CSP: Self-Supervised Contrastive Spatial Pre-Training for Geospatial-Visual Representations

– Towards a Multimodal Foundation Model for GeoAI

Dr. Gengchen Mai

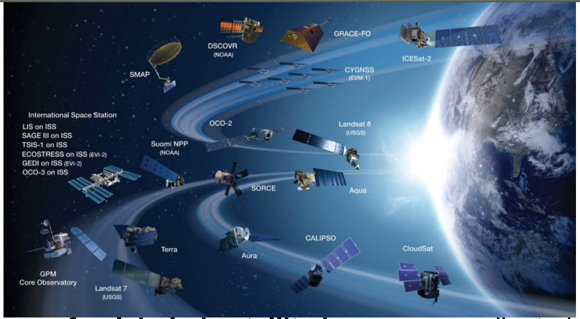
[2022 - Now] Assistant Professor at *Department of Geography, University of Georgia*

https://gengchenmai.github.io/

Acknowledgement:

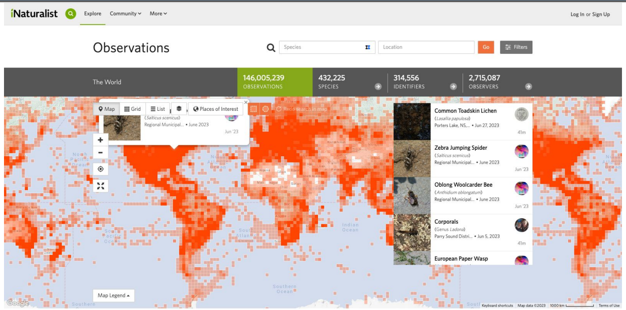
Massive Unlabeled Geo-tagged Image Datasets

Unlabeled RS Images



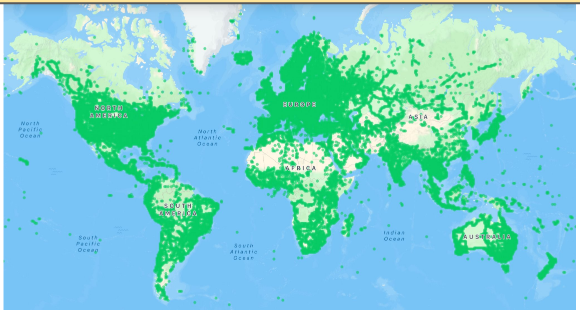
Billions of unlabeled satellite images are collected from various sensors everyday (Figure from [NASA Website])

Unlabeled iNaturalist Images



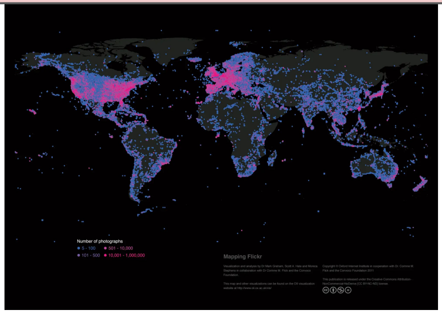
Millions of unlabeled geo-tagged species images are collected everyday (Figure from [iNaturalist Website])

Unlabeled StreetView Images



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

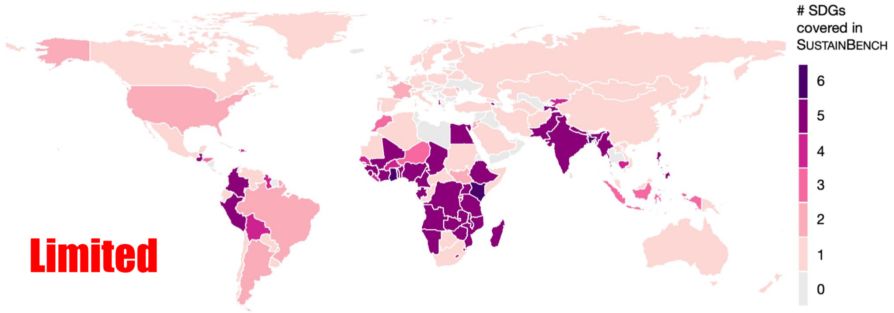
Unlabeled Flickr Images



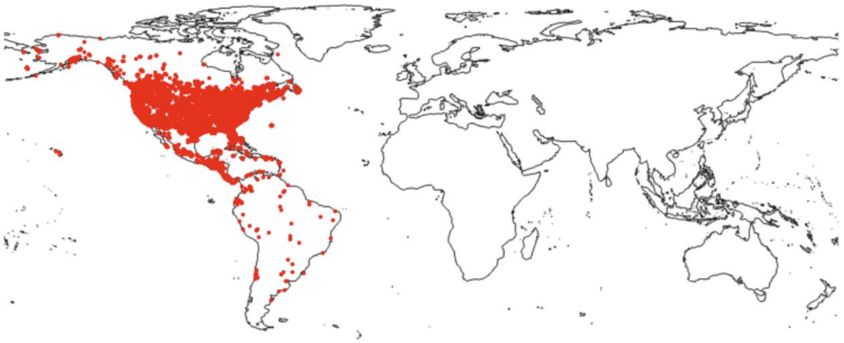
Billions of unlabeled Flickr images are uploaded everyday (Figure from [Oxford Internet Institute])

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)

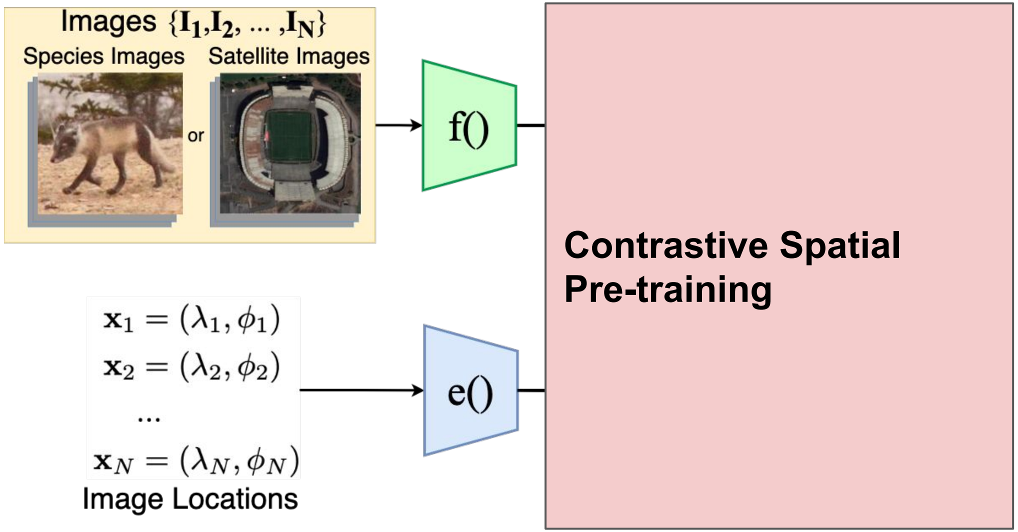


Geographic coverage of labeled species fine-grained recognition dataset – NABird (Mai et al., 2023)

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

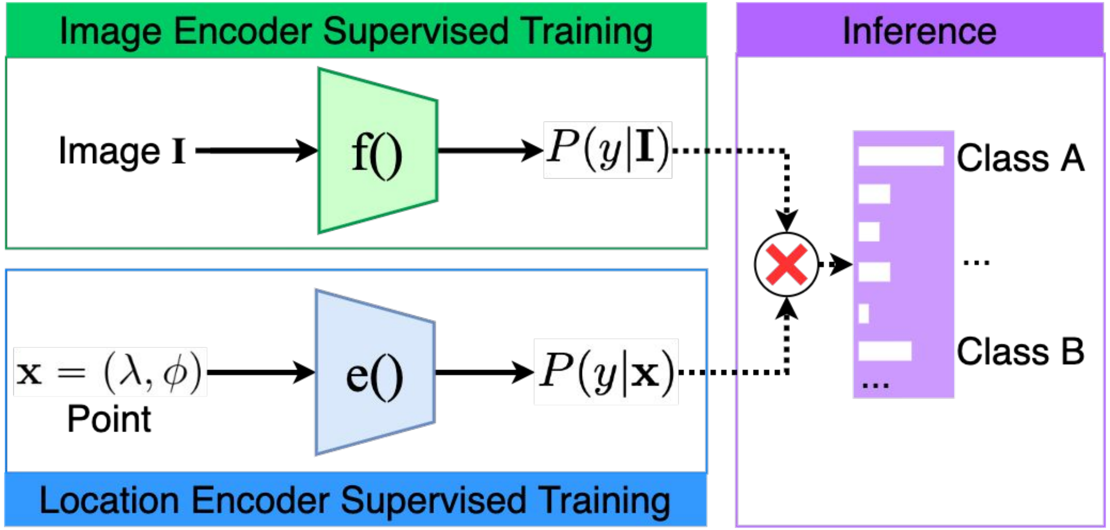


Figure 2(a) Sup. Only: Geo-aware Supervised Learning (Mac Aodha et al., 2019; Mai et al., 2020b; Mai et al., 2023)

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.

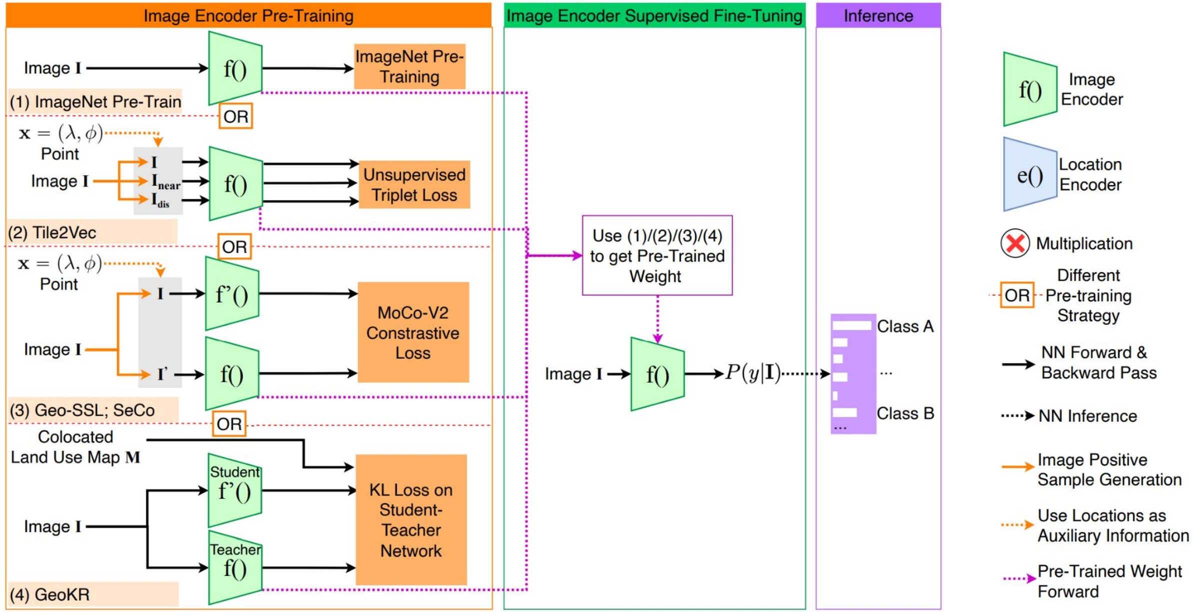


Figure 2(b) Img. Only: Image Encoder Pre-Training with Geographic Knowledge

Contrastive Spatial Pre-Training (CSP)

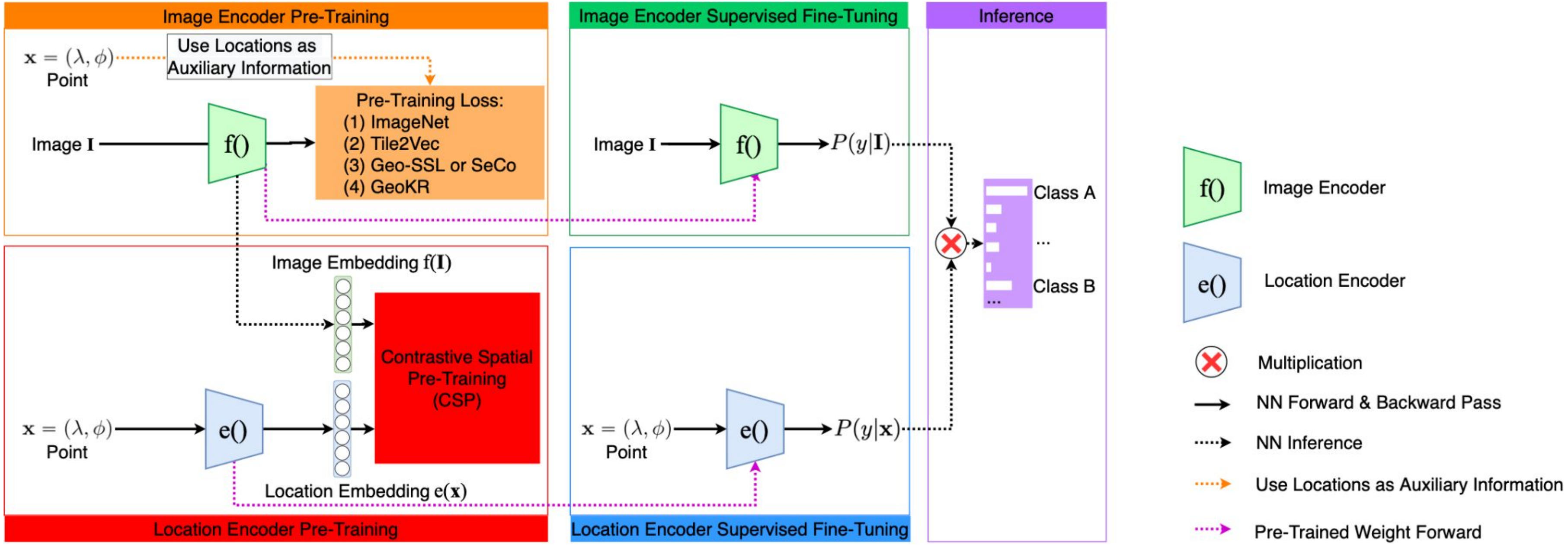
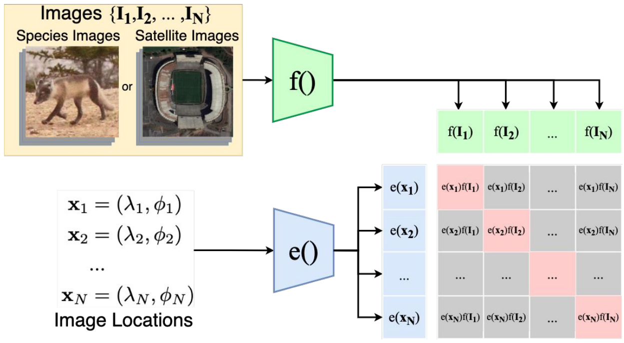


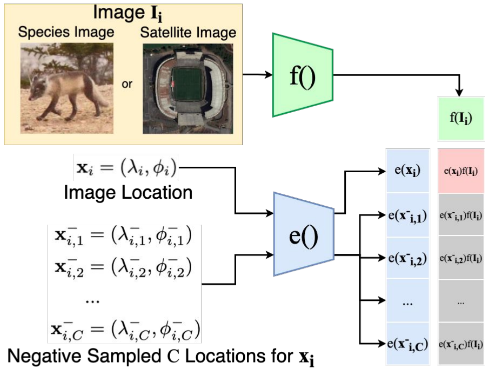
Figure 2(c) Contrastive Spatial Pre-Training (CSP)

Contrastive Spatial Pre-Training (CSP)

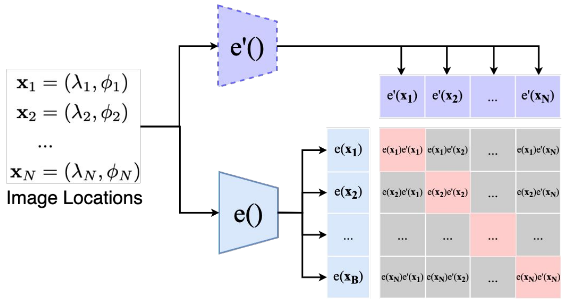
Contrast the representations between **geo-locations** and **images** in a self-supervised learning manner in three ways:



(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 $l_{\mathrm{MC}}^{D}(\mathbb{X})$ 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/

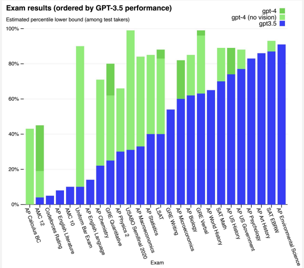
ArXiv: https://arxiv.org/abs/2305.01118

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

Foundation Models (FMs) in Different Domains

Natural Language Processing

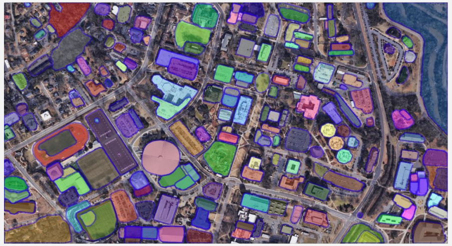
Stanford Alpaca



ChatGPT/GPT-4 (OpenAI. 2023)

Computer Vision

Imagen (Saharia et al. 2022)

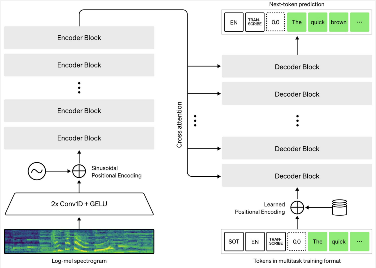


Segment Anying (Kirillov et al, 2023)

Reinforcement Learning

Gato (Reed et al. 2022)

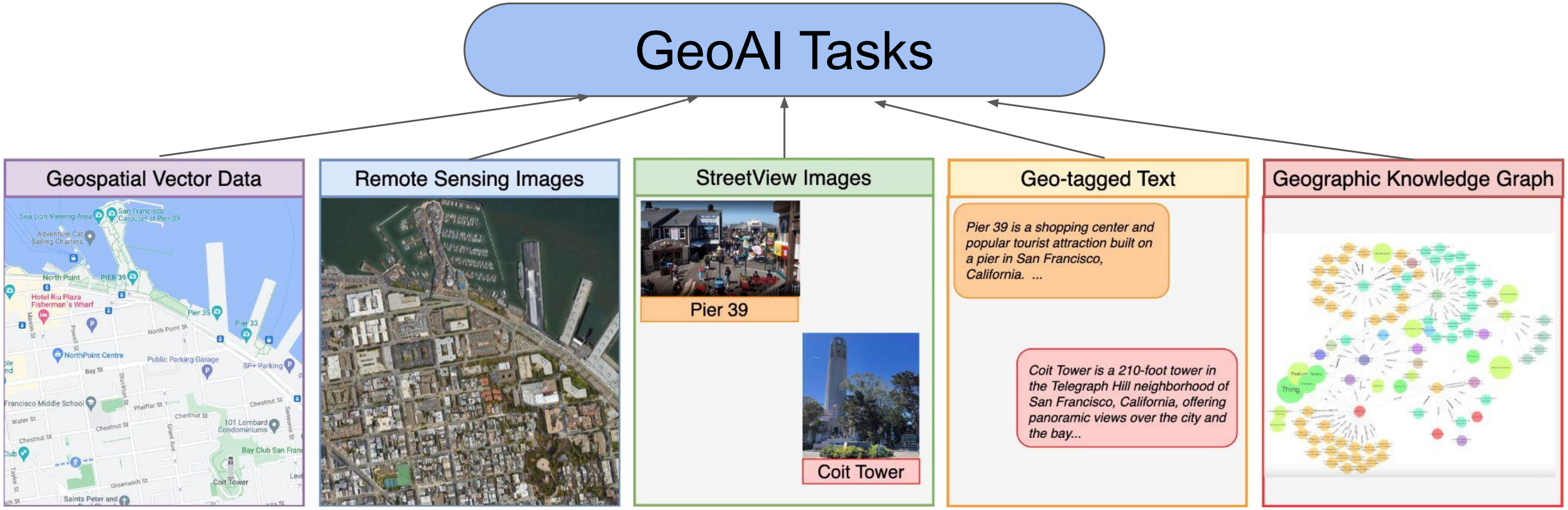
Signal Processing



Whisper (Radford et al. 2022)

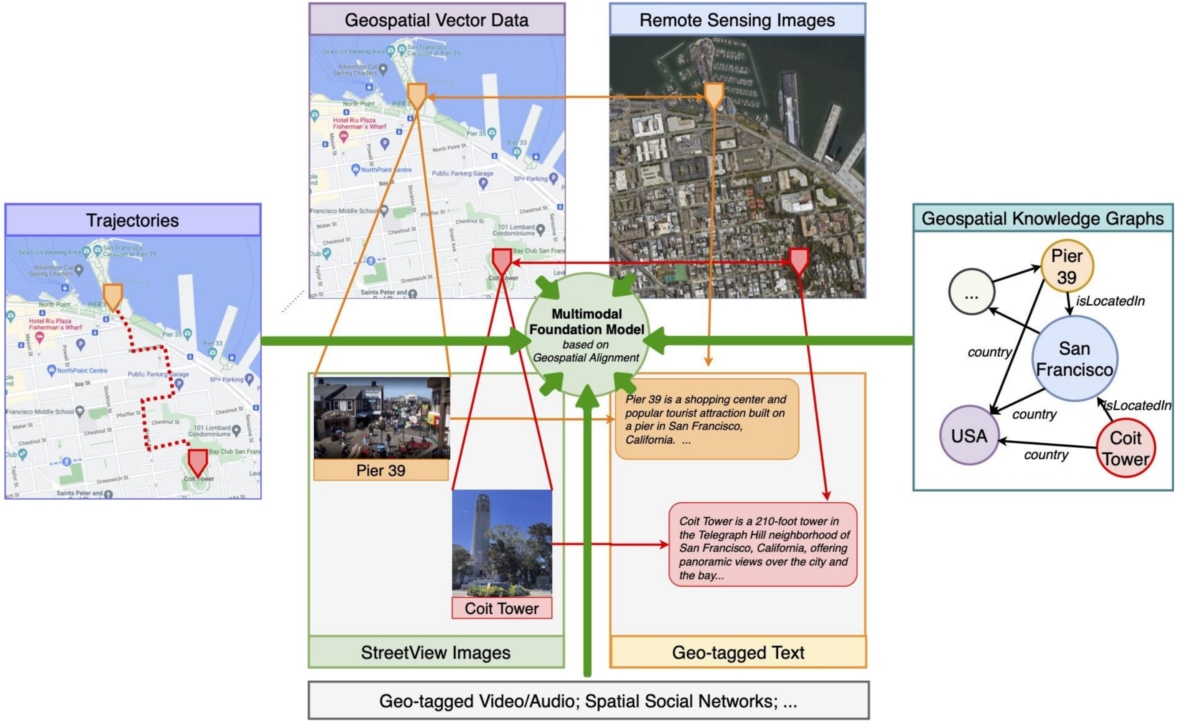
Unique Challenges of GeoAI for FMs

- **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

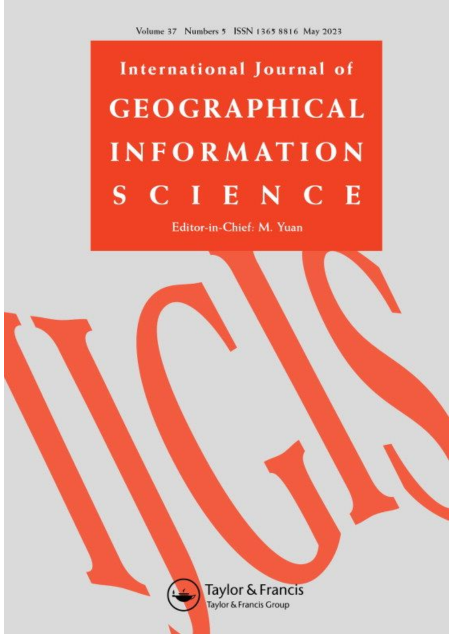
Interested authors should first submit a short abstract (250 words max) to Krzysztof Janowicz (krzysztof.janowicz@univie.ac.at) and Gengchen Mai (gengchen.mai25@uga.edu) before September 23th, 2023.

Important Dates

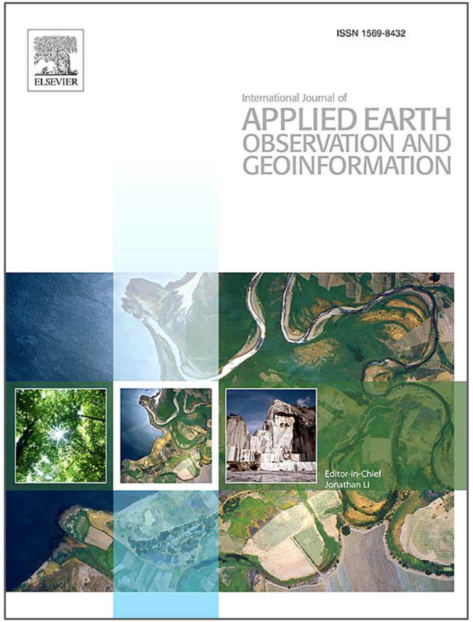
- Abstracts (no more than 250 words) Due: **Sep. 23, 2023**
- Decisions on abstracts: **Sep. 30, 2023**
- Full manuscripts Due: **Nov. 30, 2023**

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) 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:

NSF | NIH National Institutes of Health | aws | ALFRED P. SLOAN FOUNDATION | GEORGIA Department of Geography | Stanford ARTIFICIAL INTELLIGENCE