# Streamlining Geospatial Machine Learning with SRAI

Szymon Woźniak, Kamil Raczycki, Kacper Leśniara,
Piotr Gramacki and Piotr Szymański

Wrocław University of Science and Technology, Poland
Kraina AI Research Group

## Introduction

Spatial Representations for Artificial Intelligence (*srai*) is a Python library for working with geospatial data. The library can download geospatial data, split a given area into micro-regions using multiple algorithms and train an embedding model using various architectures. It includes baseline models as well as more complex methods from published works. Those capabilities make it possible to use *srai* in a complete pipeline for geospatial task solving. We hope our library will take the first steps to standardize the geospatial AI domain. The library is fully open-source and published under Apache 2.0 license.

Library

About us

Tutorial

## Library design

The *srai* library is organized around four main components:

1. Loader - loading geospatial features from a given source and pre-processing them,
2. Regionalizer - splitting a given area into micro-regions,
3. Joiner - matching loaded features with regions based on a given spatial predicate,
4. Embedder - embedding regions into a vector space based on features matched to them.

Those components create a complete pipeline for learning representations for geospatial data. Solving the final task can be obtained using any machine learning or deep learning library.

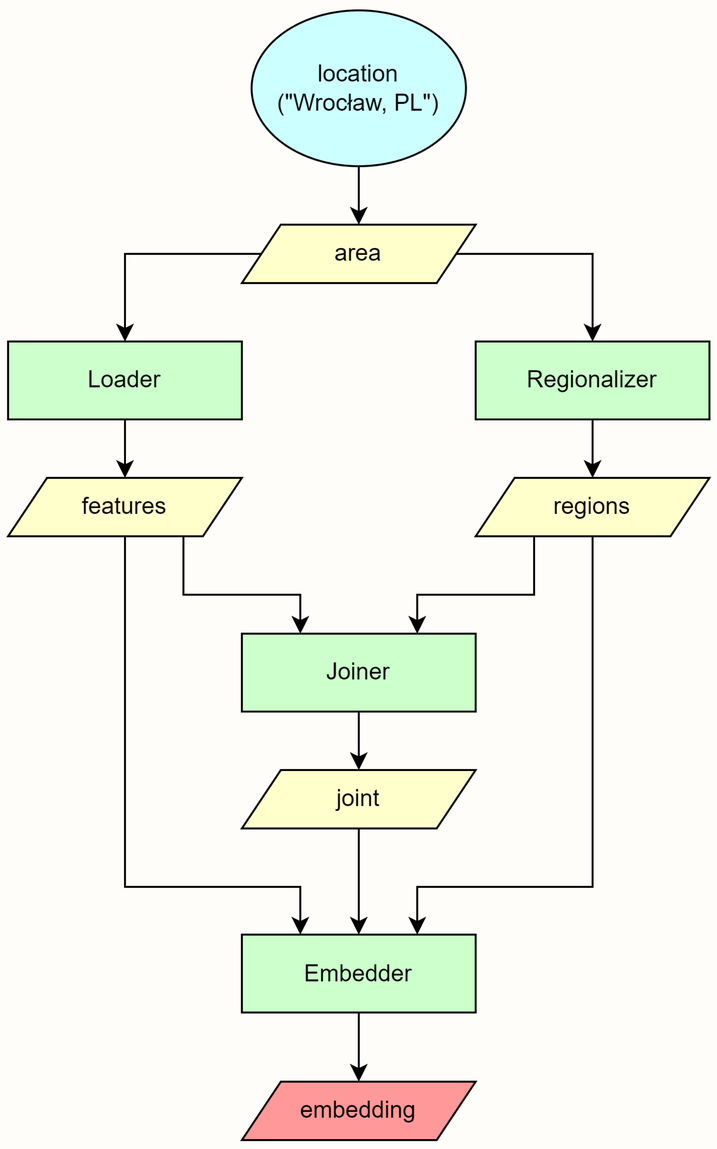


Figure 1: Overview of library design

## Loaders

- OSM tags - downloading tags specified by a filter for a given area.
- OSM networks - downloading structured networks as a graph.
- OSM tiles - downloading tiles with maps as images.
- GTFS - loading data from GTFS feeds and computing public transport offer for stops.
- GeoParquet - loading geospatial data from *geoparquet* files.

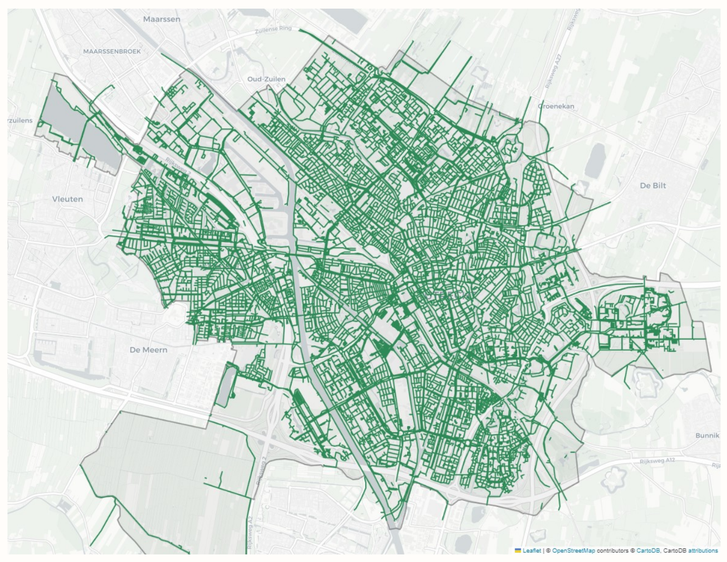


Figure 3: Road network obtained with OSM networks loader

## Regionalizers

- H3 and S2 - hierarchical spatial indices.
- Voronoi - data-driven based on Voronoi cells.
- Administrative boundaries - based on administrative boundaries from OSM.
- Slippy map - regions, which match with OSM map tiles.

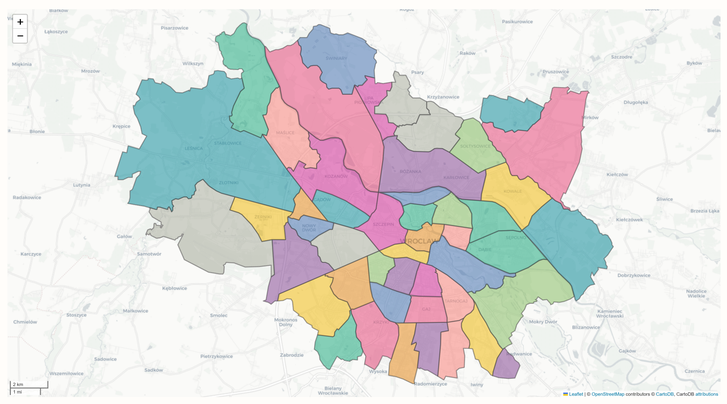

Figure 4: Regionalization based on administrative boundaries

## Pre-trained models

Some of the embedders are trainable and we provide an option to save a pre-trained embedder to a file and quickly load it. We also share a selection of pre-trained hex2vec models and intend to extend this list in time.

## Embedders

- Count - feature counts in regions,
- Contextual Count - contextualized feature counts[1],
- Hex2Vec - OSM tags[2],
- Highway2Vec - road network segments[3],
- GTFS2Vec - public transport offer[4].

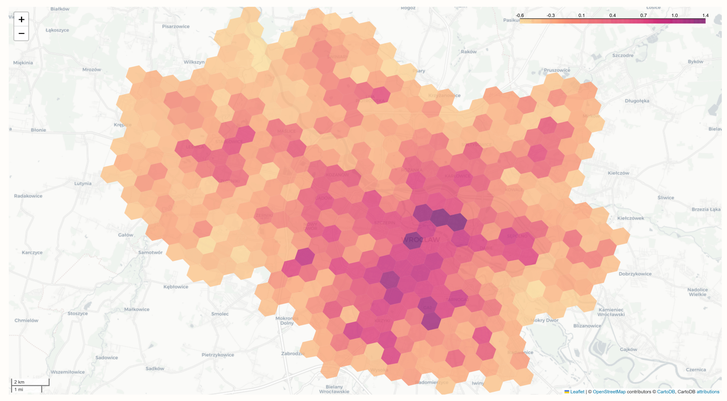

Figure 5: Embeddings for H3 cells obtained with hex2vec

## Future improvements

***Pre-calculated embeddings***

***Fine-tuning***

***Datasets and benchmarks***

## Summary and conclusions

We present a new library for geospatial data processing and machine learning. We cover the whole pipeline for most geospatial tasks based on OSM data. We intend to expand it in multiple areas, bringing us closer to the primary goal of geospatial AI domain standardization. We believe our work will benefit the whole community and be the foundation for geospatial algorithms and models' unification, reproducibility, and reusability.

## *References*

[1] Kamil Raczycki and Piotr Szymański. 2021. Transfer learning approach to bicycle-sharing systems' station location planning using OpenStreetMap data. In *Proceedings of the 4th ACM SIGSPATIAL International Workshop on Advances in Resilient and Intelligent Cities (ARIC '21)*.

[2] Szymon Woźniak and Piotr Szymański. 2021. Hex2vec: Context-Aware Embedding H3 Hexagons with OpenStreetMap Tags. In *Proceedings of the 4th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery (GEOAI '21)*.

[3] Kacper Leśniara and Piotr Szymański. 2022. Highway2vec: representing OpenStreetMap microregions with respect to their road network characteristics. In *Proceedings of the 5th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery (GeoAI '22)*.

[4] Piotr Gramacki, Szymon Woźniak, and Piotr Szymański. 2021. Gtfs2vec: Learning GTFS Embeddings for comparing Public Transport Offer in Microregions. In *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Searching and Mining Large Collections of Geospatial Data (GeoSearch'21)*.