The University of Texas at Austin
Department of Geography & the Environment
College of Liberal Arts

SPATIALLY EXPLICIT AI

TEXAS
The University of Texas at Austin

Spatial Representation Learning

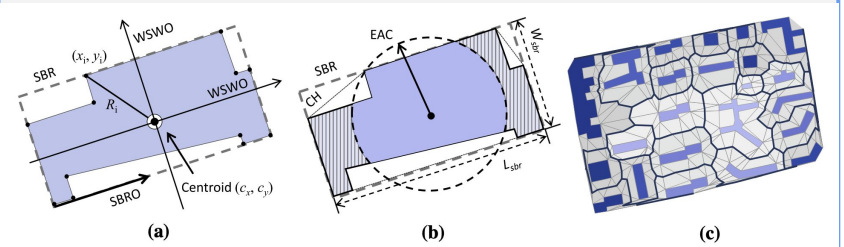
Gengchen Mai

Department of Geography and the Environment
University of Texas at Austin

Oct. 31st, 2024

Comprised Approaches Due to the Lack of SRL

Feature Engineering



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

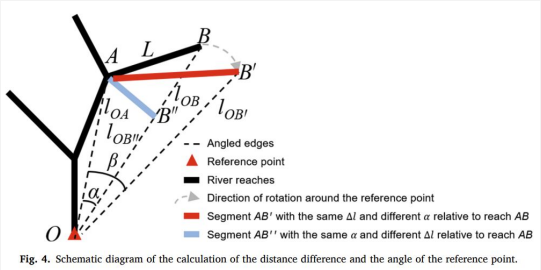
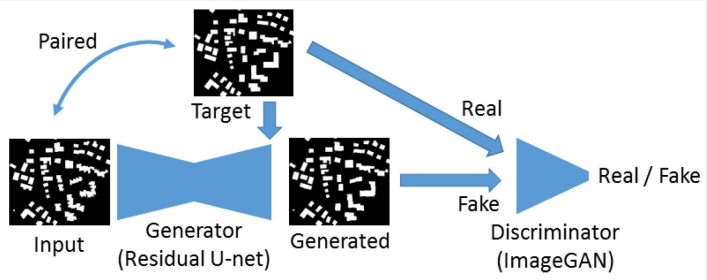


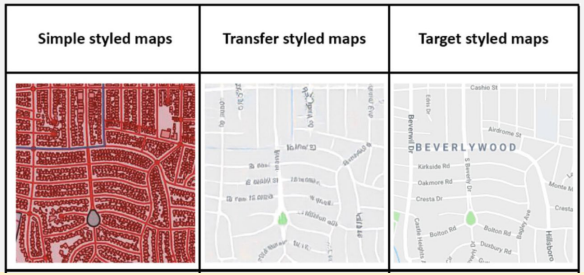
Fig. 4. Schematic diagram of the calculation of the distance difference and the angle of the reference point.

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

Data Conversion



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



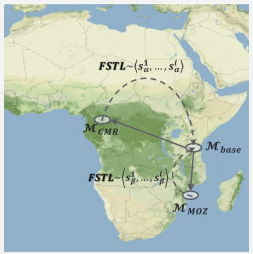
Map vector files to raster images (Kang et al, 2019)

Comprised Approaches Due to the Lack of SRL

Feature Engineering

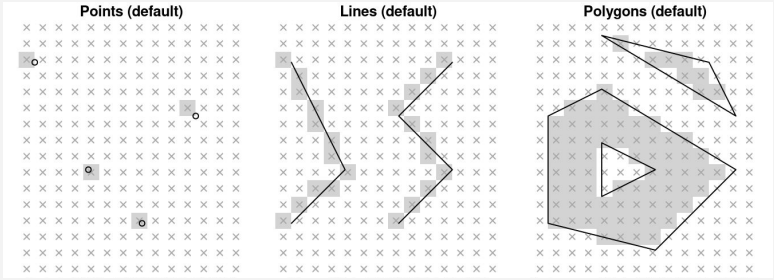
- Heavily relies on domain knowledge

- Hard to generalize to new regions and tasks

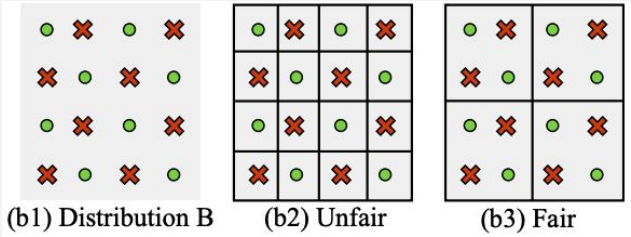


Data Conversion

- Reduced data precision and increase data storage requirement

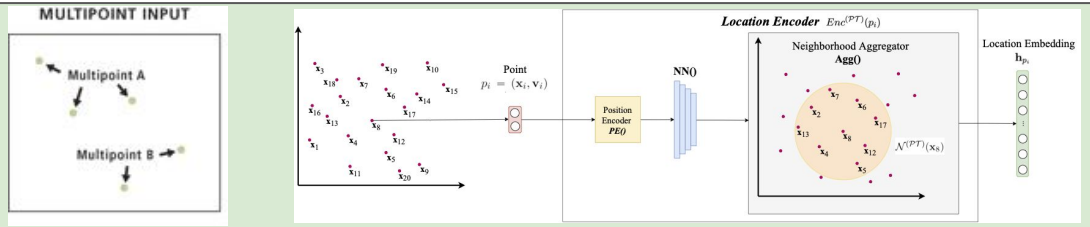
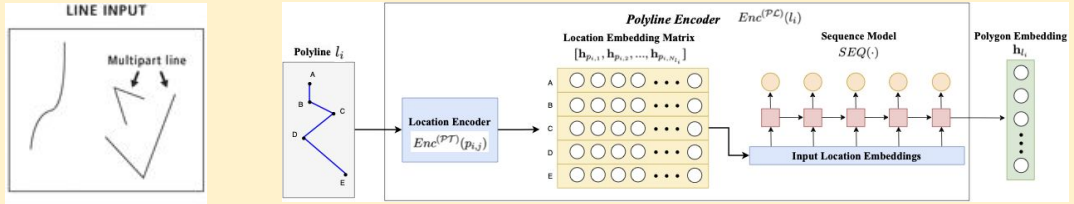
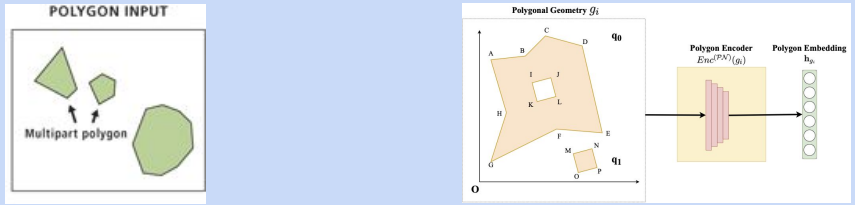


- Modifiable areal unit problem (MAUP)



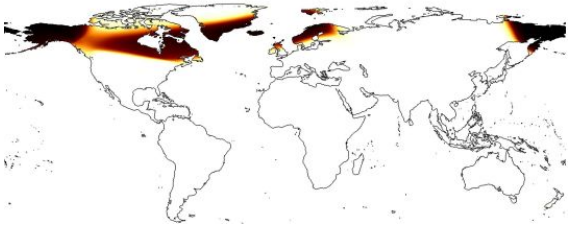
Spatial Representation Learning (SRL)

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

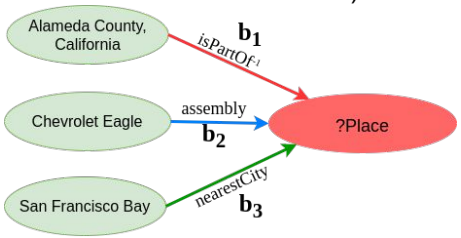
Location Encoding		• Qi et al., CVPR 2017 • Aodha et al. CVPR 2019 • Mai et al., ICLR 2020 • Mai et al., TGIS 2020 • Mai et al., IJGIS 2022 • Mai et al., ICML 2023
Polyline Encoding		• Alahi et al., CVPR 2016 • Rao et al., GIScience 2020 • Yu et al., IJGIS 2022 • …
Polygon Encoding		• Yan et al., ISPRS PHOTO 2021 • Jiang et al., ICCV 2021 • Mai et al., Geoinformatica 2023 • …

Various Geospatial Tasks

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

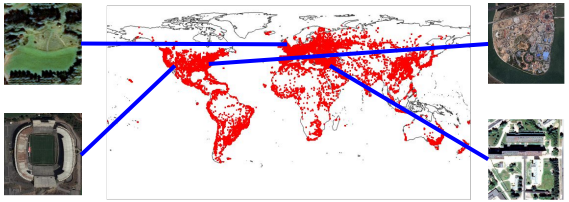


Geospatial Semantics:
Geographic Question Answering
(Mai et al., TGIS 2020; Mai et al., GeoInformatica 2023)

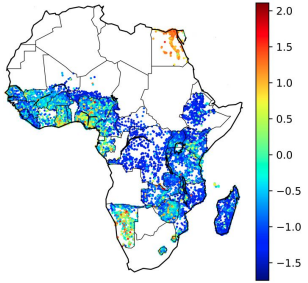


Earth System Science:
Weather Forecasting
(Nguyen et al., ICML 2023)

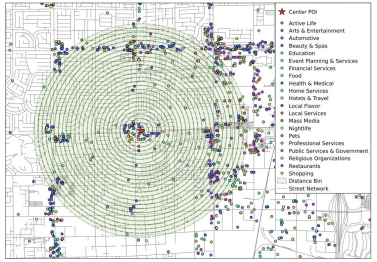
Remote Sensing:
RS Image Classification
(Mai et al., ISPRS PHOTO 2023; Mai et al. ICML 2023; Li et al., SIGSPATIAL 2023)



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



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



Unlabeled v.s. Labeled Geospatial Image Datasets

Restricted Labeled Datasets

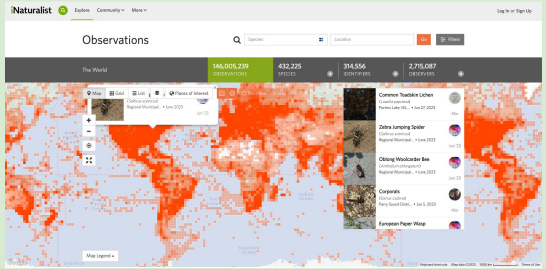
Well-curated satellite images, in contrast, have **limited sizes**, **imbalanced geographic coverage**, and **potentially oversimplified label distributions**

Geographic coverage of labeled species fine-grained recognition dataset – NABird

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

Massive Unlabeled Datasets

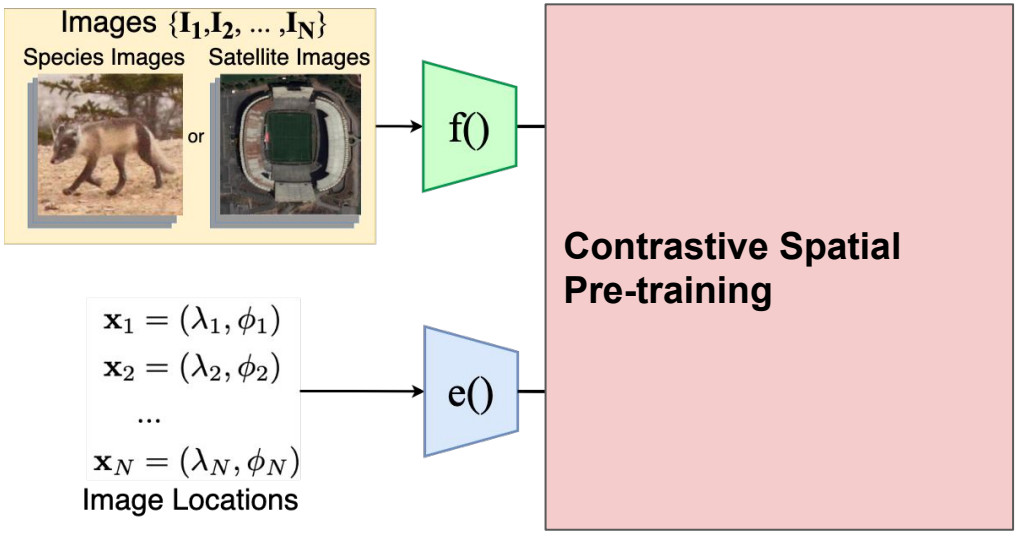
Billions of unlabeled satellite images are collected from various sensors everyday (Figure from NASA Website)



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

A Multimodal Pre-training Objective for GeoAI

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



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

Contrastive Spatial Pre-Training (CSP)

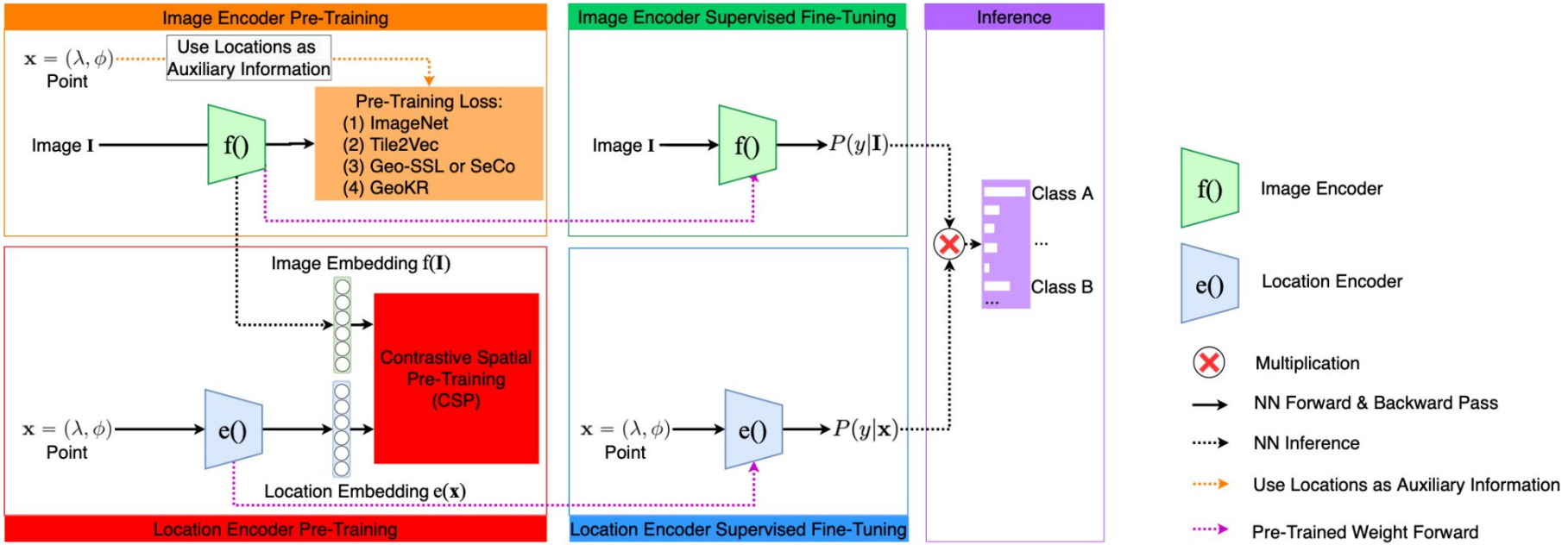


Figure 2(c) Contrastive Spatial Pre-Training (CSP)

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**.

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)**

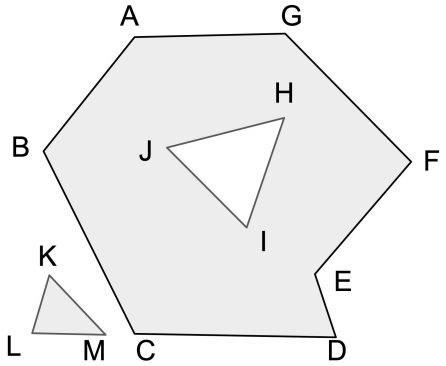
Satellite image scene classification on fMoW dataset

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.

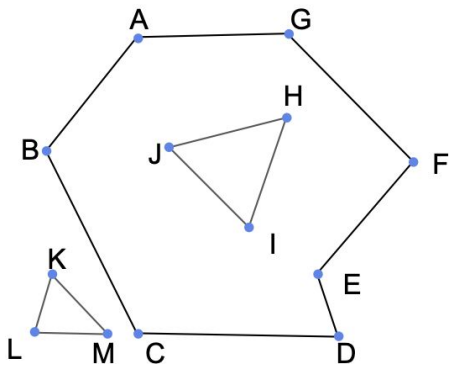
Ratio $\lambda\%$	5%	10%	20%	100%
Img. Only (Tile2Vec) (Jean et al., 2019)	59.41 (0.23)	61.91 (0.31)	62.96 (0.51)	64.45 (0.37)
Img. Only (Geo-SSL) (Ayush et al., 2021)	65.22 (-)	66.46 (-)	67.66 (-)	69.83 (-)
Sup. Only (wrap) (Mac Aodha et al., 2019)	66.67 (0.03)	68.22 (0.01)	69.45 (0.01)	70.30 (0.02)
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)**

SRL: A Continuous Neural Representation

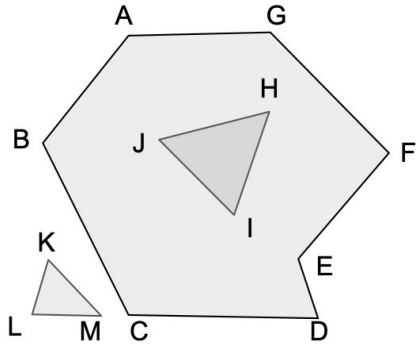
- Although spatial data are serialized in a **discretized format**, they represent **continuous objects** conceptually
- learn **continuous neural representations** of spatial data



(a) Multipolygon **p.** Triangle $\triangle HIJ$ is a **hole** of polygon **p**



(b) Vertices of **p**



(c) Multipolygon **p'.** Triangle $\triangle HIJ$ is a **part** of polygon **p'**

Multipolygon **p** is a continuous bounded surface

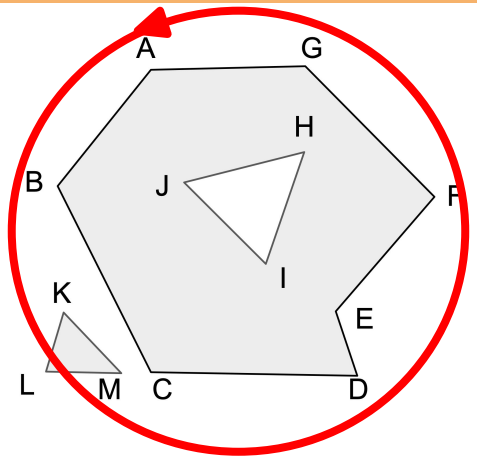
Most previous work such as **PolyFormer** and **RoomFormer** treat as **a finite list of vertices**

They can not differentiate **p** and **p'**

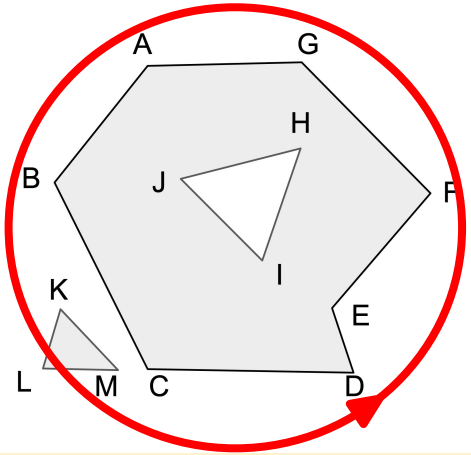
SRL: A Shape-Centric Neural Representation

- Allowing the learned representations to focus on the **shape-level information** while being **invariant of shape-invariant transformations**
- **1) Vertex loop transformation invariance:** the learned spatial representations should be invariant under vertex loop transformation

The exterior of a part of **p** can be represented by **different ordered list of vertices**



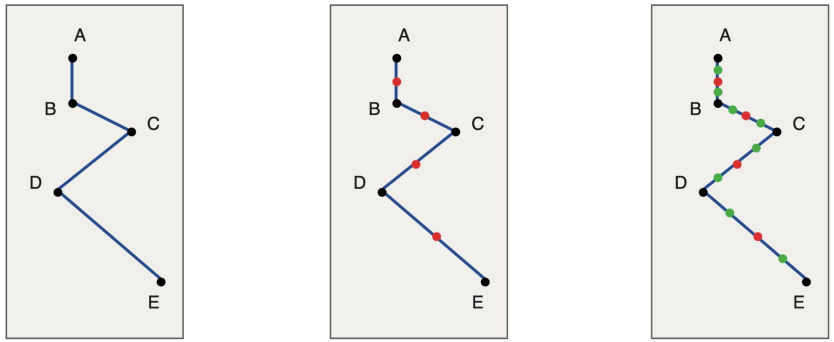
M1 = $[A,B,C,D,E,F,G]$



M2 = $[D,E,F,G,A,B,C]$

SRL: A Shape-Centric Neural Representation

- **2) Trivial vertex invariance**: adding or deleting **trivial vertices** to/from the spatial data will not change the learned spatial representations by SRL

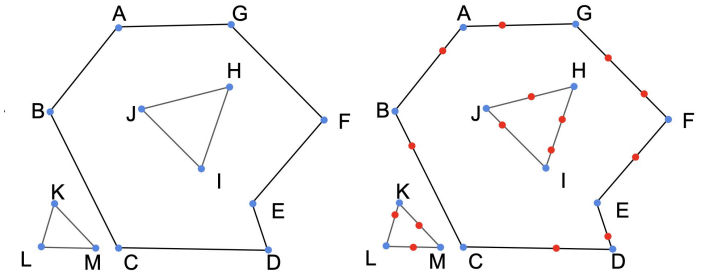


(a) The original coordinate sequence

(b) Adding some trivial vertices

(c) Adding more trivial vertices

One linear feature can be represented as polylines with different vertices

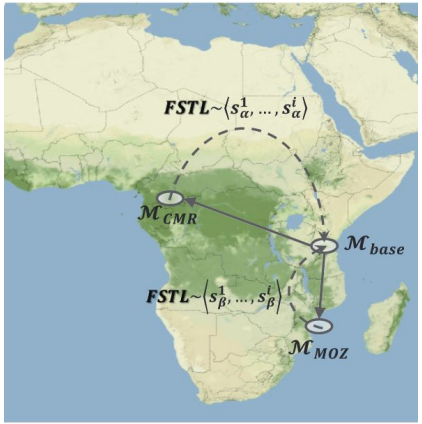


(b) p's vertices (c) Trivial vertices

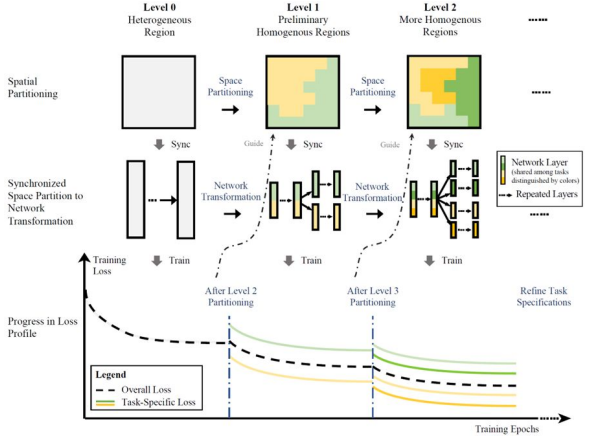
One regional feature can be represented as polygons with different vertices

SRL: A Heterogeneity-Aware Neural Representation

- In reality, the **functional relationships** between inputs and outputs tend to vary across geographic regions
- developing **heterogeneity-aware frameworks** to embed **location-induced heterogeneity** into the learned representations

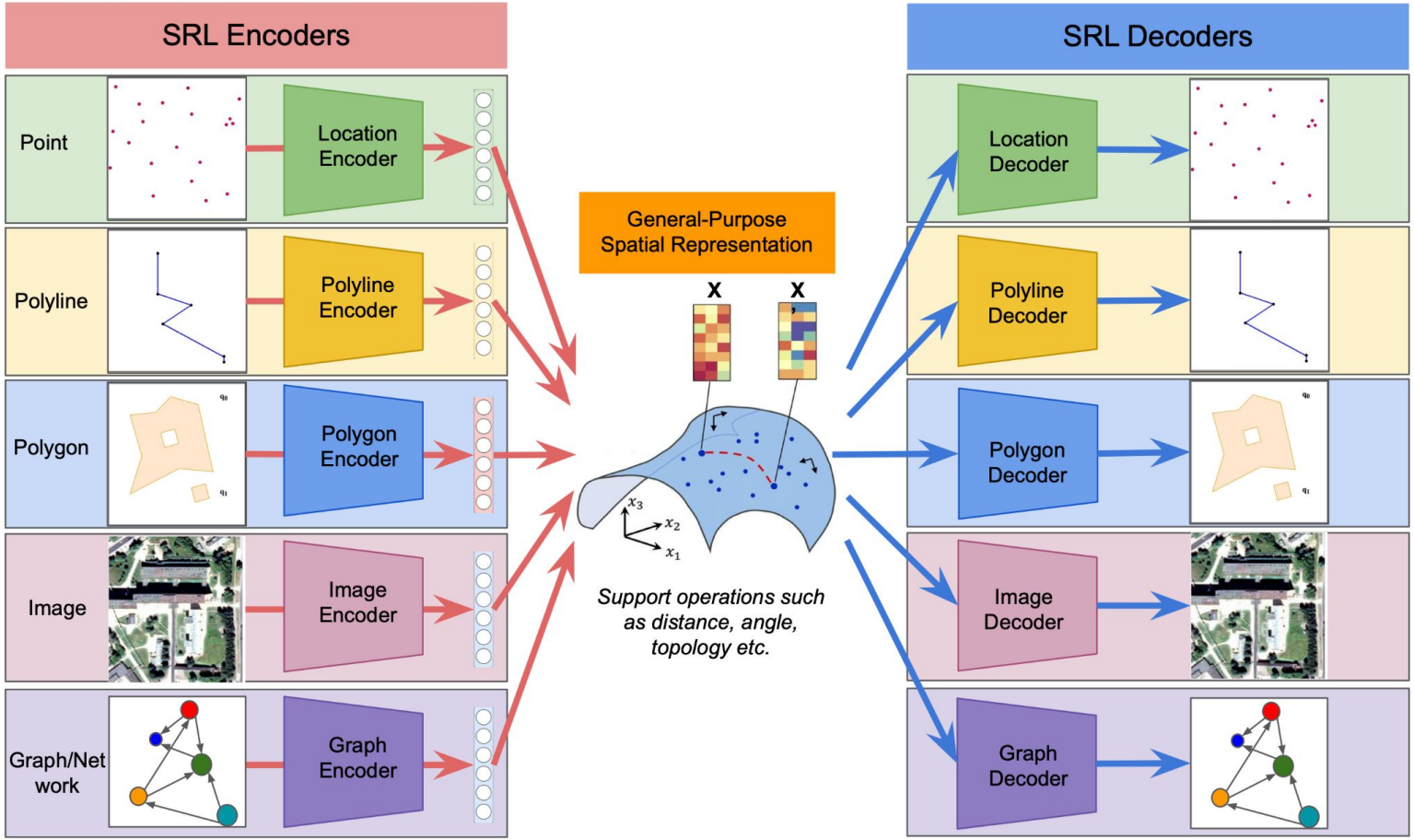


When performing transfer learning from Region A to Region B, we usually see performance degradation



Hierarchically partition the study area into different regions and learn different models

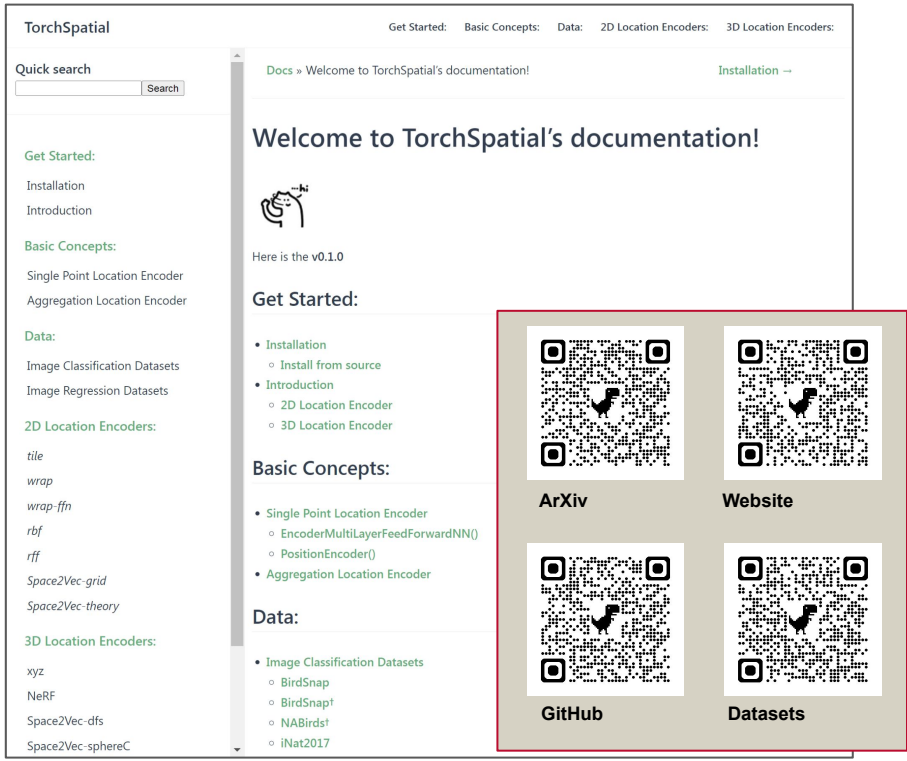
SRL: A Unified Representation Learning Framework



SRL: A Sharable SRL Model Framework

TorchSpatial:

- **A model framework** that consolidates **15 location encoders** and necessary model building blocks for future location encoder.
- **A benchmark** which contains 7 geo-aware image classification and 10 image regression datasets.
- **A set of evaluation metrics** to quantify location encoders' overall model performance as well as their **geographic bias**.



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