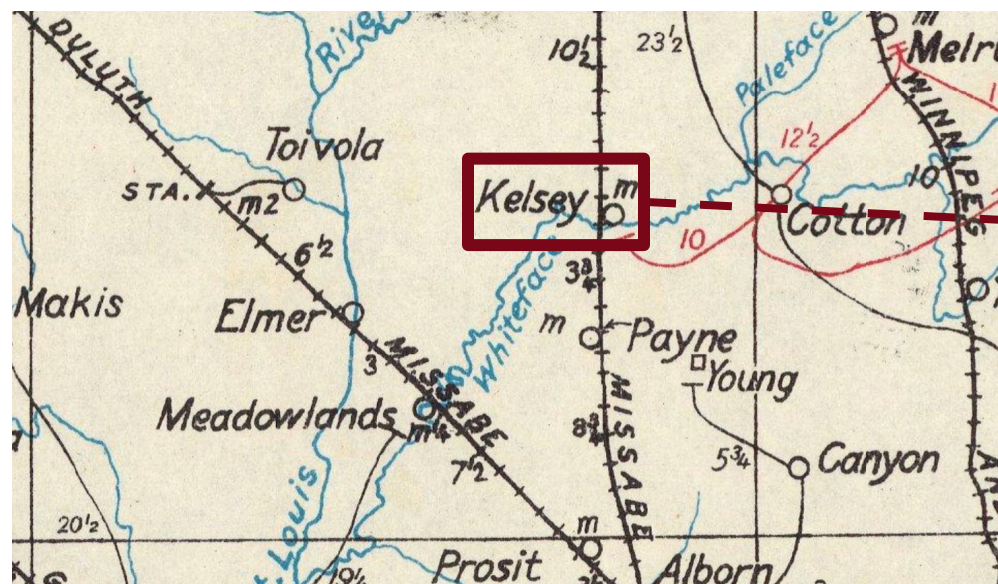# Generating Geospatial Linked Data from Text Labels on Historical Maps

Ph.D. Student Jina Kim
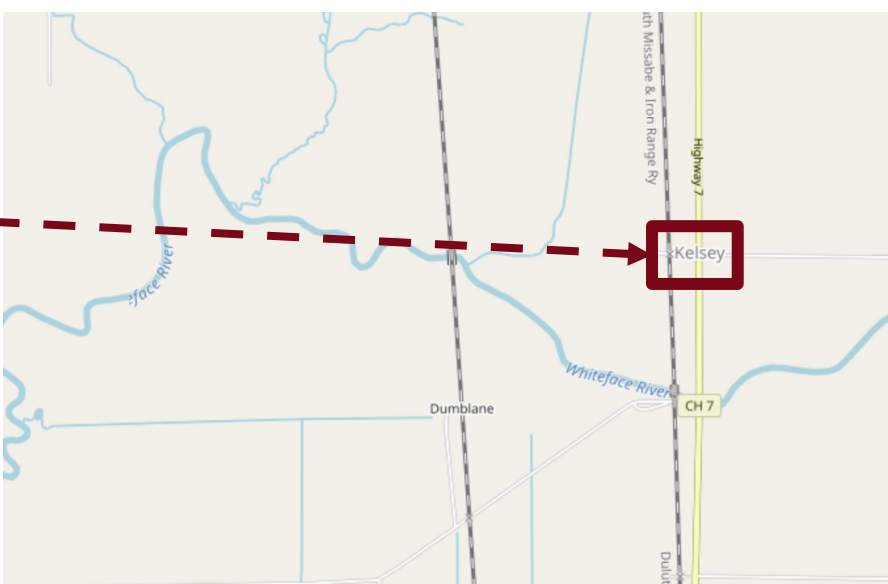**Advisor: Prof. Yao-Yi Chiang**

Department of Computer Science and Engineering, University of Minnesota, Minneapolis, USA
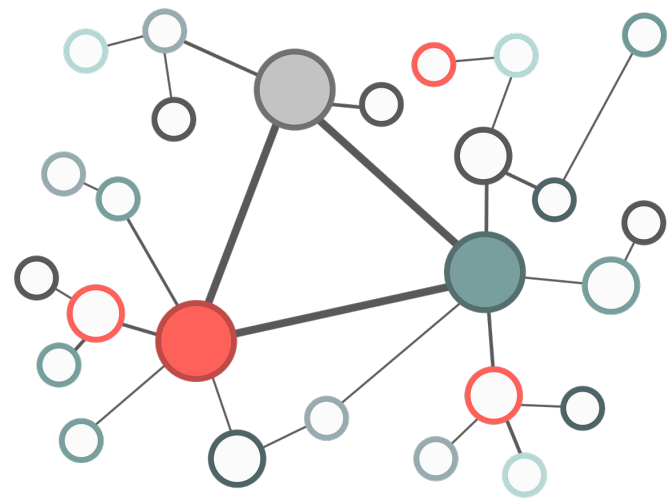kim01479@umn.edu | jina-kim.github.io

Introduction

Link text labels to entities in external knowledge bases

Map Image from David Rumsey Map Collection

OpenStreetMap

Geospatial Linked Data

Challenges to link map text and entities in external knowledge bases

(Problem #1) Placement offset
The location of a text label does not represent the actual geolocation of its corresponding geographic feature

(Problem #2) Multiple geo-entities with same name
The same name might appear multiple times within a map where each name belongs to different semantic types
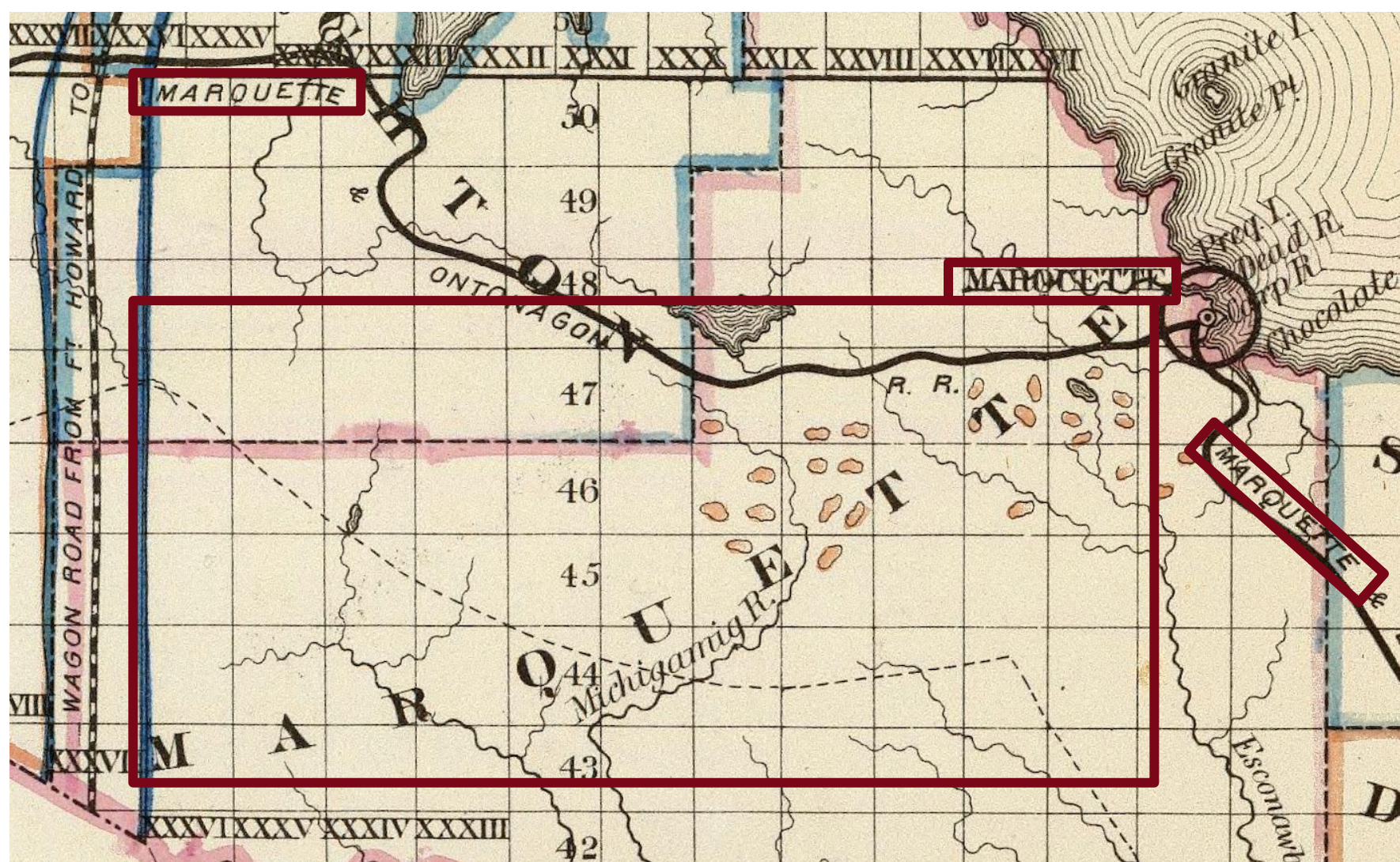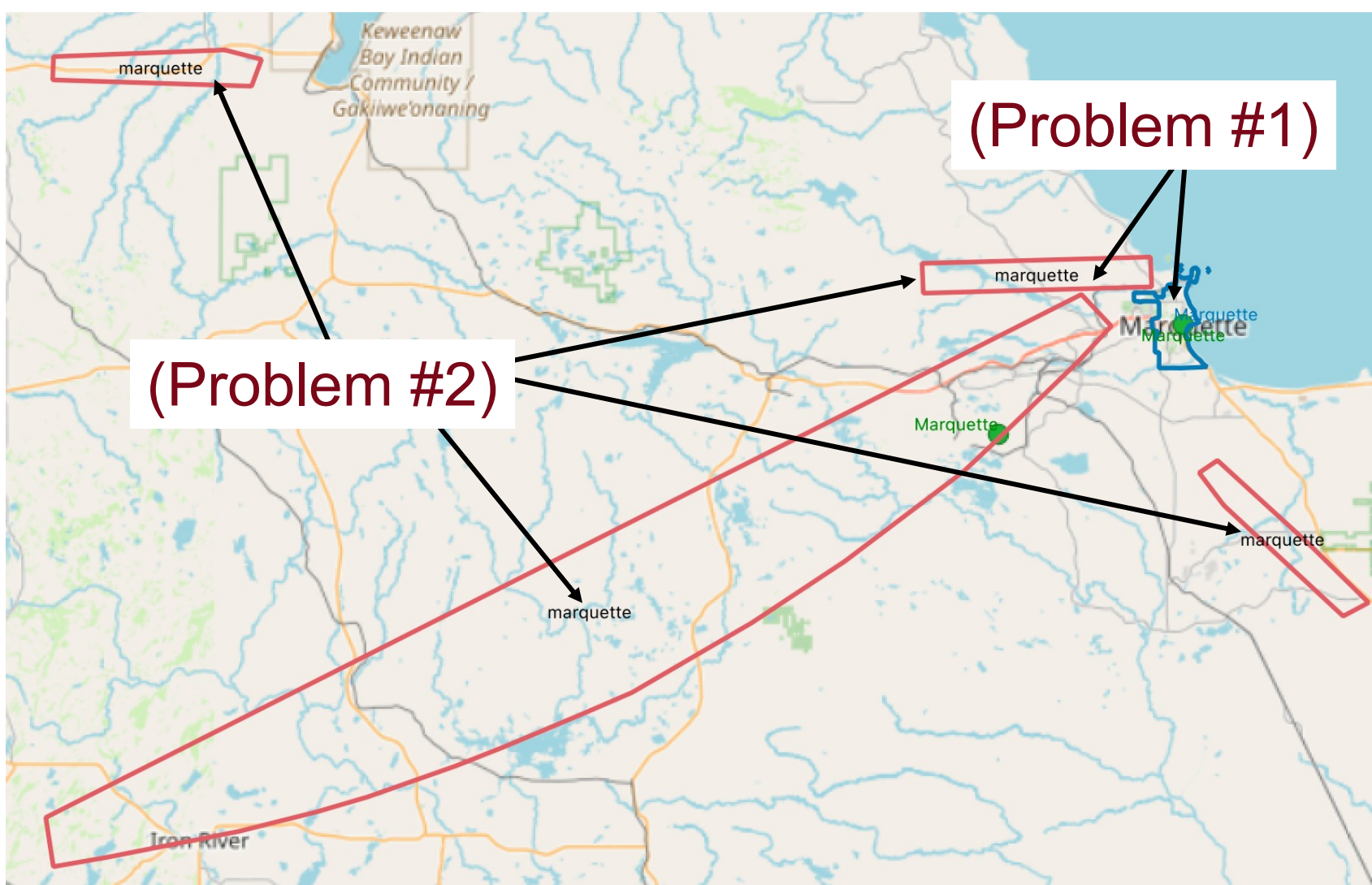

Image from David Rumsey Map Collection; duplicated text labels "Marquette" are shown in one map image


Visualization of map text (in red color) and OpenStreetMap features named "Marquette" (in green and blue color) using QGIS

RQ. How can we leverage cartographic principles of label styles and placement to automatically predict the semantic type of corresponding map text?

Proposed Approach

Synthetic data generation using OpenStreetMap and QGIS

Generated three files:
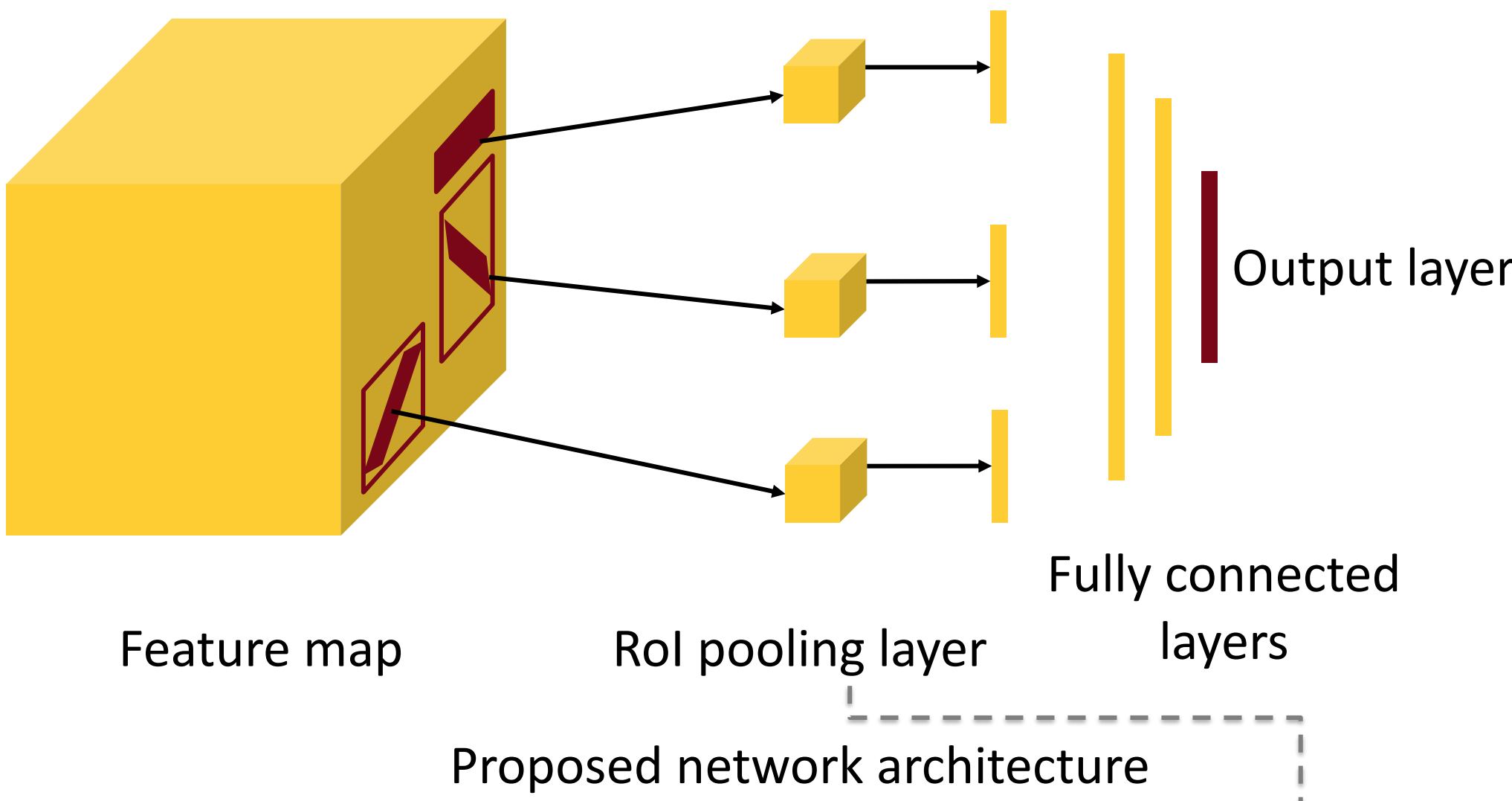- Raster map image (.png)
- Metadata (.pgw)
- Map text information (.csv)
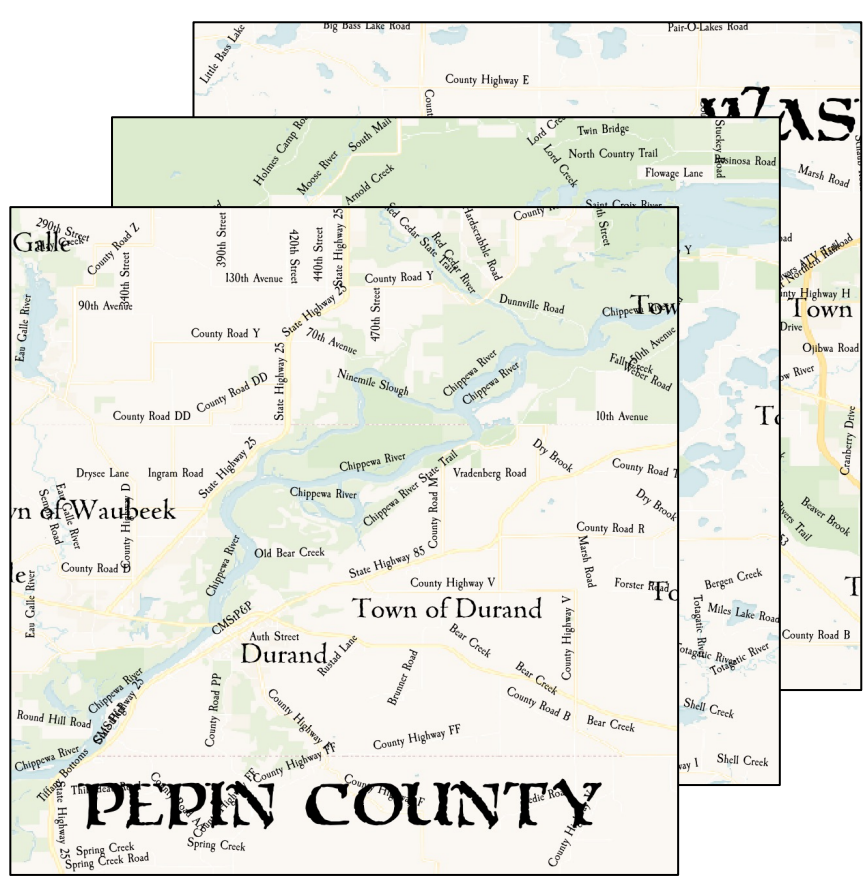
Examples of synthetic map

Semantic type classification

Predict semantic type of map text by employing pretrained text detection model (Baek et al., 2019)

Assume that intermediate layer of text detection model could capture cartographic patterns of map text (e.g., curvature, font style, font size, font spacing)

Feature map

RoI pooling layer

Fully connected layers

Output layer

Proposed network architecture

RoI pooling layer is to generate a fixed-size representation vector for each map text

Results

Synthetic Dataset	Sementic Type	# of Data	Accuracy for each class
Minnesota (zoom level 12)	City	1,099	94.7%
	County	2,236	97.8%
	Road	31,895	98.5%
	Township	1,857	93.1%
	Waterway	11,338	95.5%

Predicted semantic types of map text