Road Network Representation Learning with the Third Law of Geography

In NeurIPS 2024

H Zhou, W Huang, Y Chen, T He, G Cong, Y-S Ong
Nanyang Technological University

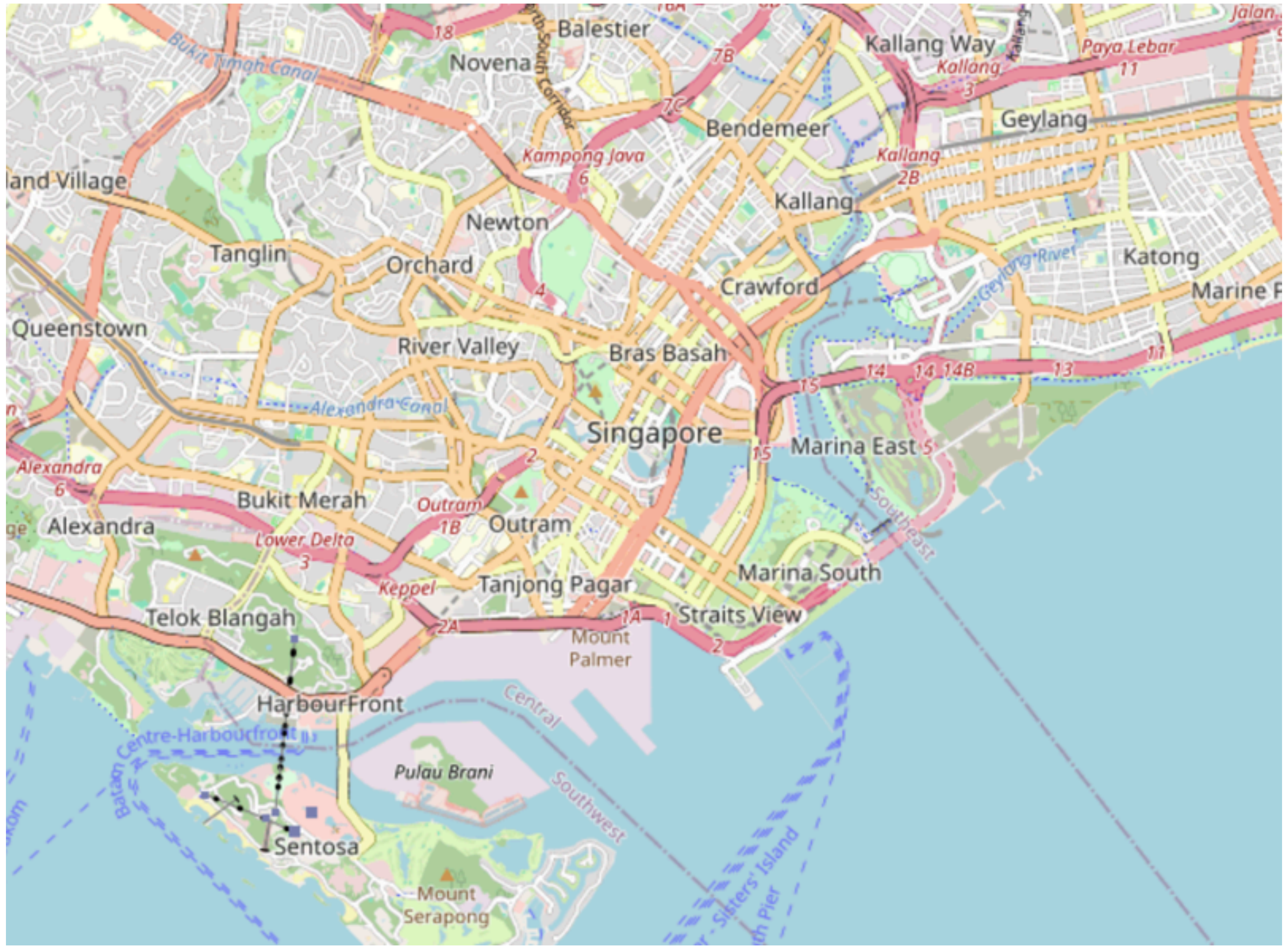
Table of Contents

Introduction

Road Network

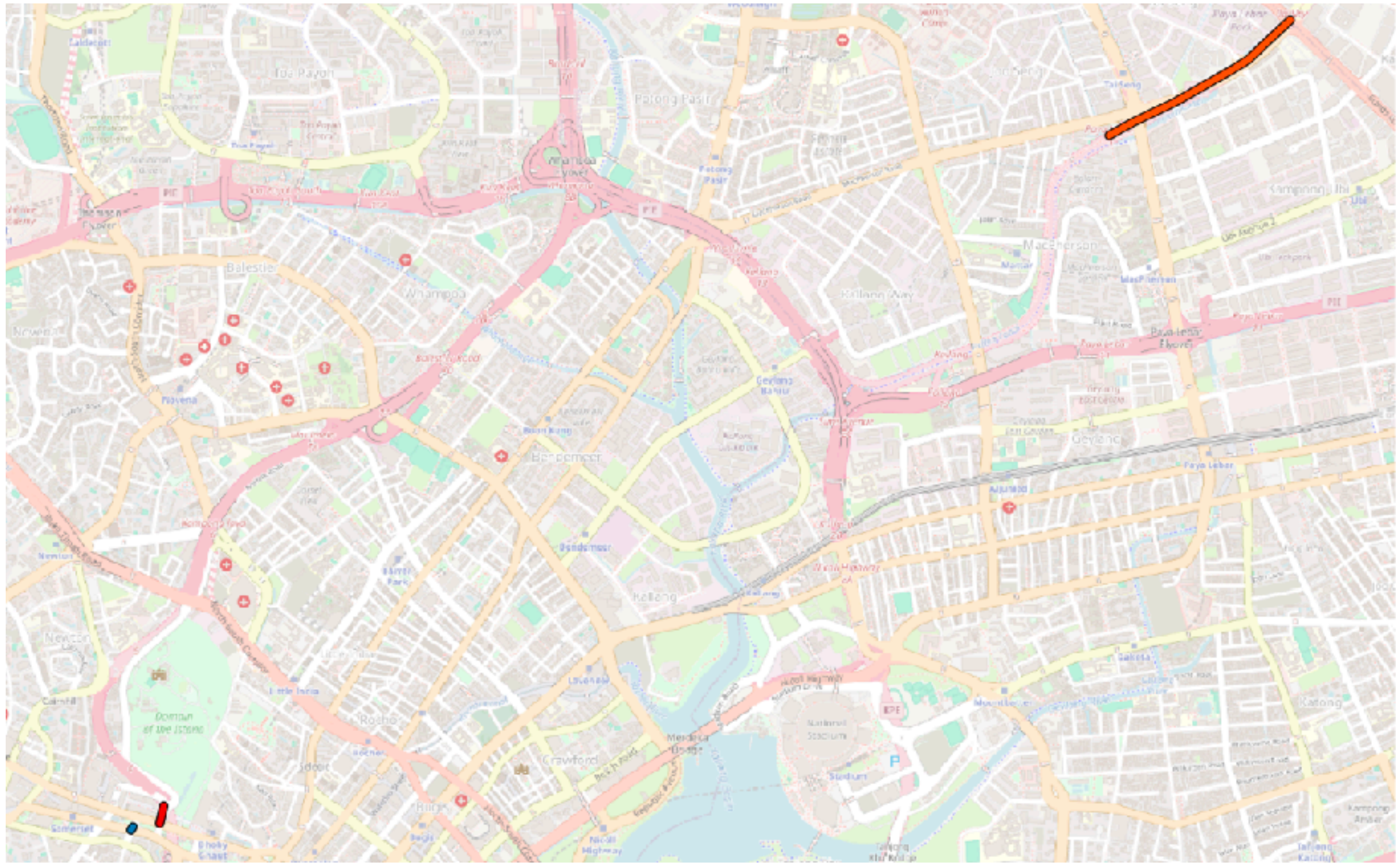
- Road network: road segments & intersections

- Road network representation learning: learn embedding for each road segment



Tobler's First Law of Geography

- Everything is related to everything else, but **near** things are more related than distant things. [1]

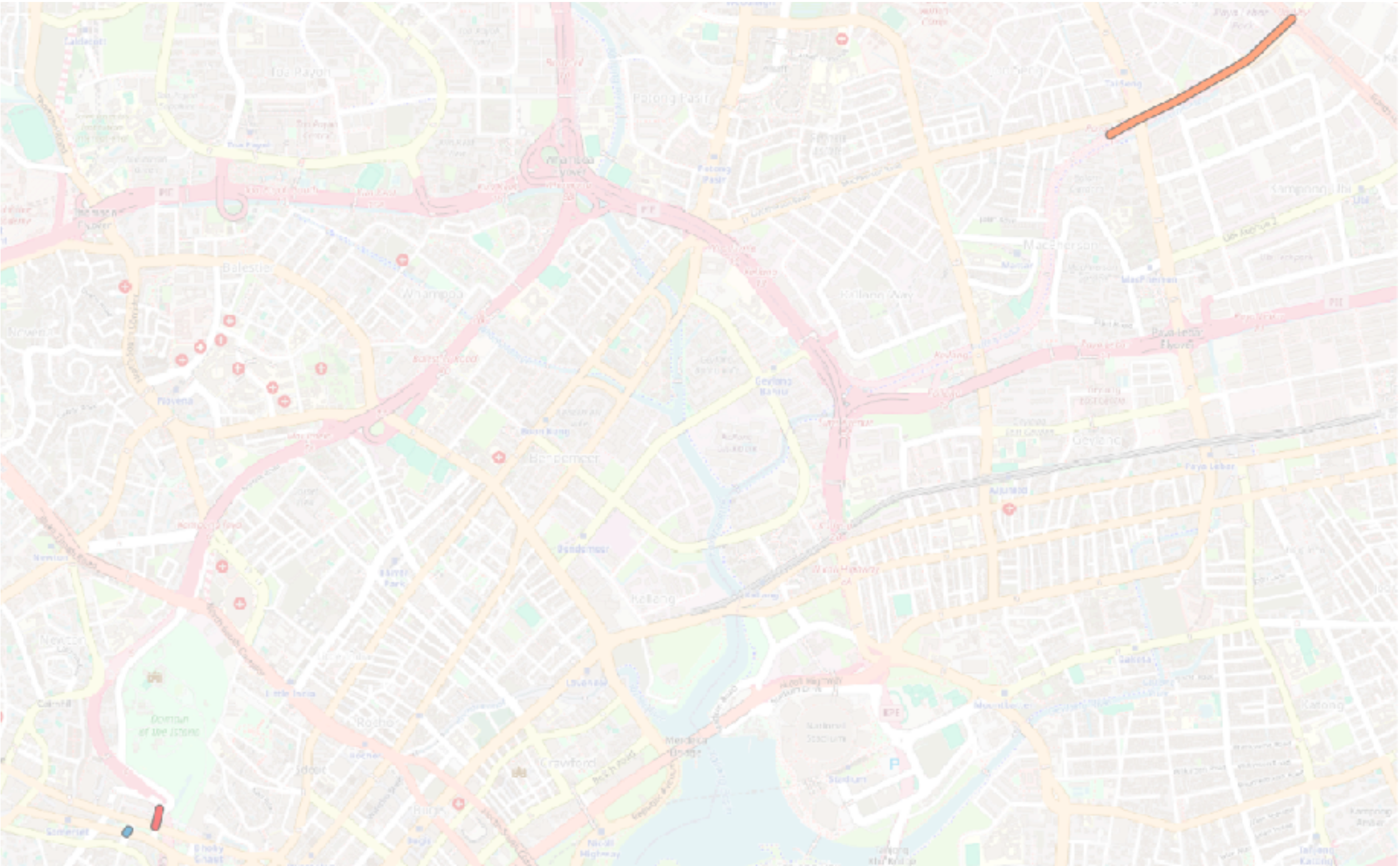


[1] Harvey J Miller. Tobler's first law and spatial analysis. Annals of the association of American geographers

Tobler's First Law of Geography

- Everything is related to everything else, but **near** things are more related than distant things. [1]

- Previous literature is primarily based on the First Law of Geography.



[1] Harvey J Miller. Tobler's first law and spatial analysis. Annals of the association of American geographers

The Third Law of Geography

- The more similar <u>geographic configurations</u> of two points (areas), the more similar the values (processes) of the target variable at these two points (areas). [1]

[1] A-Xing Zhu et al. Spatial prediction based on third law of geography. Annals of GIS, 2018

The Third Law of Geography

- The more similar geographic configurations of two points (areas), the more similar the values (processes) of the target variable at these two points (areas). [1]

- Geographic configuration refers to the description of spatial neighborhood (or context) around a point (area)

- Street view images as the proxy for geographic configuration. (We get SVI from Google Maps.)

[1] A-Xing Zhu et al. Spatial prediction based on third law of geography. Annals of GIS, 2018

The Difference between Two Laws: A case study

Research Challenges

- To effectively & efficiently model the third law.
  - Effectiveness: model the mapping from similarity relationships of geographic configurations to similarity relationships of embeddings.
  - Efficiency: the mapping is from a domain of $O(n^2)$ to another domain of $O(n^2)$.
- To fuse the two laws, which may be conflicting.

Methods

- How to model the third law:
  - Geographic configuration-aware graph augmentation + spectral negative sampling
- Fuse the two laws: dual contrastive learning objective + parameter sharing.

Contributions

- The **first** work to analyze road network representation with Geographic Laws. The **first** work to use the Third Law of Geography in road network representation learning and spatial temporal data mining.

- A novel graph contrastive learning framework to incorporate the Third Law of Geography, with non-trivial theoretical support.

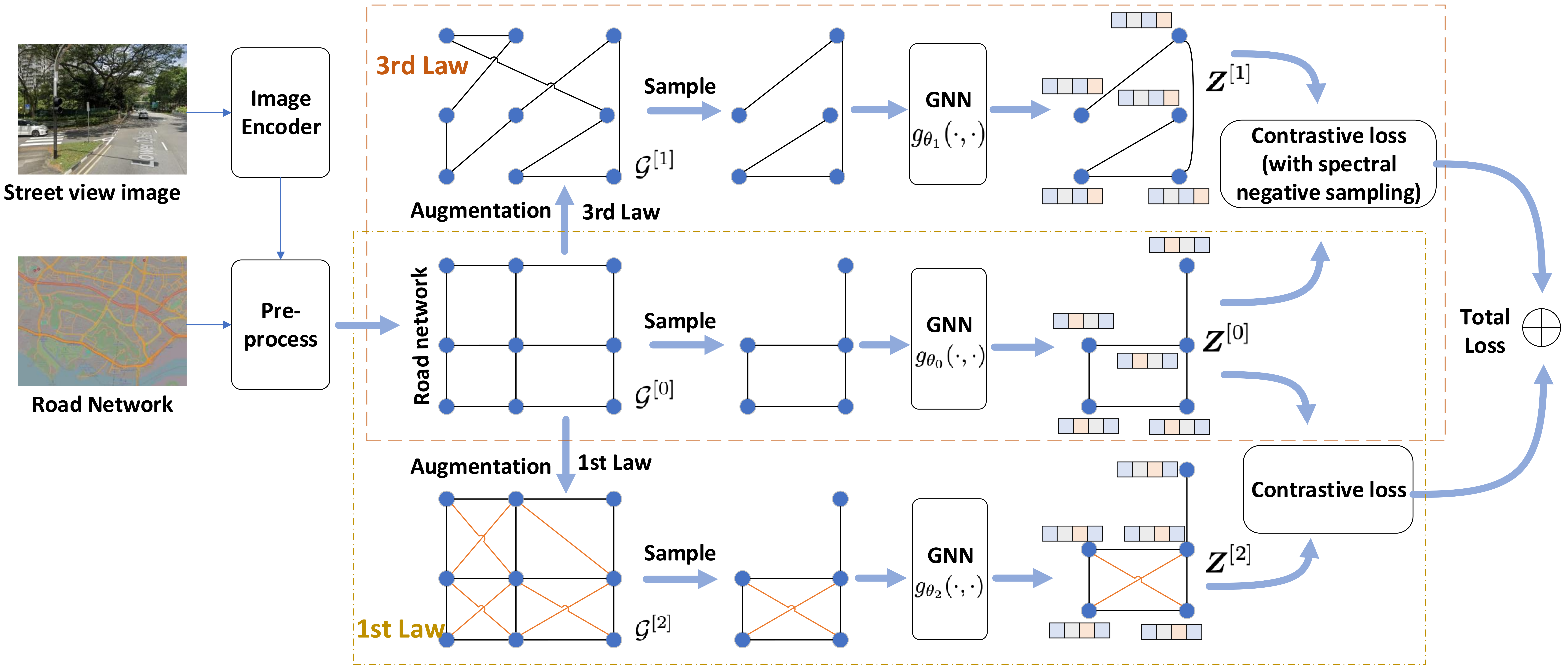
- Extensive downstream tasks to evaluate the effectiveness of the proposed method.

Method

Preliminaries

- Graph: $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \boldsymbol{X})$, (node set, edge set, feature matrix). n, m as the number of nodes and edges. $\boldsymbol{A}$ as the adjacency matrix. $\boldsymbol{D}$ is the diagonal degree matrix, where $\boldsymbol{D}_{i,i}$ is the degree of node i .
- Graph Laplacian matrix: $\boldsymbol{L} := \boldsymbol{D} - \boldsymbol{A}$
- Road network: road segments as nodes.

Model Architecture



Geographic Configuration Aware Graph Augmentation and Graph Encoder

- Goal: generate similar embedding for road segments with similar geographic configurations, with affordable computational cost.

- Build a KNN graph, according to the similarity of geographic configurations.

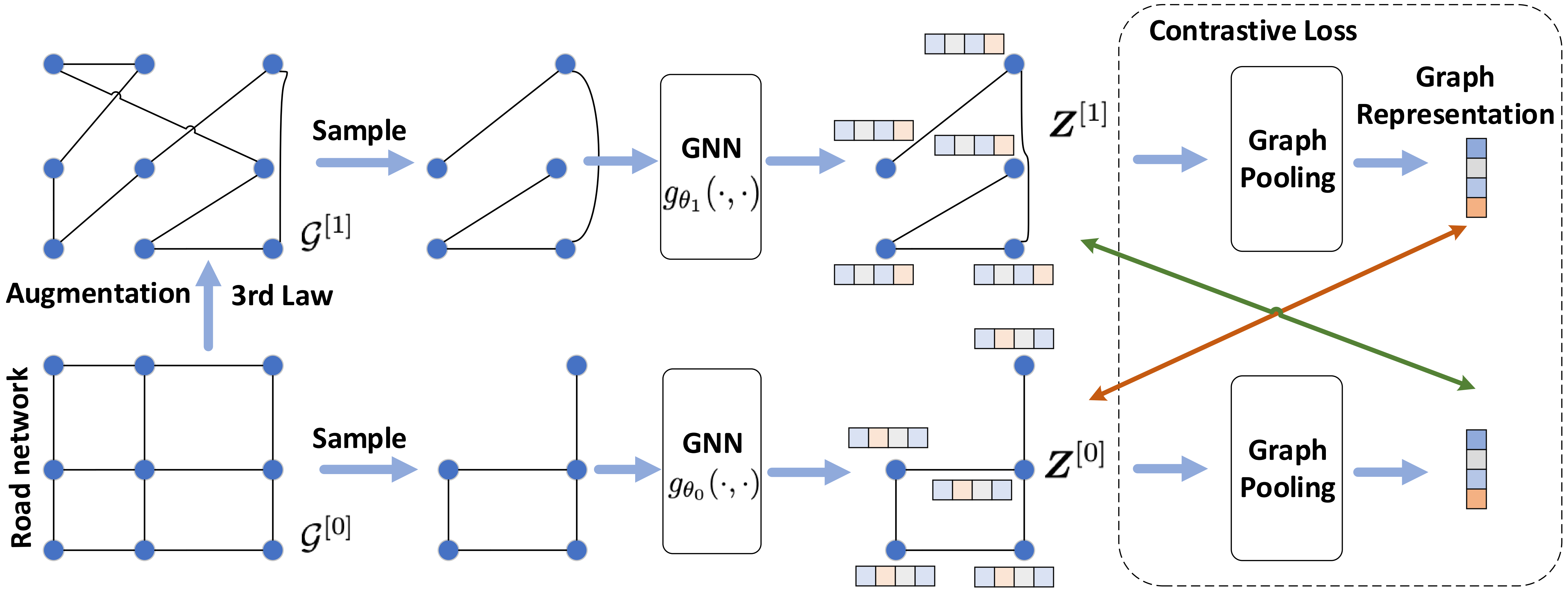
- Employ a SGC (Simple Graph Convolution) encoder on the KNN graph. This will generate similar embedding for connected nodes.

Contrastive Loss

- Maximize the mutual information between the original graph and the augmented graph. This can fuse the information between the two views.

$$\mathcal{L}_1 = -\frac{1}{|\mathcal{V}|}\sum_{i=1}^{|\mathcal{V}|}\left\{\mathrm{MI}(\boldsymbol{Z}_i^{[0]}, \boldsymbol{Z}_g^{[1]}) + \mathrm{MI}(\boldsymbol{Z}_i^{[1]}, \boldsymbol{Z}_g^{[0]})\right\}$$

$\mathrm{MI}(\cdot,\cdot)$ is the mutual information estimator based on JS divergence.



Contrastive Loss

- Maximize the mutual information between the original graph and the augmented graph. This can fuse the information between the two views.

$$\mathcal{L}_1 = -\frac{1}{|\mathcal{V}|}\sum_{i=1}^{|\mathcal{V}|}\left\{\mathrm{MI}(\boldsymbol{Z}_i^{[0]}, \boldsymbol{Z}_g^{[1]}) + \mathrm{MI}(\boldsymbol{Z}_i^{[1]}, \boldsymbol{Z}_g^{[0]})\right\}$$

$\mathrm{MI}(\cdot,\cdot)$ is the mutual information estimator based on JS divergence.

$$\mathrm{MI}(\boldsymbol{Z}_i^{[0]}, \boldsymbol{Z}_g^{[1]}) = \mathbb{E}_{(\mathcal{G}^{[0]},\mathcal{G}^{[1]})}\left[\log \mathcal{D}(\boldsymbol{Z}_i^{[0]}, \boldsymbol{Z}_g^{[1]})\right] + \mathbb{E}_{(\bar{\mathcal{G}}^{[0]},\mathcal{G}^{[1]})}\left[\log(1-\mathcal{D}(\bar{\boldsymbol{Z}}_i^{[0]}, \boldsymbol{Z}_g^{[1]}))\right]$$

$$\mathrm{MI}(\boldsymbol{Z}_i^{[1]}, \boldsymbol{Z}_g^{[0]}) = \mathbb{E}_{(\mathcal{G}^{[1]},\mathcal{G}^{[0]})}\left[\log \mathcal{D}(\boldsymbol{Z}_i^{[1]}, \boldsymbol{Z}_g^{[0]})\right] + \mathbb{E}_{(\bar{\mathcal{G}}^{[1]},\mathcal{G}^{[0]})}\left[\log(1-\mathcal{D}(\bar{\boldsymbol{Z}}_i^{[1]}, \boldsymbol{Z}_g^{[0]}))\right]$$

- Negative sample is required.

Negative Sampling

- The graph augmentation with SGC encoder achieves the Third Law.
- The negative sampling can achieve the reverse version of Third Law "roads with dissimilar geographic configurations should have dissimilar representations."
- A negative graph should have the same feature matrix as the original graph, but mainly connects road segments with dissimilar geographic configurations.
- A choice is the complete graph (fully connected graph). In the paper I derive the complete graph from the objective of the sparsest cut problem.

Negative Sampling

- The negative sampling can achieve the reverse version of Third Law "roads with dissimilar geographic configurations should have dissimilar representations."
- A choice is the complete graph.
- The complete graph has $O(n^2)$ edges, with very high computational cost in large graphs.
- Then I approximate the complete graph with a spectral graph sparsifier: a d-regular graph, where each node has d edges.

Math of Negative Sampling

- The design is inspired by the sparsest cut problem in spectral graph theory, which seeks to find cuts that minimize the number of edges between subsets of nodes. $usc_g := \min_S \frac{\mathcal{E}(S, \mathcal{V} - S)}{|S||\mathcal{V} - S|}$

- The objective can be computed as

$$usc_{\mathcal{G}} = \min_{\boldsymbol{x} \in \{0,1\}^n - \{\boldsymbol{0},\boldsymbol{1}\}} \frac{\sum_{(i,j)\in\mathcal{E}} (\boldsymbol{x}_i - \boldsymbol{x}_j)^2}{\sum_{(i,j)} (\boldsymbol{x}_i - \boldsymbol{x}_j)^2} = \min_{\boldsymbol{x} \in \{0,1\}^n - \{\boldsymbol{0},\boldsymbol{1}\}} \frac{\boldsymbol{x}^T \boldsymbol{L}_{\mathcal{G}} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{L}_{\mathcal{K}} \boldsymbol{x}}$$

- We apply a continuous relaxation in this formula and extend it to the matrix form $\min_{\boldsymbol{Z} \in \mathbb{R}^{n \times f}} \frac{\mathrm{tr}(\boldsymbol{Z}^T \boldsymbol{L}_{\boldsymbol{S}} \boldsymbol{Z})}{\mathrm{tr}(\boldsymbol{Z}^T \boldsymbol{L}_{\mathcal{K}} \boldsymbol{Z})}$

- The negative sample is based on z and complete graph $\mathcal{K}$.

Math of Negative Sampling

- The negative sample in the loss can be computed as $\bar{\boldsymbol{Z}} = g_{\theta_1}(\hat{\boldsymbol{A}}_{\mathcal{K}}, \boldsymbol{Z})$ (Details can be found in the paper.)

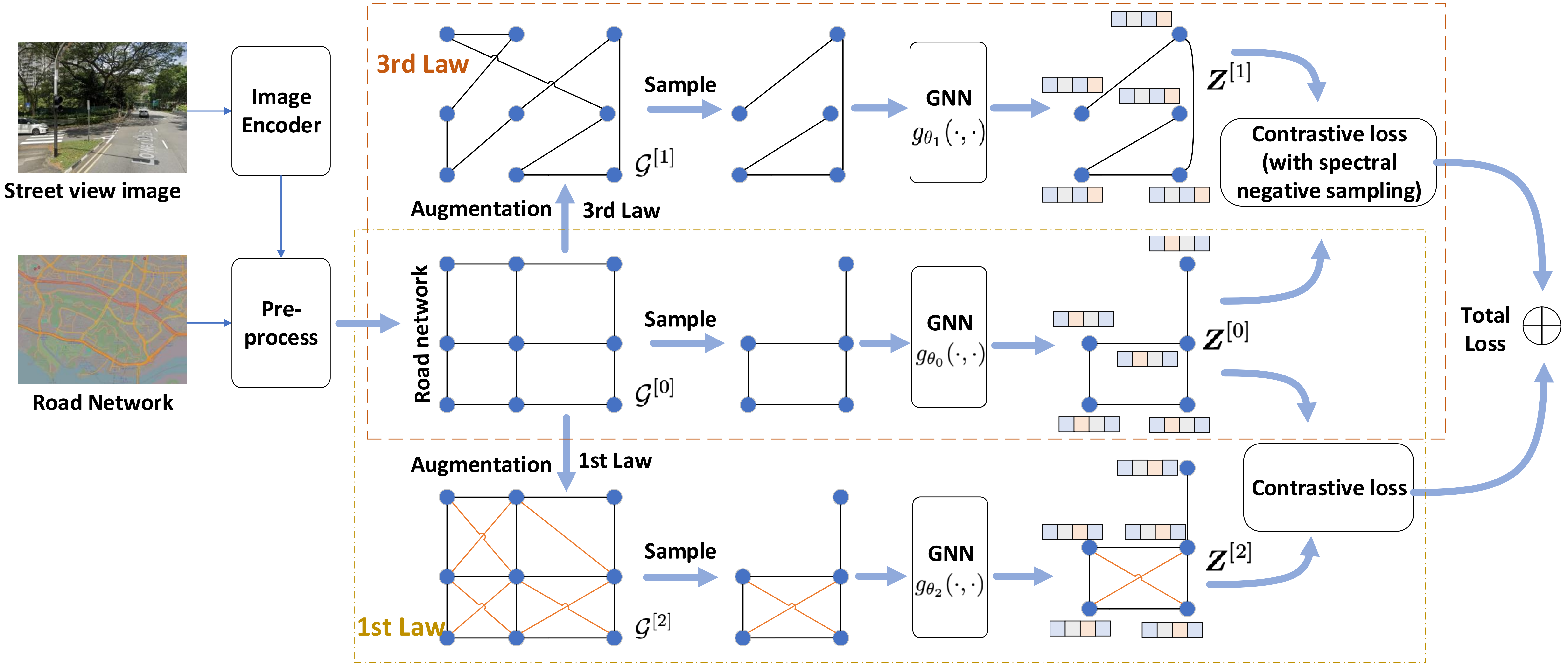
- However, the computational cost of employing a complete graph is very high ($O(n^2)$), especially in large graphs.

- Thus we derive a spectral graph sparsifier to approximate it
$$(1 - \frac{2\sqrt{d-1}}{d}) \operatorname{tr}(\boldsymbol{Z}^T \boldsymbol{L}_{\widetilde{\mathcal{K}}} \boldsymbol{Z}) \leq \operatorname{tr}(\boldsymbol{Z}^T \boldsymbol{L}_{\mathcal{K}} \boldsymbol{Z}) \leq (1 + \frac{2\sqrt{d-1}}{d}) \operatorname{tr}(\boldsymbol{Z}^T \boldsymbol{L}_{\widetilde{\mathcal{K}}} \boldsymbol{Z}),$$
where $\widetilde{\mathcal{K}}$ is a d-regular graph.

- We replace $\mathcal{K}$ with $\widetilde{\mathcal{K}}$.

Fuse the Third and the First Law

- To learn the first law, we design an graph augmentation with graph diffusion according to the road connectivity. And follow similar graph contrastive learning as the Third Law.
- We use parameter sharing between the two learning objectives.
- We add the two losses as the final learning objective.



Experiments

Experimental Setups

- Dataset: road networks (Singapore & New York) from OpenStreetMap, street view images from Google.

- Downstream Task: road function prediction, road traffic inference, and visual road retrieval.

- Hyper-parameter settings: do not tune a lot, same setting on two dataset in different tasks.

Table 1: Dataset Statistics

City	# Roads	# Edges	# SVIs
Singapore	45,243	138,843	136,399
NYC	139,320	524,565	254,239

Experimental Results - 1

Table 2: Results in Road Function Prediction, with the best in **bold** and the second best underlined

Methods	Singapore Micro-F1 (%) ↑	Singapore Macro-F1 (%) ↑	Singapore AUROC (%) ↑	NYC Micro-F1 (%) ↑	NYC Macro-F1 (%) ↑	NYC AUROC (%) ↑
Deepwalk	62.76 ± 0.49	13.30 ± 0.10	63.23 ± 0.47	78.09 ± 0.18	14.62 ± 0.02	58.49 ± 0.33
MVGRL	66.61 ± 0.50	30.67 ± 0.66	74.34 ± 0.46	78.23 ± 0.23	17.39 ± 0.23	69.96 ± 0.35
CCA-SSG	64.28 ± 0.37	22.55 ± 0.49	70.26 ± 0.37	78.20 ± 0.24	15.97 ± 0.15	68.15 ± 0.24
GGD	64.21 ± 0.39	20.58 ± 0.40	68.97 ± 0.40	78.14 ± 0.25	15.75 ± 0.16	66.11 ± 0.33
RFN	62.75 ± 0.44	12.85 ± 0.06	54.64 ± 0.44	oom	oom	oom
SRN2Vec	64.02 ± 0.45	22.47 ± 0.37	71.18 ± 0.40	oom	oom	oom
SARN	66.49 ± 0.47	22.59 ± 0.51	72.74 ± 0.50	78.14 ± 0.21	14.62 ± 0.02	68.54 ± 0.30
Garner	**81.40** ± 0.30	**62.45** ± 0.64	**93.27** ± 0.22	**82.97** ± 0.16	**47.22** ± 0.42	**89.30** ± 0.21

"oom" means out-of-memory.

Experimental Results - 2

Table 3: Results in Road Traffic Inference, with the best in **bold** and the second best underlined

Methods	Singapore MAE ↓	Singapore RMSE ↓	Singapore MAPE ↓	NYC MAE ↓	NYC RMSE ↓	NYC MAPE ↓
Deepwalk	3.43 ± 0.03	4.31 ± 0.05	0.721 ± 0.038	4.31 ± 0.03	5.92 ± 0.05	0.267 ± 0.002
MVGRL	$\underline{3.04} \pm 0.04$	$\underline{3.82} \pm 0.04$	$\underline{0.629} \pm 0.041$	$\underline{3.91} \pm 0.02$	$\underline{5.16} \pm 0.03$	$\underline{0.243} \pm 0.001$
CCA-SSG	3.31 ± 0.03	4.15 ± 0.04	0.674 ± 0.037	4.03 ± 0.03	5.34 ± 0.04	0.253 ± 0.003
GGD	3.37 ± 0.03	4.27 ± 0.04	0.684 ± 0.039	4.80 ± 0.03	6.63 ± 0.06	0.267 ± 0.002
RFN	3.54 ± 0.03	4.48 ± 0.04	0.717 ± 0.046	oom	oom	oom
SRN2Vec	3.44 ± 0.04	4.47 ± 0.05	0.569 ± 0.025	oom	oom	oom
SARN	3.40 ± 0.03	4.32 ± 0.05	0.697 ± 0.038	4.66 ± 0.04	6.39 ± 0.07	0.262 ± 0.002
`Garner`	$\mathbf{2.80} \pm 0.03$	$\mathbf{3.52} \pm 0.04$	$\mathbf{0.579} \pm 0.030$	$\mathbf{3.30} \pm 0.02$	$\mathbf{4.40} \pm 0.03$	$\mathbf{0.207} \pm 0.002$

"oom" means out-of-memory.

Case Study

- Case: top 10 most similar roads of an anchor found by the First Law only and both laws.

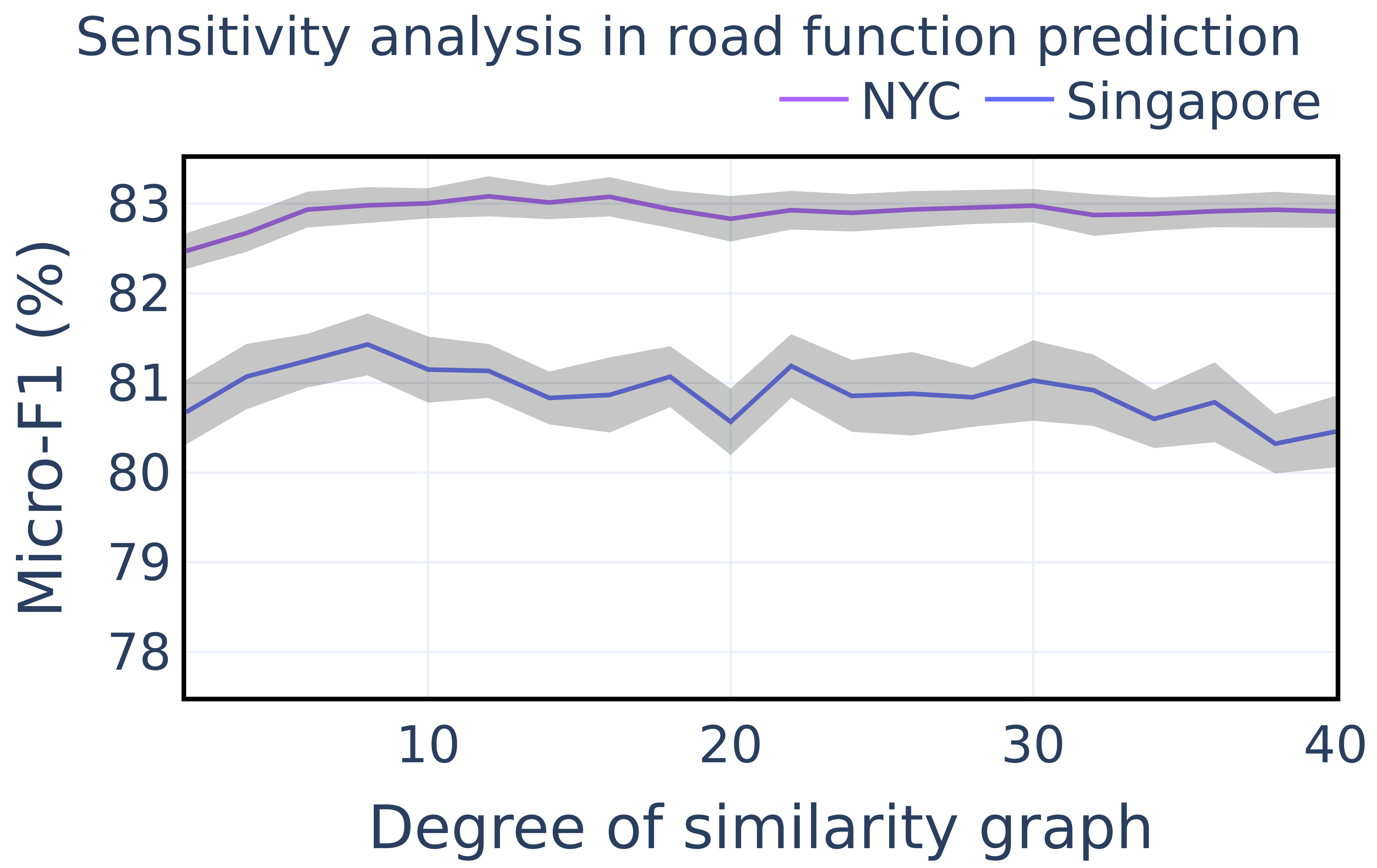
Ablation Study

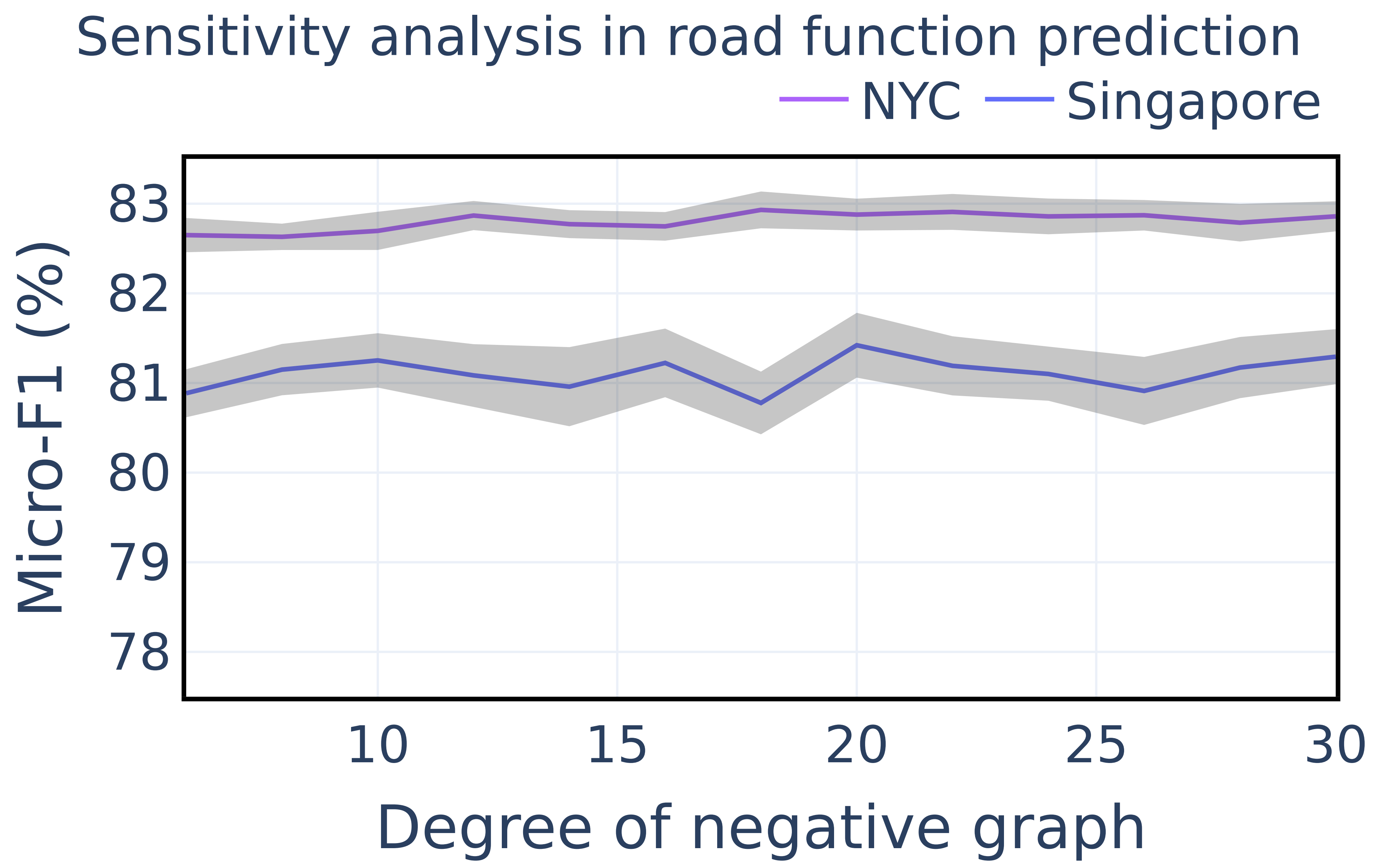
Table 5: Ablation studies on Road Function Prediction

Methods	Singapore			NYC		
	Micro-F1 (%) ↑	Macro-F1 (%) ↑	AUROC (%) ↑	Micro-F1 (%) ↑	Macro-F1 (%) ↑	AUROC (%) ↑
`Garner` - sns - aug - SVI	66.61 ± 0.50	30.67 ± 0.66	74.34 ± 0.46	78.23 ± 0.23	17.39 ± 0.23	69.96 ± 0.35
`Garner` - sns - aug	74.78 ± 0.32	50.21 ± 0.60	88.21 ± 0.30	80.64 ± 0.22	37.14 ± 0.44	85.30 ± 0.25
`Garner` - sns	80.65 ± 0.31	60.57 ± 0.68	92.46 ± 0.23	82.62 ± 0.19	45.78 ± 0.52	88.61 ± 0.18
`Garner`	81.40 ± 0.30	62.45 ± 0.64	93.27 ± 0.22	82.97 ± 0.16	47.22 ± 0.42	89.30 ± 0.21

"- sns" means to generate negative samples only with feature shuffling. "- aug" means without geographic configuration aware graph augmentation. "- SVI" means without street view images as inputs.

Parameter Sensitivity Test





- Sensitivity test on the degree of similarity graph and negative graph.

The End