From Points to Places: Towards Human Mobility-Driven Spatiotemporal Foundation Models via Understanding Places

Mohammad Hashemi
mohammad.hashemi@emory.edu
Emory University
Atlanta, USA

Andreas Züfle
azufle@emory.edu
Emory University
Atlanta, USA

ABSTRACT

Capturing human mobility is essential for modeling how people interact with and move through physical spaces, reflecting social behavior, access to resources, and dynamic spatial patterns. To support scalable and transferable analysis across diverse geographies and contexts, there is a need for a generalizable foundation model for spatiotemporal data. While foundation models have transformed language and vision, they remain limited in handling the unique challenges posed by the spatial, temporal, and semantic complexity of mobility data. This vision paper advocates for a new class of spatial foundation models that integrate geolocation semantics with human mobility across multiple scales. Central to our vision is a shift from modeling discrete points of interest to understanding *places*: dynamic, context-rich regions shaped by human behavior and mobility that may comprise many places of interest. We identify key gaps in adaptability, scalability, and multi-granular reasoning, and propose research directions focused on modeling places and enabling efficient learning. Our goal is to guide the development of scalable, context-aware models for next-generation geospatial intelligence. These models unlock powerful applications ranging from personalized place discovery and logistics optimization to urban planning, ultimately enabling smarter and more responsive spatial decision-making.

CCS CONCEPTS

• **Information systems → Geographic information systems**; **Location based services**; • **Computing methodologies → Neural networks**; **Learning latent representations**.

KEYWORDS

Human Mobility, Spatiotemporal Foundation Model, Spatial Representation Learning
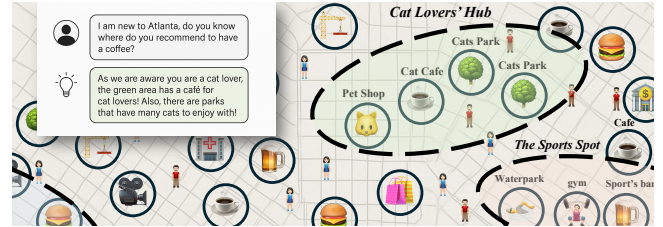
Figure 1: An illustration of the distinction between *points* and *places*. While points such as POIs represent fixed geographic references without awareness of their surrounding context or human mobility information, places reflect socially or personally meaningful regions that may span or transcend these predefined boundaries.

1 INTRODUCTION

Analyzing human mobility and geolocation data offers a wealth of applications that touch nearly every aspect of modern life [22, 23]. To truly understand human mobility, it is essential to uncover the underlying relationships between *places* and the activities they attract, as these connections shape when, where, and why people move [35]. This scenario, also illustrated in Figure 1, reflects a common experience: While exploring a new city, someone who loves cats might visit nearby cafés and restaurants but miss out on a unique experience, such as a local cat café, simply because they weren't aware it existed. This highlights the potential of intelligent systems capable of inferring personal interests and mobility patterns to provide relevant, personalized recommendations and help people discover places that matter to them without actively searching. This experience showcases a broader need: *Models that aim to understand spatial context and connect it to individual behavior and intent.*

Places vs. Points: In the analysis of human mobility and geolocation data, Points of Interest (POIs) serve as the finest-level spatial units, offering detailed insights into the structure and function of the built environment [24, 26]. However, POIs such as cafés, museums, or parks do not, by themselves, capture the full meaning of a place. Foundational theories in social sciences and geography distinguish between *points*, which represent abstract coordinates or labeled locations, and *places*, which carry social or personal significance and may not conform to predefined geographic boundaries such as POIs, cities, or states [3, 12]. A *place*, from a human perspective, can be fluid, shaped by personal experience, routine, or cultural meaning. For instance, someone might consider "My weekend walking route" as a meaningful place, even though it spans multiple parks, streets, and cafés, none of which, individually, capture the essence of that

place. Similarly, a community might view a set of adjacent businesses and gathering spots as a single neighborhood hub, despite those POIs being labeled separately in datasets. These examples illustrate that places are often emergent, defined through human behavior and connection, rather than strictly bounded by spatial labels.

Although some studies [19] focus on learning general-purpose representations of locations, sometimes treating POIs as the finest-grained spatial units, such approaches fall short of capturing the lived experiences and contextual meanings tied to these locations. Therefore, understanding mobility patterns requires moving beyond learning static features of geographic units to modeling how individuals perceive and engage with these locations as meaningful places. This vision paper takes a step in that direction by proposing a human mobility-driven spatiotemporal foundation model via understanding places through the dynamics of how people move and interact with their environments, ultimately enabling downstream applications as personalized as guiding cat lovers to places where they can find cat cafés and other places they may like.

2 RELATED WORKS

Research shows that understanding the features of geolocation data can be a more powerful indicator of long-term health outcomes than genetic factors [13]. This recognition has driven efforts to collect large-scale data capturing the interplay between human activity and the surrounding environment [22, 23]. To effectively harness this wealth of data, robust methods for understanding geographic entities have been developed. By analyzing how people move through space and time, we can transform raw location data into actionable insights, uncovering meaningful mobility patterns that drive a wide range of advanced applications [22].

Machine learning has been applied to a wide range of data sources in various modalities for geospatial modeling [8]: For instance, web search data is used to predict influenza trends [10]. Search queries are leveraged to model global economic indicators [5]. Also, satellite imagery data is utilized to estimate factors like forest cover and housing prices [27]. From a task-driven standpoint, these approaches frequently employ statistical and deep learning models, such as CNNs and RNNs, to extract fine-grained spatiotemporal patterns [8], and the majority concentrate on particular domains like internet data, satellite imagery, or map-based datasets. Also, a range of prior approaches have also aimed to develop general-purpose geographic encoders [17, 30].

To address the limitations of task-specific approaches and the reliance on manually crafted features to encode geolocation data and human mobility, there is a growing need for foundation models that can understand geolocation data at varying levels of granularity while simultaneously capturing human mobility patterns [19]. Foundation models have already become a dominant paradigm in domains such as computer vision and natural language processing, where models like CLIP [1] and GPT-4 [1] demonstrate strong transferability across tasks and data distributions. For spatiotemporal reasoning, recent work has focused on developing task-adaptive and region-agnostic foundation models for either **(i) trajectory prediction** [6, 20, 32, 36] or **(ii) geolocation representation learning** [2, 4, 18, 29, 33, 34].

2.1 Foundation Models for Trajectory Prediction

Foundation models for trajectory prediction (TP) are designed to capture general sequential patterns from trajectory data [6, 19, 20, 32, 36]. For example, PMT[32] introduces a transformer-based foundation model for human mobility prediction, representing trajectories as sequences of Census Block Groups(CBG). Trained autoregressively, the model captures spatiotemporal patterns to support trajectory prediction tasks. Similarly, UniTraj[36] proposes an encoder-decoder pretraining approach to obtain a POI sequence encoder that can be fine-tuned for trajectory-related downstream tasks. TrajFM [20] introduces a trajectory foundation model pretrained on vehicle trajectories from multiple cities, using trajectory masking and autoregressive recovery to enable both regional and task transferability.

2.2 Foundation Models for Geolocation Representation Learning

Another type of spatial foundation model focuses on geolocation representation learning (GRL), aiming to generate general-purpose embeddings for geographic entities [2, 4, 18, 19, 29, 33, 34]. Some studies have leveraged large language models(LLMs) and vision-language models(VLMs) to learn location embeddings. For example, GeoVectors [29], and SpaBERT [18] utilize open-source data such as OpenStreetMap[1], while G2PTL [31] is trained on large-scale logistics delivery data.

The most closely related work, PDFM [2], leverages a pretraining stage to integrate diverse, globally accessible geospatial data, such as maps, activity levels, and aggregated search trends, alongside environmental indicators like weather and air quality. This approach involves building a heterogeneous geospatial graph, where counties and postal codes serve as nodes, and edges are defined based on spatial proximity. A graph neural network (GNN) is then used to learn meaningful embeddings for these nodes.

Despite progress in geolocation foundation models, current methods still struggle to capture human mobility across multiple spatial scales and often fail to understand *places*, locations defined by human meaning and behavior. These models typically rely on fixed units like POIs or administrative boundaries, which do not reflect how people experience space. In Section 3, we delve into these limitations, which motivate the core objectives of this vision paper. Also, a general qualitative comparison of these existing foundation models is summarized in Table 1.

3 LIMITATIONS & MOTIVATIONS

1. Lack of Mutual Awareness in Mobility and Location Models

A key limitation lies in the lack of explicit human mobility information integrated into geolocation data. While current foundation models [2, 4, 18, 29] synthesize rich representations of geographic entities, they fall short in capturing who visits these locations, how they get there, and when these movements occur. This omission restricts the model's ability to fully understand and represent dynamic human behavior. Without incorporating mobility patterns, such as inflow and outflow volumes, visit frequency, and temporal

---

[1]https://www.openstreetmap.org/

Method	Understand Places	Mobility Utilization	Temporal Utilization	Capability	Granularity	Pretraining
TrajFM [20]	×	×	✓	TP	POI	Transformers
PMT [32]	×	✓	✓	TP	CBG	Transformers
UniTraj [36]	×	×	✓	TP	POI	Encoder-Decoder
PDFM [2]	×	×	×	GRL	ZIP codes/County	GNN
CityFM [4]	×	×	×	GRL	POI	Contrastive Learning
SpaBERT [4], G2PTL [31]	×	×	×	GRL	POI	LM/LLM
ReFound [33], UrbanCLIP [34]	×	×	×	GRL	POI	VLM

Table 1: General qualitative comparison of existing spatial foundation models. "Understand Places": the ability to reason about *places*, "Mobility Utilization": the ability to utilize human mobility information, "Temporal Utilization": the reliance on temporal information, "Capability": the goal of foundation model which is either TP as for Trajcetory Prediction, or GRL as for Geographic Representation Learning, "Granularity": level(s) of granularity that the model can infer, "Pretraining": the pretraining framework

visit distributions, the learned representations remain static and disconnected from real-world usage. On the other hand, existing foundation models for trajectory prediction [6, 20, 32, 36] do not integrate the rich information from geolocation data into sequence training, resulting in the loss of location semantics across different levels of granularity. This gap highlights the need for models that go beyond static spatial features by integrating both geolocation-level data and mobility patterns. By combining where *places* are with how people move through them, we can begin to truly understand the complex, lived experience of *places*.

2. Lack of Temporal Dynamics in Modeling Mobility

Another limitation of existing foundation models for understanding geolocations, e.g. PDFM [2], is their handling of temporal information. The input data sources mostly exhibit misaligned temporal granularity, which can affect the model's consistency. Moreover, PDFM generates static geolocation embeddings, failing to capture the dynamic and time-evolving nature of human mobility. Incorporating temporal alignment and explicitly modeling temporal dynamics could significantly improve its effectiveness in real-world mobility scenarios.

3. Scalability Issues

Another key limitation concerns the scalability of pretraining foundation models, which typically rely on massive datasets. For instance, UniTraj [36] is pretrained on 2.45 million trajectories using an encoder-decoder architecture. Similarly, a transformer-based model has been trained on nearly 42 million sequences of location-based service (LBS) data to support tasks like next-location prediction and mask imputation [32]. While these models demonstrate strong performance, such large-scale training paradigms demand significant computational resources, posing barriers to accessibility, reproducibility, and deployment in resource-constrained environments.

4. Limitation in Granularity Flexibility

Another limitation of current foundation models is their design for single-granularity inference. For instance, PMT [32] represents trajectories as sequences of CBGs, and all downstream tasks, such as next location prediction, are performed at this level. Similarly, PDFM [2] learns general-purpose embeddings for U.S. ZIP codes and counties, limiting inference to those granularities and lacking the ability to operate at finer levels such as POIs or CBGs. However, a *place* might be interpreted by multiple geographic entities at different granularity levels. This raises a critical question: *Can we develop*

a foundation model that integrates information across multiple spatial scales and supports inference at any desired granularity?

4 RESEARCH DIRECTIONS

A human mobility-driven spatial foundation model must capture rich, multi-dimensional context to support diverse applications. In this section, we highlight essential contextual signals rather than prescribing specific training methods.

4.1 Towards Understanding Places

To reason about human mobility in a meaningful way, it is essential to move beyond point-based representations and adopt a structured notion of *places*. In this context, we define a place as a semantically meaningful spatial unit that may correspond to or span multiple geographic entities, such as POIs, postcodes, neighborhoods, or administrative regions.

Definition 4.1 (Place). A *place* P is defined as a non-empty set of spatial entities:

$$P = \{e_1, e_2, \ldots, e_n\}, \quad \text{where } e_i \in \mathcal{E}$$

and $\mathcal{E} = \mathcal{G} \cup \mathcal{P}$ denotes the universe of geographic entities $\mathcal{G}$ (e.g., POIs, postcodes, cities) and existing places $\mathcal{P}$. This definition allows a place to be hierarchically composed of both primitive spatial entities and other places, enabling recursive and multi-scale representations. This formalism allows flexibility in representing places across scales and contexts. For instance, a place might be as specific as a single cat café or as broad as a neighborhood known for pet-friendly venues.

Challenges: Learning to represent places introduces several challenges. First, the data associated with many geographic entities is often sparse or incomplete. For example, detailed mobility traces or POI metadata may be missing for underrepresented regions. Second, the notion of place is inherently subjective and dynamic, evolving with user preferences, temporal context, and social factors.

4.2 Spatiotemporal Representations for Human Mobility Understanding

An important step toward understanding human mobility is the ability to represent the built environment in a way that reflects how people interact with it. Rather than viewing places in isolation, future research must consider how their roles, proximities, and

interactions shape movement behavior. This includes modeling how certain places attract recurring visits, how regions are organized hierarchically, and how connectivity patterns vary across urban and rural landscapes, all of which can be effectively captured using heterogeneous graph representations. By capturing such spatial structures, researchers can better ground human mobility patterns in the environments that produce them, enabling models that are both generalizable and context-aware.

Challenges: While some approaches may rely solely on geographic attributes such as latitude, longitude, and spatial distance, richer representations often require incorporating semantic information like POI categories, functional roles, or contextual metadata. Determining which features to include and how to encode them remains an open research question that significantly impacts the expressiveness and utility of the resulting spatial representations.

4.3 Scalable and Multi-Granular Representation Learning

As geolocation data becomes increasingly fine-grained, particularly at the POI level, new challenges emerge around scalability and granularity. Once rich POI datasets are constructed by integrating data sources, training foundation models on such large-scale datasets becomes computationally intensive and time-consuming.

Assuming the constructed POI dataset is represented using a graph modality, addressing the challenges of scalability and granularity will require innovation on multiple fronts. From a model-centric perspective, researchers may explore more efficient deep learning architectures that scale to massive graphs. From a data-centric view, graph condensation [11, 14, 15] techniques offer promising pathways to reduce the size of graph datasets while retaining key spatiotemporal structures, enabling efficient training without sacrificing model performance.

Challenges: Future research must address the challenge of learning flexible representations across varying spatial granularities. Depending on the context, a system may need to infer user intent at the level of a neighborhood, city, or even state. Supporting such seamless multi-scale inference remains a key open problem in geospatial AI.

4.4 Model Pre-training

LLMs and VLMs as vision and language foundation models are typically pretrained on static corpora like Wikipedia or ImageNet [7], where data remains relevant over long periods. In contrast, human mobility data is highly dynamic, shaped by infrastructure updates, policy changes, and events such as pandemics. This temporal fluidity makes static pretraining unsuitable for spatiotemporal graph data. To address this, future research should develop continual and online pretraining strategies that allow spatial foundation models to adapt to evolving mobility patterns without forgetting previous knowledge. These approaches must detect distribution shifts, efficiently update representations, and ensure that models remain aligned with current movement trends.

Challenges: A key challenge lies in designing architectures that can effectively model the inherently multimodal nature of geolocation data. This includes integrating spatial, temporal, and semantic modalities such as coordinates, timestamps, movement patterns, transportation modes, POIs, and environmental context.

5 REAL-WORLD APPLICATIONS

Human mobility-driven spatial foundation models can significantly improve spatiotemporal understanding and enable diverse applications. The following examples highlight their transformative potential.

5.1 Improved Spatiotemporal Analysis

Foundation models that combine insights from both geolocation data and human mobility can simplify geospatial analysis and shorten the path from concept to deployment. By capturing where people go and how they spend their time, these models can support a range of use cases, from identifying ideal sites for new businesses and analyzing real estate trends to optimizing logistics and supply chains. They can also enhance socioeconomic studies and benefit sectors like hospitality, especially when tailored to specific populations like travelers.

5.2 Personalized Place Discovery

A personalized geospatial recommender system [9] is a key application of human mobility-driven spatial foundation models. When visiting a new city with no knowledge of the area, users often struggle to find places aligned with their interests. A foundation model trained on mobility patterns and being able to understand places can infer how similar users move through a city and what places they visit. This allows the system to provide tailored, context-aware suggestions, like a hidden café or a popular local spot, without requiring active search. It enhances discovery, supports tourism, and improves user experience in unfamiliar environments.

5.3 Urban Planning

Integrating geolocation embeddings with human mobility information allows urban planners to move beyond static maps and better understand how people actually interact with space over time. This fusion reveals which areas experience high foot traffic, how people transition between neighborhoods, and where bottlenecks or service gaps emerge [16, 21]. They can also play a critical role in disaster preparedness and response, such as predicting evacuation patterns or modeling the impact of events like earthquakes to improve emergency infrastructure and resource allocation [25, 28]. As a result, planners can make more informed decisions about where to place amenities, how to design transit routes, and how to adapt urban spaces to the dynamic needs of residents, ultimately creating smarter, more responsive cities.

6 CONCLUSION

In this vision paper, we highlighted critical limitations in existing spatial foundation models and outlined research directions for building human mobility-driven models that integrate geolocation semantics across multiple granularities. By capturing how people move through space and interact with places over time, future foundation models can unlock new capabilities for personalized recommendations, dynamic urban planning, and fine-grained spatial analysis.

REFERENCES

[1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774* (2023).

[2] Mohit Agarwal, Mimi Sun, Chaitanya Kamath, Arbaaz Muslim, Prithul Sarker, Joydeep Paul, Hector Yee, Marcin Sieniek, Kim Jablonski, Yael Mayer, et al. 2024. General Geospatial Inference with a Population Dynamics Foundation Model. *arXiv preprint arXiv:2411.07207* (2024).

[3] John Agnew. 2011. Space and place. *Handbook of geographical knowledge* 2011 (2011), 316–331.

[4] Pasquale Balsebre, Weiming Huang, Gao Cong, and Yi Li. 2024. City foundation models for learning general purpose representations from openstreetmap. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*. 87–97.

[5] Hyunyoung Choi and Hal Varian. 2012. Predicting the present with Google Trends. *Economic record* 88 (2012), 2–9.

[6] Shushman Choudhury, Abdul Rahman Kreidieh, Ivan Kuznetsov, and Neha Arora. 2024. Towards a Trajectory-powered Foundation Model of Mobility. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Spatial Big Data and AI for Industrial Applications*. 1–4.

[7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 248–255.

[8] Peijun Du, Xuyu Bai, Kun Tan, Zhaohui Xue, Alim Samat, Junshi Xia, Erzhu Li, Hongjun Su, and Wei Liu. 2020. Advances of four machine learning methods for spatial data handling: A review. *Journal of Geovisualization and Spatial Analysis* 4 (2020), 1–25.

[9] Quan Fang, Changsheng Xu, M Shamim Hossain, and Ghulam Muhammad. 2016. Stcaplrs: a spatial-temporal context-aware personalized location recommendation system. *ACM Transactions on Intelligent systems and technology (TIST)* 7, 4 (2016), 1–30.

[10] Jeremy Ginsberg, Matthew H Mohebbi, Rajan S Patel, Lynnette Brammer, Mark S Smolinski, and Larry Brilliant. 2009. Detecting influenza epidemics using search engine query data. *Nature* 457, 7232 (2009), 1012–1014.

[11] Shengbo Gong, Mohammad Hashemi, Juntong Ni, Carl Yang, and Wei Jin. 2025. Scalable Graph Condensation with Evolving Capabilities. *arXiv preprint arXiv:2502.17614* (2025).

[12] Michael F Goodchild. 2010. Formalizing place in geographic information systems. In *Communities, neighborhoods, and health: Expanding the boundaries of place*. Springer, 21–33.

[13] Garth N Graham. 2016. Why your ZIP code matters more than your genetic code: promoting healthy outcomes from mother to child. *Breastfeeding Medicine* 11, 8 (2016), 396–397.

[14] Mohammad Hashemi, Shengbo Gong, Juntong Ni, Wenqi Fan, B Aditya Prakash, and Wei Jin. 2024. A comprehensive survey on graph reduction: Sparsification, coarsening, and condensation. *IJCAI* (2024).

[15] Wei Jin, Lingxiao Zhao, Shichang Zhang, Yozen Liu, Jiliang Tang, and Neil Shah. 2021. Graph condensation for graph neural networks. *arXiv preprint arXiv:2110.07580* (2021).

[16] Anirudh Ameya Kashyap, Shravan Raviraj, Ananya Devarakonda, Shamanth R Nayak K, Santhosh KV, and Soumya J Bhat. 2022. Traffic flow prediction models–A review of deep learning techniques. *Cogent Engineering* 9, 1 (2022), 2010510.

[17] Konstantin Klemmer, Esther Rolf, Caleb Robinson, Lester Mackey, and Marc Rußwurm. 2025. Satclip: Global, general-purpose location embeddings with satellite imagery. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 4347–4355.

[18] Zekun Li, Jina Kim, Yao-Yi Chiang, and Muhao Chen. 2022. Spabert: a pre-trained language model from geographic data for geo-entity representation. *arXiv preprint arXiv:2210.12213* (2022).

[19] Yuxuan Liang, Haomin Wen, Yutong Xia, Ming Jin, Bin Yang, Flora Salim, Qingsong Wen, Shirui Pan, and Gao Cong. 2025. Foundation Models for Spatio-Temporal Data Science: A Tutorial and Survey. *arXiv preprint arXiv:2503.13502* (2025).

[20] Yan Lin, Tonglong Wei, Zeyu Zhou, Haomin Wen, Jilin Hu, Shengnan Guo, Youfang Lin, and Huaiyu Wan. 2024. TrajFM: A vehicle trajectory foundation model for region and task transferability. *arXiv preprint arXiv:2408.15251* (2024).

[21] Boris Medina-Salgado, Eddy Sánchez-DelaCruz, Pilar Pozos-Parra, and Javier E Sierra. 2022. Urban traffic flow prediction techniques: A review. *Sustainable Computing: Informatics and Systems* 35 (2022), 100739.

[22] Mohamed Mokbel, Mahmoud Sakr, Li Xiong, Andreas Züfle, Jussara Almeida, Taylor Anderson, Walid Aref, Gennady Andrienko, Natalia Andrienko, Yang Cao, et al. 2024. Mobility data science: Perspectives and challenges. *ACM Transactions on Spatial Algorithms and Systems* 10, 2 (2024), 1–35.

[23] Mohamed F Mokbel, Mahmoud Attia Sakr, Li Xiong, Andreas Züfle, Jussara M Almeida, Taylor Anderson, Walid G Aref, Gennady L Andrienko, Natalia V Andrienko, Yang Cao, et al. 2022. Mobility Data Science: Dagstuhl Seminar 22021. *Dagstuhl reports* 12, 1 (2022).

[24] Haifeng Niu and Elisabete A Silva. 2021. Delineating urban functional use from points of interest data with neural network embedding: A case study in Greater London. *Computers, Environment and Urban Systems* 88 (2021), 101651.

[25] Salma Ommi and Mohammad Hashemi. 2024. Machine learning technique in the north zagros earthquake prediction. *Applied Computing and Geosciences* 22 (2024), 100163.

[26] A Psyllidis, S Gao, Y Hu, EK Kim, G McKenzie, R Purves, M Yuan, and C Andris. 2022. Points of Interest (POI): A commentary on the state of the art, challenges, and prospects for the future. Computational Urban Science, 2 (1), 20.

[27] Esther Rolf, Jonathan Proctor, Tamma Carleton, Ian Bolliger, Vaishaal Shankar, Miyabi Ishihara, Benjamin Recht, and Solomon Hsiang. 2021. A generalizable and accessible approach to machine learning with global satellite imagery. *Nature communications* 12, 1 (2021), 4392.

[28] Hanlin Sheng, Xinming Wu, Xu Si, Jintao Li, Sibo Zhang, and Xudong Duan. 2025. Seismic foundation model: A next generation deep-learning model in geophysics. *Geophysics* 90, 2 (2025), IM59–IM79.

[29] Nicolas Tempelmeier, Simon Gottschalk, and Elena Demidova. 2021. GeoVectors: a linked open corpus of OpenStreetMap Embeddings on world scale. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 4604–4612.

[30] Vicente Vivanco Cepeda, Gaurav Kumar Nayak, and Mubarak Shah. 2023. Geoclip: Clip-inspired alignment between locations and images for effective worldwide geo-localization. *Advances in Neural Information Processing Systems* 36 (2023), 8690–8701.

[31] Lixia Wu, Jianlin Liu, Junhong Lou, Haoyuan Hu, Jianbin Zheng, Haomin Wen, Chao Song, and Shu He. 2023. G2ptl: A pre-trained model for delivery address and its applications in logistics system. *arXiv preprint arXiv:2304.01559* (2023).

[32] Xinhua Wu, Haoyu He, Yanchao Wang, and Qi Wang. 2024. Pretrained mobility transformer: A foundation model for human mobility. *arXiv preprint arXiv:2406.02578* (2024).

[33] Congxi Xiao, Jingbo Zhou, Yixiong Xiao, Jizhou Huang, and Hui Xiong. 2024. ReFound: Crafting a Foundation Model for Urban Region Understanding upon Language and Visual Foundations. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 3527–3538.

[34] Yibo Yan, Haomin Wen, Siru Zhong, Wei Chen, Haodong Chen, Qingsong Wen, Roger Zimmermann, and Yuxuan Liang. 2024. Urbanclip: Learning text-enhanced urban region profiling with contrastive language-image pretraining from the web. In *Proceedings of the ACM Web Conference 2024*. 4006–4017.

[35] Wei Zeng, Chi-Wing Fu, Stefan Müller Arisona, Simon Schubiger, Remo Burkhard, and Kwan-Liu Ma. 2017. Visualizing the relationship between human mobility and points of interest. *IEEE Transactions on Intelligent Transportation Systems* 18, 8 (2017), 2271–2284.

[36] Yuanshao Zhu, James Jianqiao Yu, Xiangyu Zhao, Xuetao Wei, and Yuxuan Liang. 2024. UniTraj: Learning a Universal Trajectory Foundation Model from Billion-Scale Worldwide Traces. *CoRR* (2024).