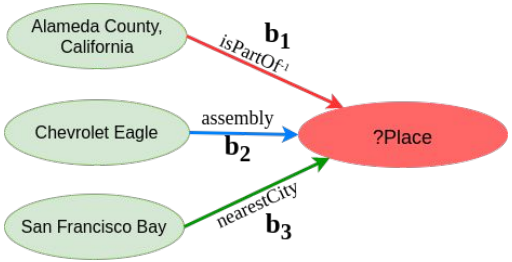


Climate Science:
Precipitation Prediction

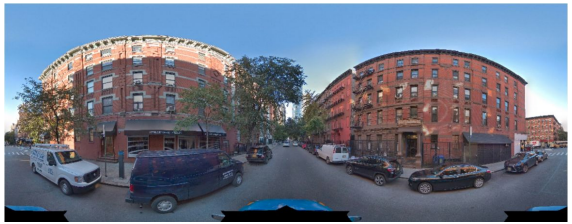
Demography:
Wealth Index Prediction
(Sheehan et al., 2019)



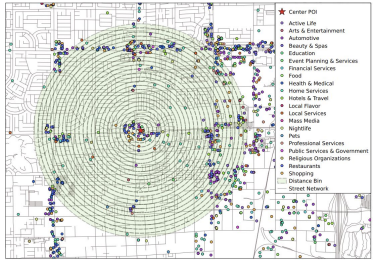
Geospatial Semantics:
Geographic Question Answering
(Mai et al., 2020)



Smart City:
Indoor/outdoor Navigation
(Mehta et al., 2020)

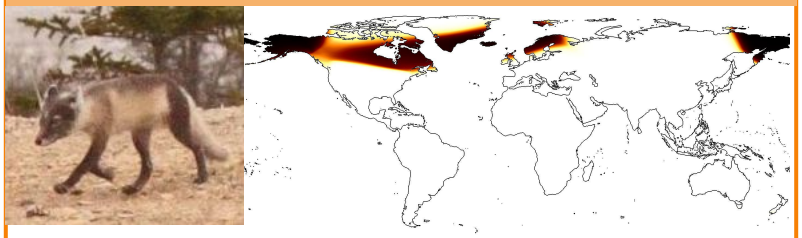


Urban Data Science:
POI Type Prediction
(Mai et al., 2020)



Species Fine-Grained Recognition: Locations Matter

Arctic Fox (North Pole)

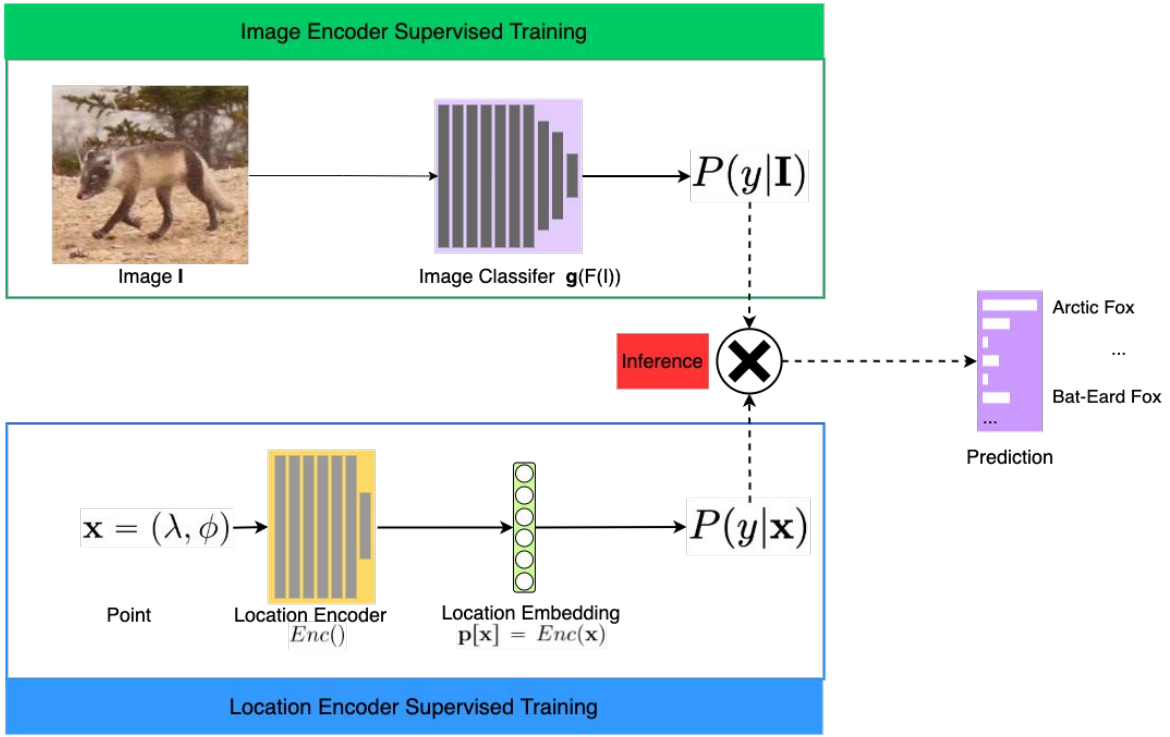


Bat-Eared Fox (South Africa)

Why -- Use image **locations** in the image classification model

Geo-Aware Fine-Grained Species Recognition

The Model Architecture:

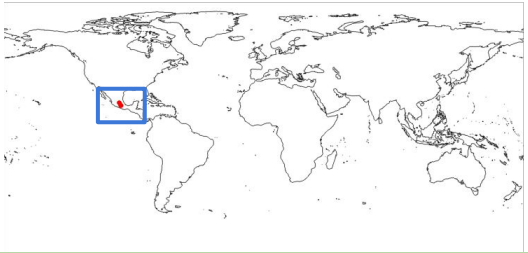


Mai, G., et al, 2020. Multi-Scale Representation Learning for Spatial Feature Distributions using Grid Cells. In **ICLR 2020**

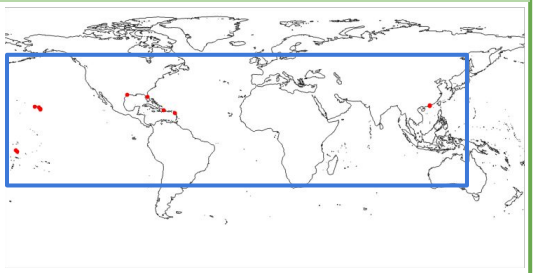
***Spotlight paper (Acceptance Rate 6%).**

The key challenges of location encoding

Morning Glory (Clustered)



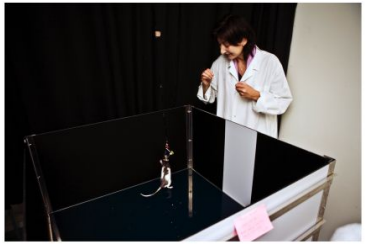
Wedelia (Wide Spread)

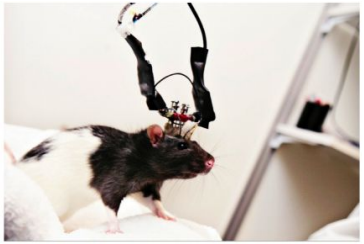


How -- Use **multi-scale representations** to joint model different distributions

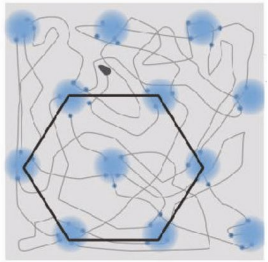
Grid Cell Based Multi-Scale Location Encoding

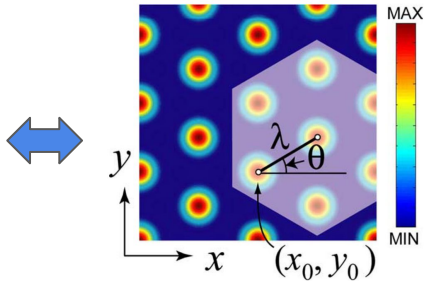
Grid cells in mammals provide a **multi-scale periodic representation** that functions as a metric for location encoding. (Banino et al., 2018)





It can be simulated by summing **three cosine grating functions** oriented 60 degree apart (**a simple Fourier model of the hexagonal lattice**). (Blair et al. 2007)





Space2Vec Location Encoder

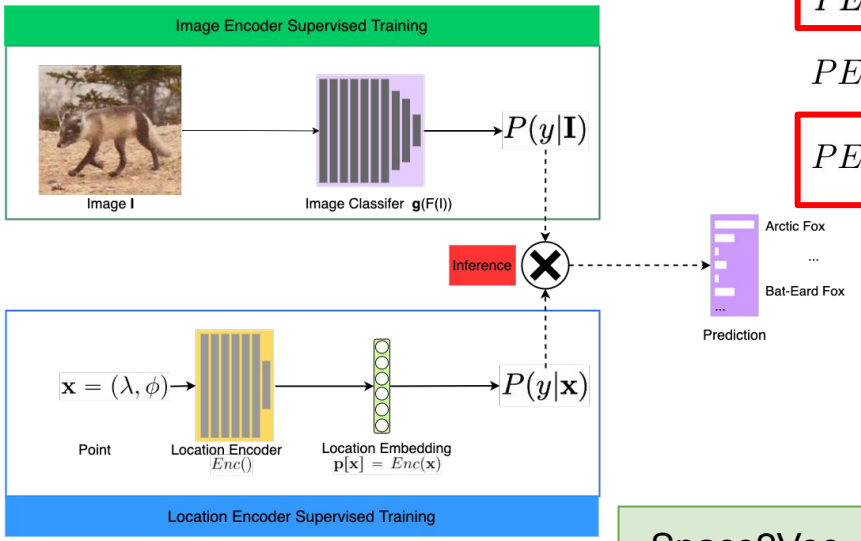
Given a location **x**, generate a **multi-scale** representation:

$$\mathbf{e}[\mathbf{x}] = Enc^{(x)}_{theory}(\mathbf{x}) = \mathbf{NN}(PE^{(t)}(\mathbf{x})) \quad (1)$$

$$PE^{(t)}(\mathbf{x}) = [PE^{(t)}_0(\mathbf{x}); ...; PE^{(t)}_s(\mathbf{x}); ...; PE^{(t)}_{S-1}(\mathbf{x})] \quad (2)$$

$$PE^{(t)}_s(\mathbf{x}) = [PE^{(t)}_{s,1}(\mathbf{x}); PE^{(t)}_{s,2}(\mathbf{x}); PE^{(t)}_{s,3}(\mathbf{x})] \quad (3)$$

$$PE^{(t)}_{s,j}(\mathbf{x}) = [\cos(\frac{\langle \mathbf{x}, \mathbf{a}_j \rangle}{\lambda_{min} \cdot g^{s/(S-1)}}); \sin(\frac{\langle \mathbf{x}, \mathbf{a}_j \rangle}{\lambda_{min} \cdot g^{s/(S-1)}})] \forall j = 1, 2, 3; \quad (4)$$



	BirdSnap†	NABirds†
No Prior (i.e. uniform)	70.07	76.08
Nearest Neighbor (num)	77.76	79.99
Nearest Neighbor (spatial)	77.98	80.79
Adaptive Kernel (Berg et al., 2014)	78.65	81.11
tile (Tang et al., 2015) (location only)	77.19	79.58
wrap (Mac Aodha et al., 2019) (location only)	78.65	81.15
rbf (σ=1k)	78.56	81.13
Space2Vec		
grid (λ_{min}=0.0001, λ_{max}=360, S = 64)	**79.44**	81.28
theory (λ_{min}=0.0001, λ_{max}=360, S = 64)	79.35	**81.59**

Mai, G., et al, 2020. Multi-Scale Representation Learning for Spatial Feature Distributions using Grid Cells. In **ICLR 2020**

***Spotlight paper (Acceptance Rate 6%).**

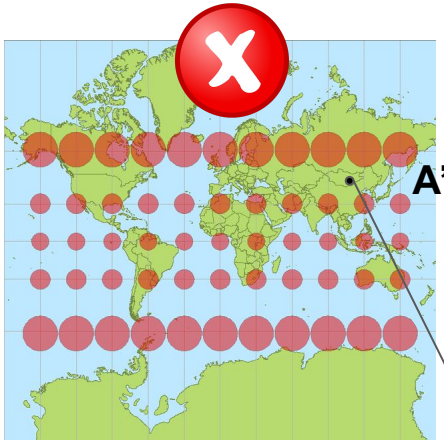
Sphere2Vec – Spherical Location Encoder

Calculating on a round planet!

(Chrisman et al., 2017)

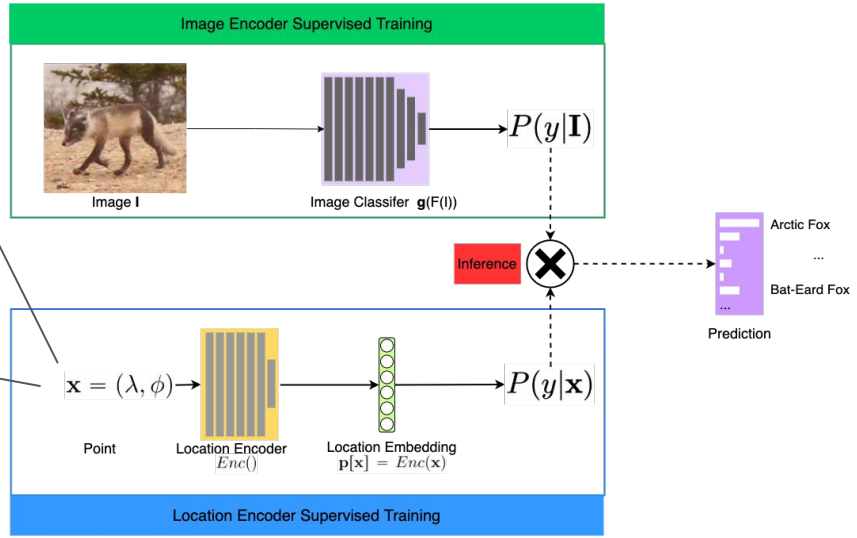
Space2Vec

- Treating lat/lon as **2D** coordinates
- **Map Projection Distortion**



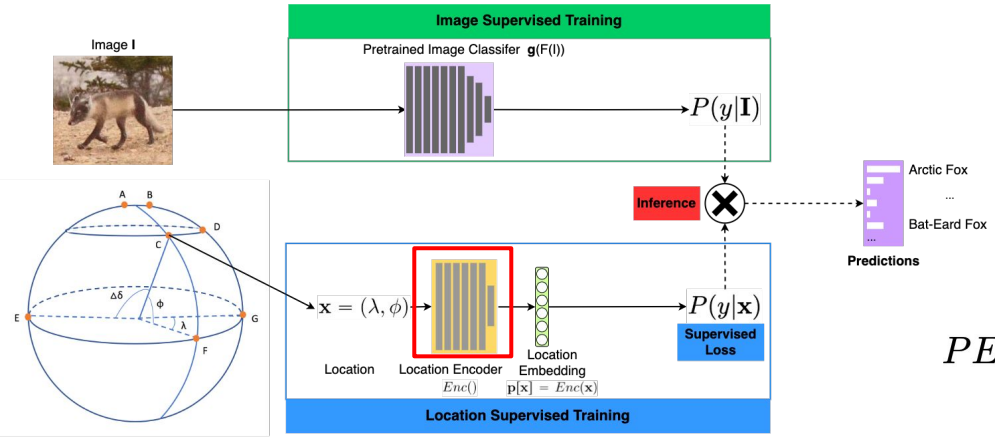
Sphere2Vec

- Treating lat/lon as **Spherical** coordinates



Sphere2Vec – Spherical Location Encoder

Sphere2Vec: Directly encode **spherical coordinates** with **Double Fourier Sphere (DFS)**



Any spherical coordinates $\mathbf{x} = (\lambda, \phi) \in \mathbb{S}^2$ can be converted to the **3D Cartesian space** by:

$$(x, y, z) = [\sin(\phi), \cos(\phi)\cos(\lambda), \cos(\phi)\sin(\lambda)] \quad (1)$$

Sphere2Vec generalizes it into a **Multi-scale periodical** representation

$$PE_{S,sphereC}(\mathbf{x}) = \bigcup_{s=0}^{S-1} [\sin\phi_s, \cos\phi_s \cos\lambda_s, \cos\phi_s \sin\lambda_s]. \quad (2)$$

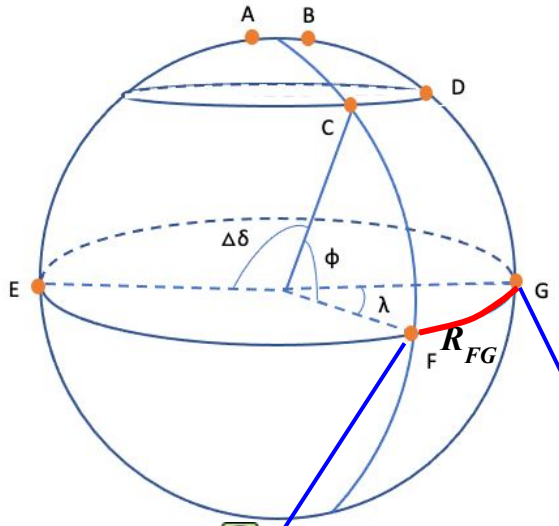
where $\lambda_s = \frac{\lambda}{f^s}, \phi_s = \frac{\phi}{f^s}, f^s = r_{min} \cdot g^{s/(S-1)}$

Mai, G., et al. 2023. Sphere2Vec: A General-Purpose Location Representation Learning over a Spherical Surface for Large-Scale Geospatial Predictions. *ISPRS Journal of Photogrammetry and Remote Sensing (2023)*

Sphere2Vec – Spherical Location Encoder

Spherical Distance Preservation:

The **cosine similarity** between two **Sphere2Vec location embeddings** is **inverse proportional to** the **spherical distance** between the corresponding points.



Sphere2Vec
Location Embedding

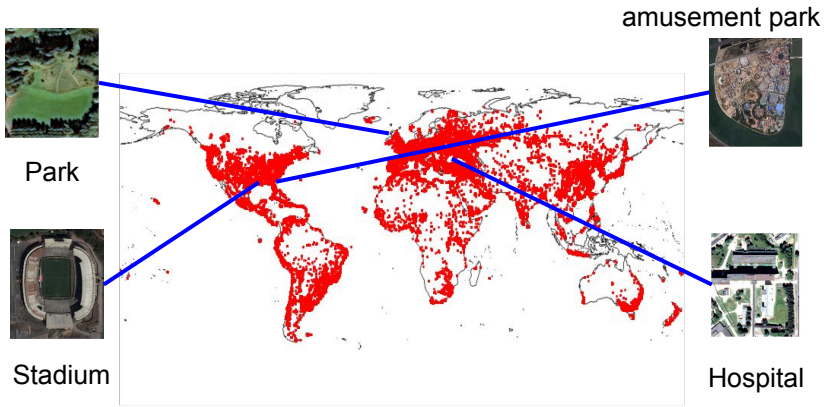
$e(\mathbf{x}_F)$ × $e(\mathbf{x}_G)$ $\propto \frac{1}{Dis(\mathbf{x}_F, \mathbf{x}_G)} = \frac{1}{R_{FG}}$

Sphere2Vec: Fine-Grained Species Recognition

	Task	Species Recognition					
	Dataset	BirdSnap	BirdSnap†	NABirds†	iNat2017	iNat2018	Avg
	P(y|x) - Prior Type	Test	Test	Test	Val	Val	-
Image Only	No Prior (i.e. image model)	70.07	70.07	76.08	63.27	60.20	67.94
Baselines	*tile* (Tang et al., 2015)	70.16	72.33	77.34	66.15	65.61	70.32
	xyz	71.85	78.97	81.20	69.39	71.75	74.63
	$wrap*$ (Mac Aodha et al., 2019)	71.66	78.65	81.15	69.34	72.41	74.64
	$wrap$	71.87	79.06	**81.62**	69.22	72.92	74.94
	$wrap + ffn$	**71.99**	79.21	81.36	69.40	71.95	74.78
	rbf (Mai et al., 2020b)	71.78	79.40	81.32	68.52	71.35	74.47
	rff (Rahimi et al., 2007)	71.92	79.16	81.30	69.36	71.80	74.71
Space2Vec	*Space2Vec-grid* (Mai et al., 2020b)	71.70	79.72	81.24	69.46	73.02	75.03
	Space2Vec-theory (Mai et al., 2020b)	71.88	**79.75**	81.30	**69.47**	**73.03**	**75.09**
NeRF	*NeRF* (Mildenhall et al., 2020)	71.66	79.66	81.32	69.45	73.00	75.02
Sphere2Vec	*Sphere2Vec-sphereC*	72.11	79.80	81.88	69.68	73.29	75.35
	Sphere2Vec-sphereC+	**72.41**	80.11	**81.97**	**69.75**	73.31	75.51
	Sphere2Vec-sphereM	72.06	79.84	81.94	69.72	73.25	75.36
	Sphere2Vec-sphereM+	72.24	**80.57**	81.94	69.67	**73.80**	**75.64**
	Sphere2Vec-dfs	71.75	79.18	81.39	69.65	73.24	75.04

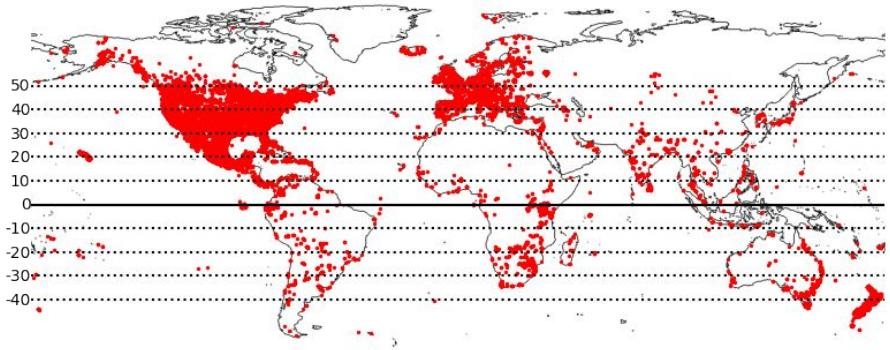
Sphere2Vec: Satellite Image Classification

	Task	RS
	Dataset	fMOW
	P(y|x) - Prior Type	Val
Image Only	No Prior (i.e. image model)	69.84
Baselines	*tile* (Tang et al., 2015)	-
	xyz	70.18
	wrap ∗ (Mac Aodha et al., 2019)	-
	wrap	70.29
	wrap + *ffn*	70.28
	rbf (Mai et al., 2020b)	70.65
	rff (Rahimi et al., 2007)	70.27
Space2Vec	*Space2Vec-grid* (Mai et al., 2020b)	**70.80**
	Space2Vec-theory (Mai et al., 2020b)	**70.81**
NeRF	*NeRF* (Mildenhall et al., 2020)	70.64
Sphere2Vec	*Sphere2Vec-sphereC*	71.00
	Sphere2Vec-sphereC+	71.03
	Sphere2Vec-sphereM	70.99
	Sphere2Vec-sphereM+	71.10
	Sphere2Vec-dfs	**71.46**

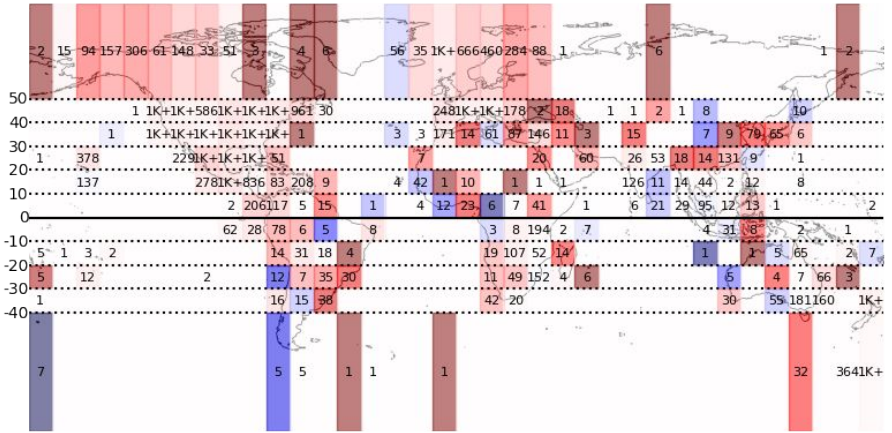


The spatial distributions of satellite image tiles from **Functional Map of the World (fMoW) dataset** (Christie et al. 2018)

Sphere2Vec: Visual Analysis



(a) Species image locations of iNat2017 validation dataset

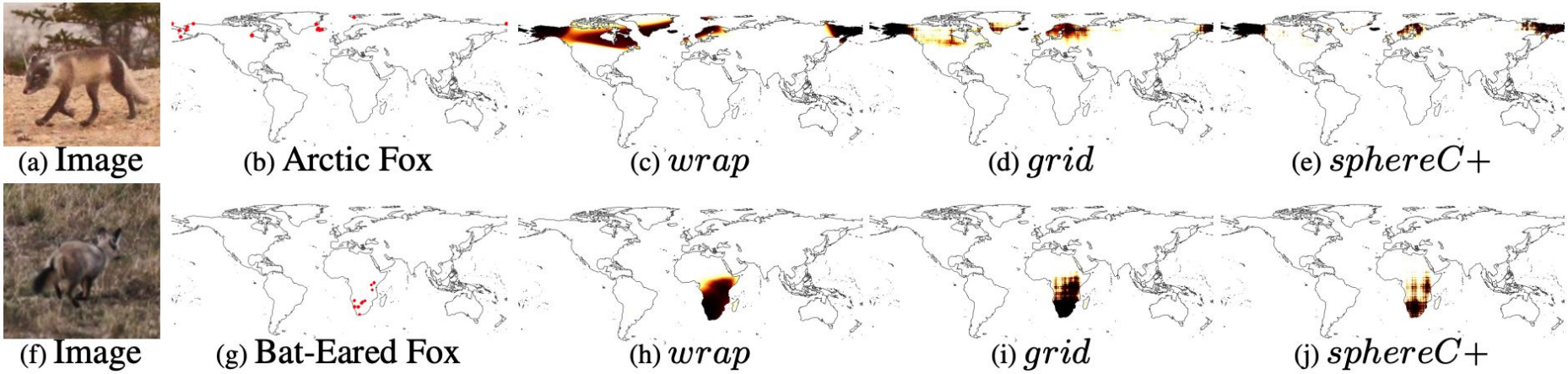


(b) ΔMRR = MRR(Sphere2Vec) - MRR(Space2Vec)

- Red: Sphere2Vec > Space2Vec
- Blue: Otherwise

Predicted Species Distribution

- **Space2Vec & Sphere2Vec** captures the **spatial heterogeneity**.
- **Sphere2Vec** considers the **map projection distortion**.



Reference

1) **Gengchen Mai**, Weiming Huang, Jin Sun, Suhang Song, Deepak Mishra, Ninghao Liu, Song Gao, Tianming Liu, Gao Cong, Yingjie Hu, Chris Cundy, Ziyuan Li, Rui Zhu, Ni Lao. On the Opportunities and Challenges of Foundation Models for Geospatial Artificial Intelligence. arXiv preprint arXiv:2304.06798 (2023).
2) Jielu Zhang, Zhongliang Zhou, **Gengchen Mai**, Lan Mu, Mengxuan Hu, Sheng Li. Text2Seg: Remote Sensing Image Semantic Segmentation via Text-Guided Visual Foundation Models. arXiv preprint arXiv:2304.10597 (2023).
3) **Gengchen Mai,** Krzysztof Janowicz, Bo Yan, Rui Zhu, Ling Cai, Ni Lao. Multi-Scale Representation Learning for Spatial Feature Distributions using Grid Cells, In: *Proceedings of ICLR 2020*.
4) **Gengchen Mai**, Ni Lao, Yutong He, Jiaming Song, Stefano Ermon. Self-Supervised Contrastive Spatial Pre-Training for Geospatial-Visual Representations, In: *Proceedings of ICML 2023*.
5) Haixing Dai, Yiwei Li, Zhengliang Liu, Lin Zhao, Zihao Wu, Suhang Song, Ye Shen, Dajiang Zhu, Xiang Li, Sheng Li, Xiaobai Yao, Lu Shi, Quanzheng Li, Zhuo Chen, Donglan Zhang, **Gengchen Mai***, Tianming Liu*. AD-AutoGPT: An Autonomous GPT for Alzheimer's Disease Infodemiology. arXiv preprint arXiv:2306.10095. ***Corresponding author**
6) **Gengchen Mai**, Yao Xuan, Wenyun Zuo, Yutong He, Jiaming Song, Stefano Ermon, Krzysztof Janowicz, Ni Lao. 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.

UNIVERSITY OF GEORGIA
Single Line Name

Contact

Prof. **Gengchen Mai**

Email: gengchen.mai25@uga.edu

Website: https://gengchenmai.github.io/

Acknowledgement: