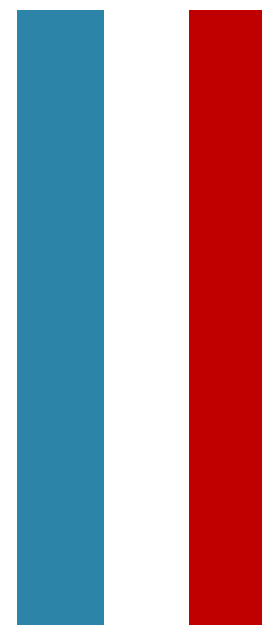


UniTR: A Unified Framework for Joint Representation Learning of Trajectories and Road Networks

发表: AAAI 2025(Oral)

代码链接: <https://github.com/csjiezhao/UniTR>



2025/10/12

I Introduction

城市时空数据表示学习的必要性

1. 可服务于多种智能应用——将时空对象转化为信息丰富的向量表示，为下游任务提供基石
2. 路网和轨迹数据本质上相互关联，它们的联合表示学习可以显著提升应用的准确性和实用性
 - 路网的拓扑结构约束了轨迹的运动，路段的多模态信息补充了轨迹的地理上下文
 - 轨迹揭示了路段间的高阶关系(**high-order relationships**), 转移频率(**transition frequencies**)凸显了不同路段的重要性

联合表示的挑战性

- **复杂交互建模**: 体现在轨迹与多个路段交互，而一个路段也被多个轨迹穿过，具有多对多关系
- **跨尺度优化**: 路网与轨迹的嵌入空间存在于不同尺度，需要同时进行同尺度和跨尺度(**both same-scale and cross-scale**)优化设计
- 路网-轨迹联合表示学习统一框架UniTR
 - 设计**分层传播机制**，建模路网与轨迹之间复杂的多对多交互，生成信息丰富的嵌入表示
 - 引入**三重对比优化模块**，系统筛选有效的正负样本，优化嵌入表示
 - 在两个真实数据集上进行实验，结果表明在**表示质量**和**下游任务性能**方面优于现有方法

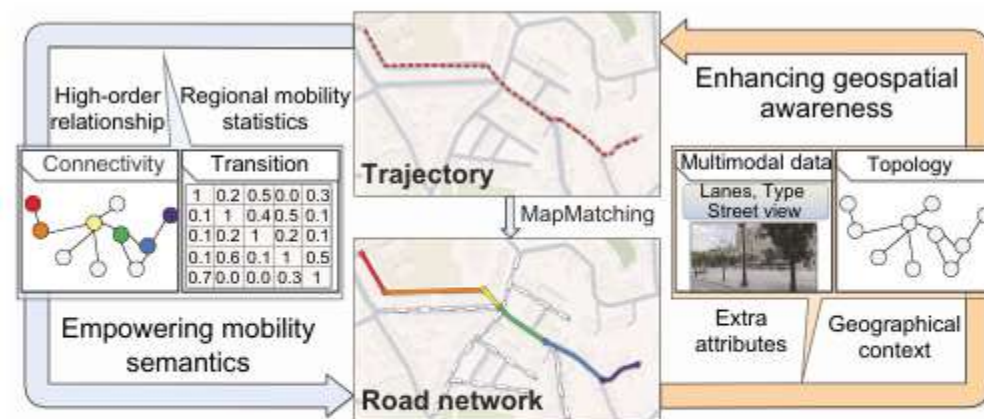


Figure 1: Illustration of mutual enhancement in joint representation of road networks and trajectories.



II Related Work



II-A Related Work

■ 联合表示学习

基于外部信息融合的方法

- **Toast**: 通过整合**辅助交通**上下文信息增强轨迹的表示学习
- **GTS**: 通过引入**兴趣点(POI)**嵌入, 为轨迹和路网提供更丰富的语义背景

评价: **间接**利用路网与轨迹的关联, 但**未能直接、显式地建模**两者之间根本、结构上的动态交互

基于自监督与对比学习的方法

- **(Fu and Lee 2020; Liang et al. 2022a; Yang et al. 2021b; Jiang et al. 2023; Mao et al. 2022; Ma et al. 2024)**: 监督信号包括**对比损失、自回归损失和掩码损失**等

现存挑战与文本动机

- 尽管取得了上述进展, 但有效地管理轨迹与道路网络之间复杂的“**多对多**”交互 (**many-to-many interactions**) 仍然是一个挑战。
- 这意味着现有方法可能在融合两种数据时**不够充分或直接**, 未能显式地建模一条轨迹与多个路段、一个路段被多条轨迹穿越的这种复杂关系。成为了UniTR框架的核心创新点

III

Method



III-A Definition

■ 定义I-路网与轨迹

1. 道路网络

- 形式化定义：有向图 $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$
- 顶点 \mathcal{V} ：代表路段，包含属性（长度、坐标）
- 边 \mathcal{E} ：代表路段间连接关系
- 邻接矩阵 (A) ：描述图结构

2. 轨迹

- 原始形式：GPS点 $\tau = \langle p_1, p_2, \dots, p_{|\tau|} \rangle$
- 路网匹配形式：通过地图匹配算法，将点序列映射为路段序列 $\langle v_1, v_2, \dots, v_n \rangle$

目标：

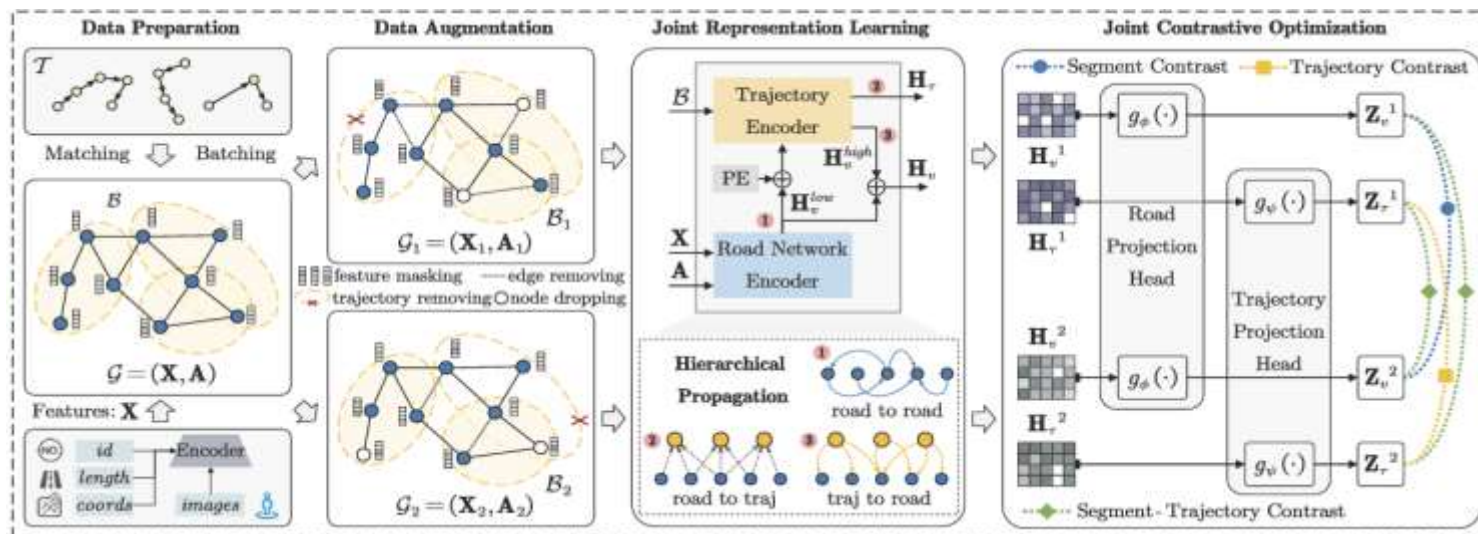
- 两个映射函数
 - 路段嵌入 $\mathcal{F}_v: \mathcal{V} \rightarrow R^d$
 - 轨迹嵌入 $\mathcal{F}_\tau: \mathcal{V} \rightarrow R^d$
- 路网图结构、路段属性、街景图像和轨迹序列作为输入，通过联合学习，将路段和轨迹学习到一个统一的低维向量表示

■ 定义II-多模态数据

1. 街景图像

- 定义：在路网特定位置拍摄的、展示周围环境的详细视觉快照。
- 数据处理：在每个路段上采集多张图像，拼接生成360度全景视图 $v.img$ 。

III-B Data Preparation and Augmentation



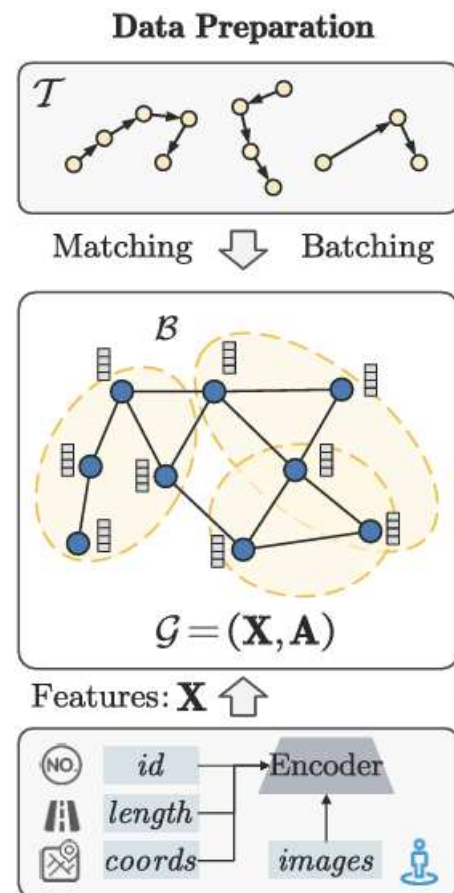
■ 数据准备阶段:

■ 路段特征提取

- ✓ 将每个道路段采集的街景图输入Swin Transformer得到高级视觉特征向量，与id、length、coords进行拼接，形成道路段的多模态特征矩阵 X

■ 轨迹预处理

- ✓ 使用地图匹配算法将原始GPS轨迹映射为路段序列。
- ✓ 组织成批次 B



III-B Data Preparation and Augmentation

■ 为何要数据增强？ ■ 核心思想：主动在训练阶段引入可控的扰动，构建困难样本

■ 路网增强-多模态特征掩码

- 每个路段包含两种特征模态：

1. 属性特征 x_i^{attr}
2. 视觉特征 x_i^{vis}

- 随机选择一种模态，特征向量置为零向量

- 表达： $m_i^{attr} \odot x_i^{attr} \parallel m_i^{vis} \odot x_i^{vis}$

✓ 模拟现实中路网信息部分缺失的场景

✓ 迫使模型不能依赖单一信息源

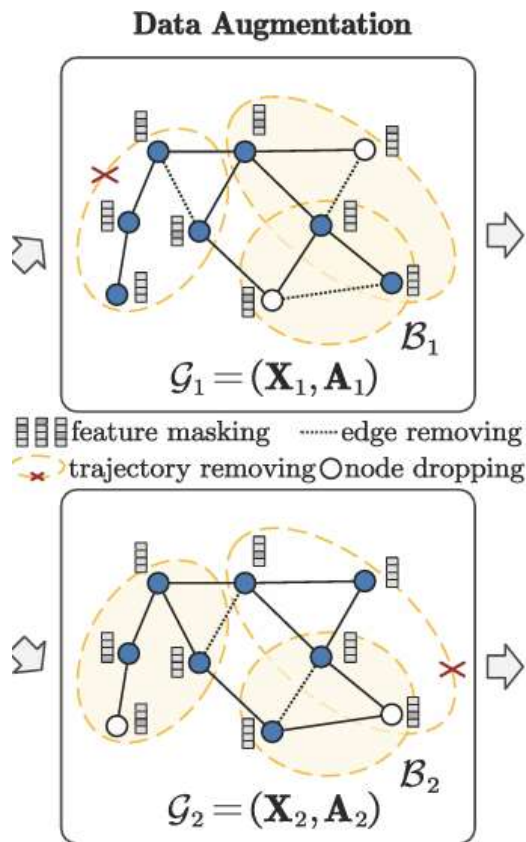
■ 轨迹增强-路段丢弃&轨迹丢弃

- 路段丢弃：随机选择轨迹中的部分路段，从序列中移除

- 轨迹丢弃：随机将整条轨迹移除

✓ 防止轨迹序列出现过拟合

✓ 保证每次迭代训练数据分布不同



■ 路网增强-基于移动性的边删除

- 计算转移概率：

根据历史轨迹数据，计算从路段 v_i 到 v_j 的概率

$$p(v_j | v_i) = \frac{\#trans(v_i \rightarrow v_j)}{\#visit(v_i)}$$

$\#trans(v_i \rightarrow v_j)$: 数据中从 v_i 直接到 v_j 的次数

$\#visit(v_i)$: 所有轨迹经过 v_i 的总次数

- 定义删除概率：边的删除概率与转移概率成反比

$$pr(e_{i,j}) = \delta_\epsilon (1 - p(v_j | v_i))$$

✓ 模拟道路的动态变化（交通管制、事故）

✓ 优先删除不重要的链接，保留主干道

III-B Joint Representation Learning

■ 路网编码器-捕获低阶结构信息

■ 采用ChebGCN

- ChebGCN使用切比雪夫多项式来近似图卷积操作，效果更好，计算更稳定

- 能够明确控制感受野大小(K-hop)

● $H_v^{low} = \sum_{k=0}^{K-1} T_k(\hat{L}) X \Theta_k$

- ✓ 学习每个路段的低阶结构表示，反映在路网中的局部拓扑位置
- ✓ 每个路段的表示，是其自身及其K-hop邻域内所有路段特征的加权求和，最终得到所有路段的低阶结构表示

■ 轨迹编码器

■ 路段->轨迹传播(Road-to-Traj Propagation)

目标：通过道路到轨迹的传播机制，聚合连续路段的表示以学习轨迹表示

1. 对于轨迹中每个路段 τ_j ，获取低阶路段表示 $\mathbf{h}_{v_i}^{low}$
2. 引入位置编码区分路段序列： $PE(i)$
3. 聚合轨迹表示：将轨迹中所有路段表示进行聚合

$$h_{\tau_j} = \sigma \left(\sum_{v_i \in \tau_j} \frac{(h_{v_i}^{low} + PE(i)) \Theta_{\tau_j}}{|\tau_j|} + b_{\tau_j} \right)$$

得到轨迹整体表示向量，蕴含轨迹的移动路径语义

■ 轨迹->路段传播(Traj-to-Road Propagation)

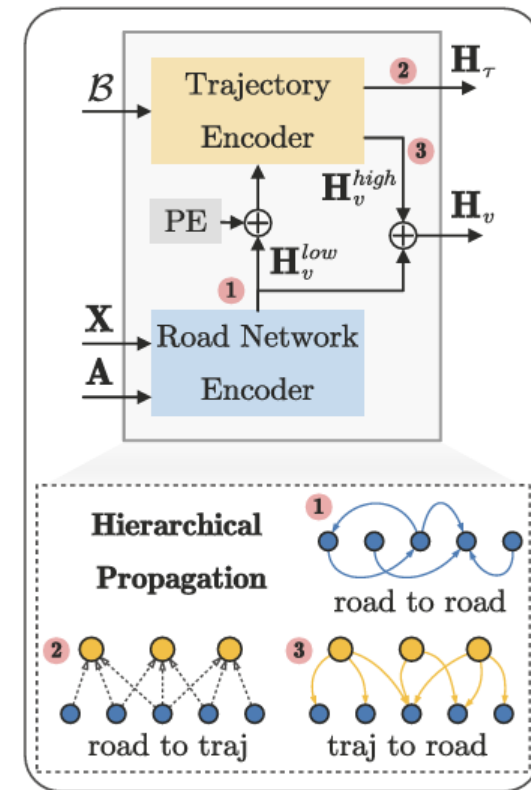
目标：利用轨迹中挖掘的高阶语义信息，对路段表示进行补充和丰富

1. 学习到的轨迹表示 \mathbf{h}_{τ_j} 通过一个专门的卷积层反馈给对应的道路段，得到路段的高阶表示：

$$h_{v_i}^{high} = \sigma \left(\sum_{\tau_j: v_i \in \tau_j} \frac{h_{\tau_j} \Theta_{v_i}}{\text{degree}(v_i)} + b_{v_i} \right)$$

最后，将低阶表示和高阶表示结合得到道路段表示 $\mathbf{H}_v^{low} + \mathbf{H}_v^{high}$

Joint Representation Learning



✓ 实现多对多建模：

- 一个轨迹通过聚合多个路段的信息形成表示（多对一）
- 一个路段通过汇聚多条轨迹的语义增强表示（多对一）
- 整体上实现了路网与轨迹集合之间的多对多交互建模

III-A Datasets & Evaluation

■ 对比学习核心思想？

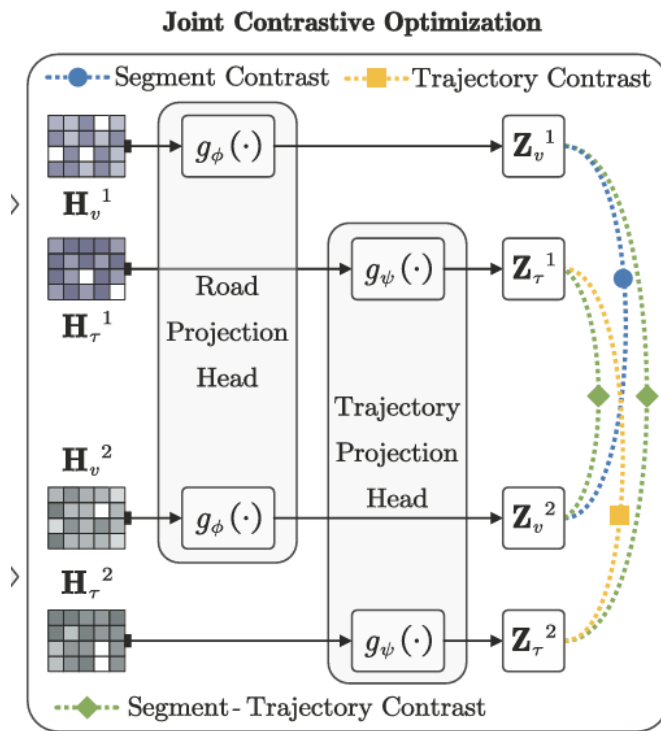
- 目标：学习一个嵌入空间，在此空间中
 - 相似的样本（正样本）的表示彼此靠近
 - 不相似的样本（负样本）的表示彼此远离
- 在UniTR中的挑战：如何定义路段和轨迹的“相似”？如何实现跨尺度的对比？

■ 预处理：投影头

- 在计算对比损失前，将编码器输出的表示通过一个简单多层感知机，即投影头 $g_v(\cdot)$ 和 $g_\tau(\cdot)$
- 目标：1. 将表示映射到更适合计算对比损失的空间； 2. 防止对比损失直接破坏编码器学到的信息（一种正则化手段）
- 然而，由于有效正负样本挖掘和嵌入空间不对齐问题，联合优化这些投影仍然具有挑战性。

■ 基于邻近性和移动性的采样策略

- ① 对于每个路段，从两个来源识别**正样本**：1. 来自特定范围内的**q-hop**邻近道路段(空间上接近->功能上会尽可能相似)； 2. 来自同一道路段在另一个视图中的表示(同一实体的不同扰动版本)
- ② 路段**负样本**：1. 不在路段的**q-hop**领域内； 2. 未在与路段同一批次的任何轨迹中共同出现的路段（现实中无车辆连续经过他们，说明移动模式上无关）
- ③ 轨迹**正样本**：1. 与轨迹Jaccard相似度超过某个阈值 θ 的轨迹； 2. 另一个视图下的同一条轨迹
- ④ 轨迹**负样本**：1. Jaccard相似度为0的轨迹(无共同路段的轨迹)



Jaccard相似度计算的是两条轨迹经过的路段集合的交并比。高的重叠意味着相似的轨迹

III-A Datasets & Evaluation

■ 三级对比损失函数

■ 对比损失通用公式(InfoNCE Loss)

- 思想：对于一个样本，让其与正样本的相似度远大于与负样本的相似度
- 相似度计算： $s(\mathbf{z}_a, \mathbf{z}_b) = e^{c o s(\mathbf{z}_a, \mathbf{z}_b) / x}$

1. 路段级对比损失 \mathcal{L}_v

1. 目标：使拓扑相近的路段在嵌入空间中也相近
2. 公式： $\mathcal{L}_v = \frac{1}{2|V|} \sum_{i=1}^{|V|} l_v(\mathbf{z}_{v,i}^1) + l_v(\mathbf{z}_{v,i}^2)$
3. 理解：拉近路段与正样本的距离，推远与所有负样本的距离。

2. 轨迹级对比损失 \mathcal{L}_τ

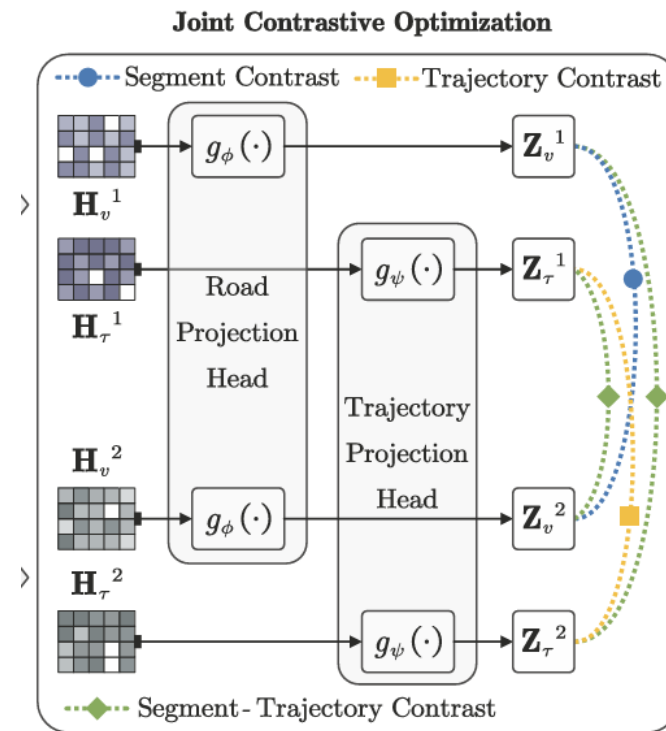
1. 目标：使路径相似的轨迹在嵌入空间中也相近
2. 公式： $\mathcal{L}_\tau = \frac{1}{2|B|} \sum_{i=1}^{|B|} l_\tau(\mathbf{z}_{\tau,i}^1) + l_\tau(\mathbf{z}_{\tau,i}^2)$

3. 跨级别对比损失 \mathcal{L}_c

1. 目标：对齐路段和轨迹两个不同尺度的表示空间
2. 方法：使用一个判别器 \mathcal{D} (双线性层+Sigmoid)计算一个路段-轨迹对是正样本的概率

$$\mathcal{L}_c = \frac{1}{2|N|} \sum_{i=1}^{|V|} \sum_{j=1}^{|B|} o l_{v_i \in \tau_j} (l_c(\mathbf{z}_{v,i}^1, \mathbf{z}_{\tau,j}^2) + l_c(\mathbf{z}_{v,i}^2, \mathbf{z}_{\tau,j}^1))$$

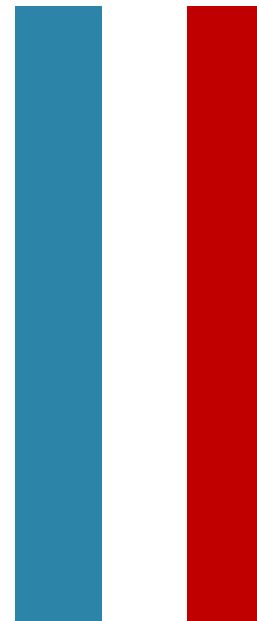
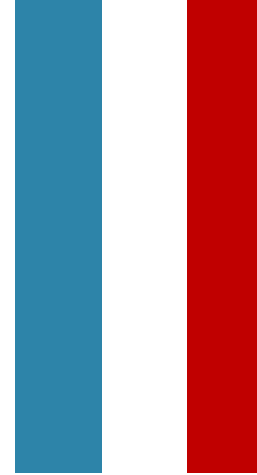
理解：双向损失。既要求从路段能找出所属的轨迹，也要求从轨迹找出其包含的路段



总体目标函数

$$\mathcal{L} = \lambda_v \mathcal{L}_v + \lambda_\tau \mathcal{L}_\tau + \lambda_c \mathcal{L}_c$$

III Experiments



III Experiments

■ 数据集

- **Porto:** 欧洲城市路网数据集
- **Chengdu:** 成都路网数据集

数据构成:

1. 道路网络: OpenStreetMap
2. **GPS**轨迹: 出租车行驶记录
3. 街景图像: Google/Baidu Map API获取

Datasets	Porto	Chengdu
# road segments	10,780	6,786
# road connections	24,980	17,542
# trajectories	1,710,670	5,819,383
# street view images	10,780	6,786

Table 1: Statistics of the datasets.

■ 四类下游任务

- ◆ 道路类型分类(Road Type Classification)
目标: 预测路段类别(主干道、辅路)
指标: Micro-F1和Macro-F1
- ◆ 交通速度推断(Traffic Speed Inference)
目标: 预测路段平均速度
指标: 平均绝对误差MAE、均方根误差RMSE
- ◆ 行程时间估计(Travel Time Estimation)
目标: 给定轨迹, 预测总行程时间
指标: 平均绝对误差MAE、均方根误差RMSE
- ◆ 轨迹相似性搜索(Trajectory Similarity Search)
目标: 给定一条查询轨迹, 找出最相似Top-k轨迹
指标: 命中率HR@10、平均排名Mean Rank

III Experiments

■ 基线模型对比

道路网络: OpenStreetMap

1. 随机游走类: Node2vec, SRN2Vec
2. GNN类: RFN、HRNR
3. 图对比学习类: SARN
4. 联合表示学习类: JCLRNT

■ 道路类型分类与交通速度推断优势

- ✓ 多模态融合中视觉特征对道路类型分类有帮助
- ✓ 路段的表示融入了动态移动模式, 对于交通速度推断至关重要

■ 行程时间估计与轨迹相似性搜索优势

- ✓ UniTR分层传播机制比JCLRNT代表的简单融合方式更有效

■ 跨数据集优势

- 在Chengdu (轨迹数: 580w, 路网规模: 6786个路段) 数据集表现远高于Porto (轨迹数: 170w, 路网规模: 10780个路段)
- ✓ 高质量、高密度的轨迹数据对提升表示学习性能更重要

Task	Road Type Classification		Traffic Speed Inference		Travel Time Estimation		Similar Trajectory Search	
Dataset	Porto	Chengdu	Porto	Chengdu	Porto	Chengdu	Porto	Chengdu
Metric	Mi-F1/Ma-F1	Mi-F1/Ma-F1	MAE/RMSE	MAE/RMSE	MAE/RMSE	MAE/RMSE	MR/HR@10	MR/HR@10
Node2Vec	0.430/0.377	0.467/0.455	3.69/4.89	7.30/8.28	202.4/293.1	137.2/203.2	25.265/0.712	43.062/0.659
SRN2Vec	0.511/0.286	0.442/0.391	2.45/3.23	3.42/4.57	202.0/294.3	226.5/312.2	11.822/0.914	14.410/0.861
RFN	0.501/0.228	0.431/0.327	3.25/4.55	6.79/7.78	163.0/256.7	162.7/235.4	31.654/0.724	54.181/0.639
HRNR	0.409/0.296	0.410/0.370	3.56/4.80	7.24/8.22	205.8/295.7	144.8/212.5	24.743/0.716	46.780/0.653
SARN	0.510/0.305	0.507/0.404	3.50/4.76	7.12/8.11	167.5/263.7	162.1/234.0	28.356/0.703	52.356/0.638
JCLRNT	0.593/0.402	0.634/0.620	2.39/3.29	3.76/4.83	159.3/251.3	120.6/180.8	5.003/0.939	7.457/0.912
UniTR	0.638/0.564	0.668/0.664	2.10/2.91	3.20/4.26	144.6/209.9	110.5/162.7	1.740/0.984	3.002/0.958

Table 2: Performance comparison on downstream tasks.

III Ablation Study

■ 消融实验分析

1. w/o vis: 移除街景视觉特征

- 受影响最大: **道路类型分类**
- ✓ 原因: 街景图像提供路宽、周围环境信息帮助远超长度等属性信息

2. w/o t2r: 移除轨迹到路段的反馈传播

- 受影响最大: **交通速度推断、轨迹相似性搜索**
- ✓ 原因: t2r连接静态结构与动态(交通速度)语义(相似性)

3. w/o \mathcal{L}_τ : 移除轨迹级对比损失

- 受影响最大: **(轨迹相关任务)旅行时间估计、相似搜索**
- ✓ 原因: 专门用于优化轨迹本身的表示空间, 确保路径相似的轨迹聚集在一起。移除导致轨迹嵌入质量下降。

4. w/o \mathcal{L}_c : 移除跨级别对比损失

- 均出现中等程度的下降
- ✓ 证明了是确保路网与轨迹表示空间对齐、有效联合学习的必要条件

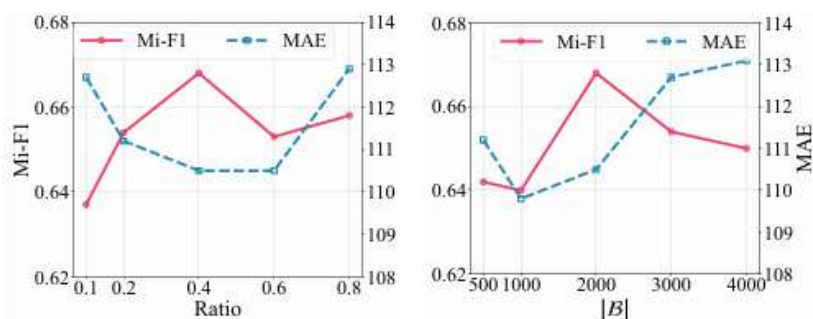
Task	Road Type Classification		Traffic Speed Inference		Travel Time Estimation		Similar Trajectory Search	
Metric	Mi-F1	Ma-F1	MAE	RMSE	MAE	RMSE	MR	HR@10
UniTR	0.638	0.564	2.101	2.909	144.6	209.9	1.740	0.984
w/o vis	0.613 ↓	0.497 ↓	1.978 ↑	3.026 ↓	142.7 ↑	211.9 ↑	2.049 ↓	0.946 ↓
w/o t2r	0.600 ↓	0.472 ↓	2.144 ↓	2.973 ↓	145.9 ↑	224.1 ↓	1.964 ↓	0.976 ↓
w/o \mathcal{L}_τ	0.613 ↓	0.505 ↓	2.221 ↓	3.108 ↓	146.6 ↑	214.1 ↓	1.742 ↓	0.983 ↓
w/o \mathcal{L}_c	0.633 ↓	0.532 ↓	2.177 ↓	3.046 ↓	142.4 ↑	209.3 ↑	1.756 ↓	0.983 ↓

Table 3: Ablation study on four downstream tasks in Porto dataset. ↓ denotes a decrease in performance, and ↑ an improvement.

III Parameter Analysis & Case Study

■ 参数分析

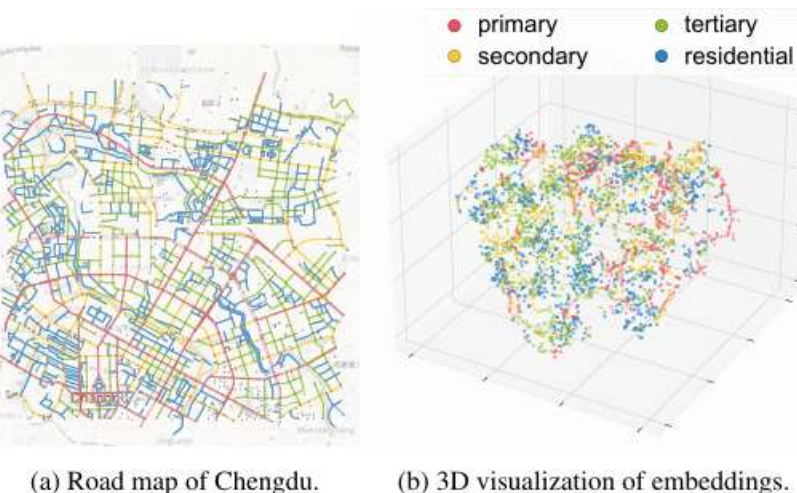
路段丢弃比率Dropping Ratio和轨迹批次大小Batch Size



1. Dropping Ratio在0.4左右，道路分类和行程时间估计性能最佳，说明适度丢弃路段有助于学习鲁棒表示；丢弃更多会破坏关键信息，预测能力下降
2. 中等规模的Batch Size（2000条）可在学习效率和性能间取得平衡；过小导致欠拟合，过大会引入噪声

■ 嵌入可视化

- 将路段嵌入降维可视化，相同类型的路段清晰的聚在一起
- ✓ 主干道、次干道、住宅道路和第三级道路清晰分离



(a) Road map of Chengdu.

(b) 3D visualization of embeddings.

Figure 4: Road types and embedding visualization

■ 轨迹相似性搜索

- 展示查询轨迹和模型检索的Top-3相似轨迹，结果显示极高的空间重叠度

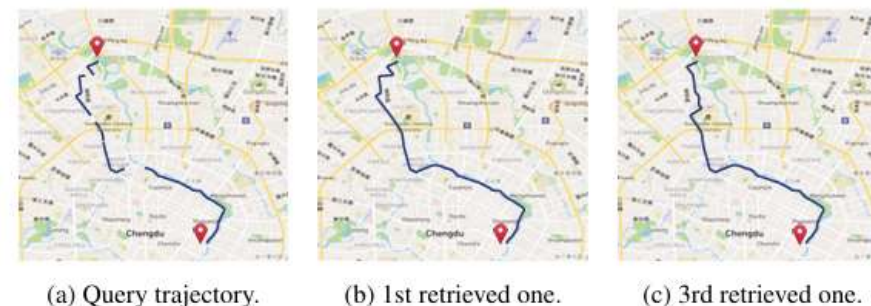


Figure 5: Illustration of trajectory search.

IV Conclusion



IV Conclusion

■ 解决的核心问题

- 如何有效对道路网络和轨迹两种内在关联的时空数据进行联合表示学习
- 现有方法的不足：大多孤立学习，或未能解决多对多交互和跨尺度优化问题

■ 核心贡献

- 提出路网-轨迹分层传播机制，实现道路->轨迹，轨迹->道路的双向信息流动
- 设计路网-轨迹样本选择策略，确保正负样本在空间和语义上的相关性
- 三级对比损失函数，确保判别性和对齐性

■ 未来工作

下游任务的扩展

- 新的下游任务：交通预测、路线规划、功能区识别、异常检测

跨城市泛化能力

- 跨城市训练模型的迁移能力、通用型城市时空大模型

谢谢！

