SRL: Towards a General-Purpose Framework for Spatial Representation Learning

Gengchen Mai
University of Texas at Austin
gengchen.mai@austin.utexas.edu

Xiaobai Yao
University of Georgia
xyao@uga.edu

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

Jinmeng Rao
Google DeepMind
jinmengrao@google.com

Hao Li
Technical University of Munich
hao_bgd.li@tum.de

Qing Zhu
Lawrence Berkeley National Lab
qzhu@lbl.gov

Ziyuan Li
University of Connecticut
ziyuan.2.li@uconn.edu

Ni Lao
Google
nlao@google.com

ABSTRACT

Representation learning (RL) techniques are widely adopted in areas such as natural language processing and computer vision, with prominent examples such as attention and ConvNet architectures. In comparison, many GeoAI works still rely on feature engineering or data conversion to represent spatial data (e.g., points, polylines, polygons, 3D building models, etc.) as features in formats that are easier for neural networks to handle. The neural network architectures remain unchanged, and the need for feature engineering has become a bottleneck for applying deep learning to new tasks in the age of big data. In this paper, we advocate the idea of developing learnable spatial representation modules, which not only enable spatial reasoning but also enable neural nets to directly consume (i.e., encoding) or generate (i.e., decoding) spatial data. We propose **Spatial Representation Learning (SRL)**, a new general-purpose representation learning framework for spatial reasoning. We discuss the key challenges of spatial representation learning including multi-scale RL, continuous RL, shape-centric RL, noise-robust RL, heterogeneity-aware RL, and fairness-aware RL. We also discuss the critical role and potential of SRL in various geospatial subdomains and how this technique can lead to a new generation of GeoAI.

CCS CONCEPTS

• **Computing methodologies** → **Neural networks**; **Learning latent representations**; • **Applied computing** → **Earth and atmospheric sciences**.

KEYWORDS

Spatial representation learning, spatially explicit artificial intelligence, location encoding, polygon encoding

ACM Reference Format:
Gengchen Mai, Xiaobai Yao, Yiqun Xie, Jinmeng Rao, Hao Li, Qing Zhu, Ziyuan Li, and Ni Lao. 2024. SRL: Towards a General-Purpose Framework for Spatial Representation Learning. In *The 32nd ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL '24), October 29-November 1, 2024, Atlanta, GA, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3678717.3691246

SIGSPATIAL '24, October 29-November 1, 2024, Atlanta, GA, USA

ACM ISBN 979-8-4007-1107-7/24/10.
https://doi.org/10.1145/3678717.3691246

1 INTRODUCTION

Representation learning, as a new paradigm of feature extraction, has become the foundation for dramatic performance improvements when solving various AI tasks such as natural language processing, computer vision, and speech recognition. Text, images, and videos data can be directly fed into dedicated neural network architecture modules to automatically learn low-level and high-level representations without the need for the classic feature engineering step. In contrast, we see a lack of such representation learning techniques for various types of spatial data (e.g., points, polylines, polygons, triangulated irregular networks (TINs), 3D LiDAR point clouds, 3D building models, etc.). To extract meaningful features from spatial data for downstream tasks, many researchers either perform feature engineering based on domain knowledge [27], or convert spatial data from their original formats (e.g., points, polylines, and polygons) into formats that are easier for neural networks to handle (e.g., point clouds to voxels [30], or map vector files into raster image tiles [9]). The first approach heavily relies on domain knowledge and can not be easily generalized to new tasks, while the latter approach suffers from reduced data precision and increased data storage requirements. Either way, the system's overall performance is limited due to the lack of end-to-end learning.

In this vision paper, we propose a new general-purpose representation learning framework for spatial data, called **Spatial Representation Learning (SRL)**, which aims at *directly learning neural spatial representations of various types of spatial data in their native data format without the need for any feature engineering or data conversion stage.* Compared with other approaches discussed above, SRL has several key advantages: 1) Less domain knowledge is required for feature extraction so the SRL network can be easily utilized on various tasks without modification; 2) It enables end-to-end training, which has great potential for deep learning models to significantly improve performances; 3) It eliminates the need for sophisticated data preprocessing and model output postprocessing.

In fact, various pioneering works have been done to explore the possibility of directly consuming or generating spatial data in its native data format such as location encoders [14–16, 24], polyline encoders [19, 21], polygon encoders [13, 18], etc. However, these works apply existing neural networks in ad hoc manners. There

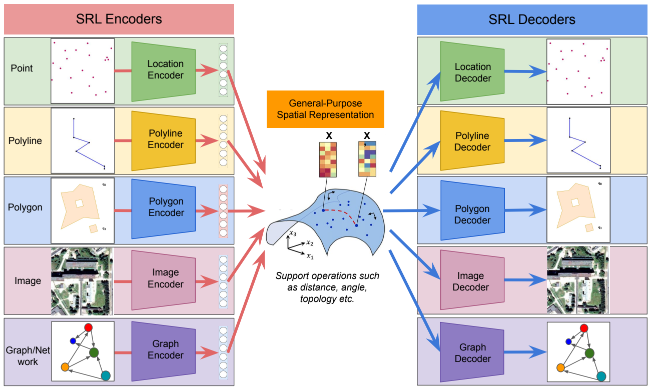


Figure 1: The overall framework of SRL. Here we demonstrate with a few representative spatial data formats.

is no theoretical framework that discusses the unique challenge SRL needs to solve, what is a **general spatial representation** that supports reasoning and is also learning-friendly, how this powerful technique means to the whole GeoAI domain, and what should be done next. In this paper, we formally discuss the advantages of SRL, the unique challenges of SRL, and how it can benefit research across various disciplines.

2 DEFINITIONS AND KEY COMPONENTS

First, we discuss definitions of SRL and its key components. Spatial Representation Learning aims at learning neural network representations of various spatial data such as points, polylines, polygons, TINs, 3D building models, raster images, etc. SRL has two key components: 1) **SRL encoders** can directly consume spatial data and output learning-friendly neural spatial representations for downstream tasks; 2) **SRL decoders** can take neural representations and directly *generate* spatial data in the dedicated format.

According to the input spatial data types, SRL encoders can be classified into location encoders, polyline encoders, polygon encoders, graph encoders, 3D mesh encoders, image encoders, etc. Similarly, SRL decoders can be classified into location decoders, polyline decoders, polygon decoders, graph decoders, 3D mesh decoders, image decoders, etc. Figure 1 illustrates the relations between these models under the SRL framework. While some of these architectures have been widely studied such as location encoders [15], image encoders, and image decoders [3] many of these directions have never been explored or significant drawbacks exist in the current solutions due to the neglect of the uniqueness of spatial data such as polygon encoders, location decoders, polygon decoders, 3D mesh generators, etc.

3 THE UNIQUE CHALLENGES IN SRL

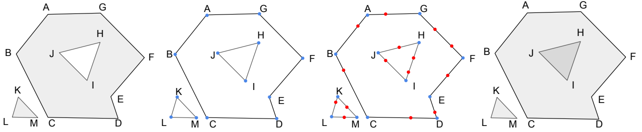
Compared with representation learning on other data types, SRL demonstrates several unique challenges that are usually ignored by pioneer research. In this section, we briefly discuss those challenges in the hope that they can guide future SRL model development.

3.1 Multi-scale Representation Learning

While sharing some similarities, representing a geographic location into an embedding space is more challenging than representing a point/location in other spaces (e.g., an indoor robot's location). One big challenge of the former is the need for *multi-scale representation learning*. Although in theory space is continuous, in many situations, we can discretize the space into regular grids so that location encoding and location decoding in these spaces becomes a simple finite location/grid embedding matrix learning [4] and spatial softmax [2] problem. Examples are determining the location of a robot within an indoor environment [2]. In fact, many pioneering works in geographic location encoding [22] and geographic location decoding [23] adopted a similar practice.

One big problem is that the geographic location representation learning problem usually requires a model to jointly consider multiple scales ranging from global scale, country scale, city scale, neighborhood scale, etc. For example, an image geolocalization model needs to start from the whole globe to precisely determine where the given image was taken. While many previous works [23] used hierarchical grid systems to progressively narrow down the search space, the model performance is still bounded by the smallest grid size, thus leading to a systematic model bias cannot be avoided. Currently, many multi-scale location encoders have been developed [15]. However, we have yet to see any multi-scale geographic location decoders that can precisely *regress* the precise geo-location while avoiding the discretization step. In fact, multi-scale representation learning is not only a unique challenge for location representation learning but also a problem for other more complex spatial data types.

3.2 Continuous Representation Learning



(a) Polygon p **(b) p's vertices** **(c) Trivial vertices** **(d) Polygon p′**

Figure 2: (a) Multipolygon p conceptually indicates a continuous bounded surface. (b) Many works [13, 28] only represent p as its boundary information, i.e., a list of vertices. (c) When adding trivial vertices (i.e., red dots) to the shape, conceptually the shape of p does not change. However, most current polygon representation learning models will lead to different polygon embeddings. (d) Multipolygon p′ shares the same set of vertices with p. However, instead of a hole of p, Triangle $\triangle_{HIJ}$ now is a part polygon of p′.

Another unique challenge of SRL is that although spatial data are serialized in a discretized format, they represent continuous objects conceptually. This requires SRL models to learn continuous representations of them instead of treating them as a finite list of coordinates. For example, a polyline is represented as a list of vertices while it indicates a continuous linear-shaped feature. Similarly, a polygon is usually represented as a ring of vertices or several rings (polygons with holes or multipolygons). However, it represents a continuous surface (see Figure 2a).

Most previous work failed to capture the continuous nature of spatial data but treating them as a finite list of coordinates. For example, PolyFormer [13] and RoomFormer [28] treat a polygon as a list of vertices. As shown in Figure 2, instead of interpreting Multipolygon **p** as a continuous bounded surface in Figure 2a, all above models only consider boundary vertices of **p**. This practice disables the possibility of neural networks to learn the topological nature of polygons, i.e., they can not differentiate **p** and **p′** shown in Figure 2a and 2d. So they cannot perform topological relation computation among polygons [18].

3.3 Shape-Centric Representation Learning

Another key challenge of SRL is allowing the learned representations to focus on the shape-level information while being invariant of shape-invariant transformations. Here, we list several expected properties for SRL: 1) **Vertex loop transformation invariance**: the learned spatial representations should be invariant under vertex loop transformation. For example, the exterior of a part of Multipolygon **p** in Figure 2a can be represented as an ordered vertex list $\mathbb{M}_1 = [A, B, C, D, E, F, G]$. After vertex loop transformations, $\mathbb{M}_1$ becomes $\mathbb{M}_2 = [D, E, F, G, A, B, C]$ or $\mathbb{M}_3 = [B, C, D, E, F, G, A]$. While they are different, they all indicate the same polygon shape, so the SRL model should output identical embeddings. This is not limited to polygonal features but also TIN data, 3D meshes, etc. 2) **Trival vertex invariance**: adding or deleting trivial vertices to/from the spatial data will not change the learned spatial representations by SRL models. The red dots in Figure 2c are trivial vertices since adding or deleting them will not affect the shape of **p**, thus this shape-invariant operation should not change the polygon embeddings. The same property is also expected for polyline encoders/decoders, 3D mesh encoders/decoders, etc.

3.4 Noise-Robust Representation Learning

Another challenge is to learn a noise-robust representation for spatial data. In practice, spatial data is prone to errors and noise during data collection and map editing such as sliver polygons [18]. When we compute the strict spatial relations between map features (i.e., topological relations), these errors might lead to wrong answers. Learning noise-robust spatial representations will simplify the data preprocessing and improve the representations' quality.

3.5 Heterogeneity-Aware RL

While it is ideal to learn a single general representation that works well across space, technically this may not be feasible due to spatial heterogeneity. In the context of ML, this means the functional relationships between inputs and outputs tend to vary across geographic regions [5, 25]. For example, the observable inputs in real-world problems often only cover a subset of variables that influence the target outputs. Thus, a spatially stationary function cannot reflect the variability of unobserved variables, constraining the prediction quality. Currently, most SRL methods do not consider spatial heterogeneity [16], making the results sub-optimal and harder to adapt to large-scale applications. It is important to develop heterogeneity-aware frameworks to embed location-induced heterogeneity into the learned representations. Potential directions include a separate representation of heterogeneity or a partitioning-based separation of representation learning [26].

3.6 Fairness-Aware Representation Learning

The increasing deployment of ML in applications has drawn significant attention to model fairness. This is no stranger to spatial data as geographic bias induced by ML models can easily occur without explicit fairness-aware formulations [26], and such biases might be propagated to high-stake decision-making. The need for spatial fairness also presents unique challenges for SRL. For instance, existing fairness metrics often rely on discrete groups such as racial/gender groups while space is continuous. Although we can do space partitions, the quantification of bias is affected by the modifiable areal unit problem, making the results sensitive to both the partitioning scheme and the scale of the analysis [7].While recent works have started exploring fairness issues for spatial applications [6], it remains largely unexplored for SRL. In addition, when building fairness-aware spatial representations, it is important to address challenges including the generalizability of the embedded fairness in unseen geographic regions and future time periods.

4 THE POTENTIAL OF SRL IN GEOAI

4.1 Human Mobility

Human mobility research has been leveraging a vast array of data, e.g., GPS trajectories, mobile phone records, smart card data, and social media location data, to gain insights into human dynamics. Human mobility analysis, encompassing tasks like trajectory classification, clustering, prediction, and generation, has enormous application potential in urban planning, marketing and retail, smart cities, epidemiology, and public health research and practices. SRL plays a critical role by enabling the effective encoding and decoding of complex spatiotemporal patterns inherent in movement data. Other than those challenges discussed in Section 3, we identify three key challenges when developing SRL for human mobility tasks: **1) Intent/Behavior Preservation:** The representation of movement data should capture not only the geometries of trajectories but also the underlying intent or behavior of individuals. Representations that incorporate semantic information about points of interest, transportation modes, or social interactions can benefit downstream tasks like activity recognition or trip purpose inference. **2) Geo-Privacy Preservation:** Given the sensitive nature of location data, privacy preservation is paramount. SRL should incorporate mechanisms that anonymize or obfuscate individual trajectories while still retaining the essential patterns for analysis [19], This could involve differential privacy [1], decentralized learning [20], trajectory generalization, or learning privacy-preserved representations for trajectories [21]. **3) Data Bias:** A considerable portion of popular human mobility datasets comprises passive data, indicating that data collectors have little control over their availability and frequency. Integrating such data presents significant challenges in addressing data biases and data fusion.

4.2 Remote Sensing

While general computer vision and representation learning fields have made major breakthroughs, the unique characteristics of remote sensing (RS) data limit their abilities on RS tasks. In this context, we envision a critical role of SRL in the next generation of RS-based GeoAI models. Specifically, we identify three pressing challenges: **1) Consistent Level-of-Details (LoD) representation**: similar to vector geometries, it is key to encode RS data of various spatial and temporal resolutions into a uniform and resolution-agnostic latent representation so that downstream tasks can easily handle consistent LoD without additional upsampling and downsampling. **2) Spatial-spectral data representation:** spectral imaging enables accurate analysis of objects and scenes beyond what is possible with regular RGB aerial images but hinders the adoption of state-of-the-art pre-training weights (e.g., from ImageNet) to many RS data, such as multispectral and hyperspectral imagery, thus deserves a dedicated spatial-spectral encoding strategy (e.g., 3D tensor mask in [3]). **3) Location-aware representation:** it is simply disappointing when an established GeoAI model fails entirely in a slightly different geographical setup, e.g., different countries or landscapes, which is known as replication across space

[5], geographical generalizability [12], or spatial heterogeneity [25]. A location-aware representation is extremely helpful in leveraging the location encoding and RS data representation to gain generalizability across space [17, 25]. Furthermore, it is intuitive to foresee the great potential of SRL of RS data in quantum computing and edge AI applications [31].

4.3 Cartography

As a critical field of geography, cartography presents unique challenges for SRL: **1) Multi-Scale representation for map generalization**: To ensure map quality and readability, the selection of spatial objects and their symbolization may vary depending on the map scale. Developing SRL models that generalize well across different map types and scales is a key challenge. **2) SRL for historic map digitization**: Developing SRL models for map data extraction and digitization from historic maps requires the learned representations capable of complex spatial reasoning among map symbols, text, and textures, and grounding them to a geo-knowledge base [10]. **3) 3D maps and high-dimensional visualization:** With the trend of digital twins and metaverse, the vertical dimension of maps has been explored with techniques like 3D, leading to a pressing need for robust and effective SRL in 3D space and beyond. A key challenge here is how to adapt to sparse and entangled geometry and spatial relationships in a 3D world.

4.4 Earth System Science

SRL holds substantial promise for advancing coupled Earth system modeling and enhancing our understanding of Earth system science (ESS) which can help to address grand challenges in predictive modeling, process-level interpretation, parameterization, uncertainty reduction, and long-term projections, thus leading to a more accurate, efficient, and comprehensive understanding of the coupled earth system. We identify unique challenges of SRL for ESS: **1) SRL for parameterizations derivation of sub-grid scale processes:** SRL should be able to derive these parameterizations that are not captured at typical earth grid cell scales (10 km) by learning from patterns identified from high-resolution data. This can enhance the representation of these processes in larger-scale models, leading to better predictions of local patterns and climate change impacts [8, 29]. **2) SRL for dynamic ESM simulations:** SRL should be learning efficiently to transform real-time data assimilation for dynamic ESM simulations. ML surrogate models trained on spatial data can quickly incorporate new datasets at different scales which are particularly beneficial in e.g., weather forecasting and climate disaster prediction/response, where timely and accurate forecasts are crucial [11]. **3) SRL for long-term scenario analysis:** SRL should be adapted for scenario analysis over long-term time periods (annual, decadal, and centurial scales) enabling more effective decision-making in climate policy.

5 CONCLUSION

In this vision paper, we introduce a general-purpose representation learning framework for spatial data called Spatial Representation Learning which aims at learning a general spatial representation that supports reasoning and is learning-friendly. A theoretical framework of SRL is provided which discusses the unique challenges of SRL and how SRL can benefit research across various disciplines. We believe SRL is a unique and key question in the GeoAI domain that requires community efforts.

ACKNOWLEDGMENTS

Yiqun Xie is supported in part by the National Science Foundation under Grant No. 2126474, 2147195, and 2425844, Google's AI for Social Good Impact Scholars program, and the Zaratan supercomputing cluster at the University of Maryland. Gengchen Mai is partially supported by the Microsoft Research AFMR Initiative.

REFERENCES

[1] Andrés et al. 2013. Geo-indistinguishability: Differential privacy for location-based systems. In *ACM SIGSAC 2013*. 901–914.

[2] Cheng Chi et al. 2023. Diffusion policy: Visuomotor policy learning via action diffusion. *arXiv preprint arXiv:2303.04137* (2023).

[3] Yezhen Cong et al. 2022. SatMAE: Pre-training Transformers for Temporal and Multi-Spectral Satellite Imagery. In *NeurIPS 2022*.

[4] Ruiqi Gao et al. 2022. Learning V1 Simple Cells with Vector Representation of Local Content and Matrix Representation of Local Motion. In *AAAI 2022*.

[5] Michael F Goodchild and Wenwen Li. 2021. Replication across space and time must be weak in the social and environmental sciences. *PNAS* 118, 35 (2021).

[6] Erhu He et al. 2023. Physics Guided Neural Networks for Time-aware Fairness: An Application in Crop Yield Prediction. In *AAAI 2023*.

[7] Erhu He et al. 2024. Learning With Location-Based Fairness: A Statistically-Robust Framework and Acceleration. *TKDE* (2024).

[8] Christopher Irrgang et al. 2021. Towards neural Earth system modelling by integrating artificial intelligence in Earth system science. *Nature Machine Intelligence* 3, 8 (2021), 667–674.

[9] Yuhao Kang and et al. 2019. Transferring multiscale map styles using generative adversarial networks. *International Journal of Cartography* 5, 2-3 (2019), 115–141.

[10] Jina Kim et al. 2023. The mapKurator System: A Complete Pipeline for Extracting and Linking Text from Historical Maps. In *ACM SIGSPATIAL 2023*. 35.

[11] Yochanan Kushnir et al. 2019. Towards operational predictions of the near-term climate. *Nature Climate Change* 9, 2 (2019), 94–101.

[12] Hao Li et al. 2023. Rethink Geographical Generalizability with Unsupervised Self-Attention Model Ensemble: A Case Study of OpenStreetMap Missing Building Detection in Africa. In *ACM SIGSPATIAL 2023*.

[13] Jiang Liu et al. 2023. Polyformer: Referring image segmentation as sequential polygon generation. In *CVPR 2023*. 18653–18663.

[14] Oisin Mac Aodha et al. 2019. Presence-only geographical priors for fine-grained image classification. In *ICCV 2019*. 9596–9606.

[15] Gengchen Mai et al. 2020. Multi-Scale Representation Learning for Spatial Feature Distributions using Grid Cells. In *ICLR 2020*. openreview.

[16] Gengchen Mai et al. 2022. A review of location encoding for GeoAI: methods and applications. *IJGIS* 36, 4 (2022), 639–673.

[17] Gengchen Mai et al. 2023. Csp: Self-supervised contrastive spatial pre-training for geospatial-visual representations. In *ICML 2023*. PMLR, 23498–23515.

[18] Gengchen Mai et al. 2023. Towards general-purpose representation learning of polygonal geometries. *GeoInformatica* 27, 2 (2023), 289–340.

[19] Jinmeng Rao et al. 2020. LSTM-TrajGAN: A Deep Learning Approach to Trajectory Privacy Protection. In *GIScience 2020*. 12:1–12:17.

[20] Jinmeng Rao et al. 2021. A privacy-preserving framework for location recommendation using decentralized collaborative machine learning. *TGIS* (2021).

[21] Jinmeng Rao et al. 2023. CATS: Conditional Adversarial Trajectory Synthesis for privacy-preserving trajectory data publication using deep learning approaches. *International Journal of Geographical Information Science* 37, 12 (2023), 2538–2574.

[22] Kevin Tang et al. 2015. Improving image classification with location context. In *ICCV 2015*. 1008–1016.

[23] Nam Vo et al. 2017. Revisiting im2gps in the deep learning era. In *ICCV 2017*.

[24] Nemin Wu et al. 2024. TorchSpatial: A Location Encoding Framework and Benchmark for Spatial Representation Learning. *arXiv preprint arXiv:2406.15658* (2024).

[25] Yiqun Xie et al. 2021. A statistically-guided deep network transformation and moderation framework for data with spatial heterogeneity. In *ICDM*. 767–776.

[26] Yiqun Xie et al. 2023. Harnessing Heterogeneity in Space with Statistically-Guided Meta-Learning. *KAIS* 65 (2023), 2699–2729.

[27] Xiongfeng Yan et al. 2022. A graph deep learning approach for urban building grouping. *Geocarto International* 37, 10 (2022), 2944–2966.

[28] Yuanwen Yue et al. 2023. Connecting the dots: Floorplan reconstruction using two-level queries. In *CVPR 2023*. 845–854.

[29] Keer Zhang et al. 2023. A global dataset on subgrid land surface climate (2015–2100) from the Community Earth System Model. *Geoscience Data Journal* 10, 2 (2023), 208–219.

[30] Yin Zhou and Oncel Tuzel. 2018. Voxelnet: End-to-end learning for point cloud based 3d object detection. In *CVPR 2018*. 4490–4499.

[31] Johann Maximilian Zollner et al. 2024. Satellite Image Representations for Quantum Classifiers. *Datenbank-Spektrum* (2024), 1–9.