



# 面向多源异构空间大数据的城市用地 识别与监测：一种多模态深度学习框架

答辩人：王成龙

指导老师：焦利民教授、宫兆亚助理教授



# 目 录

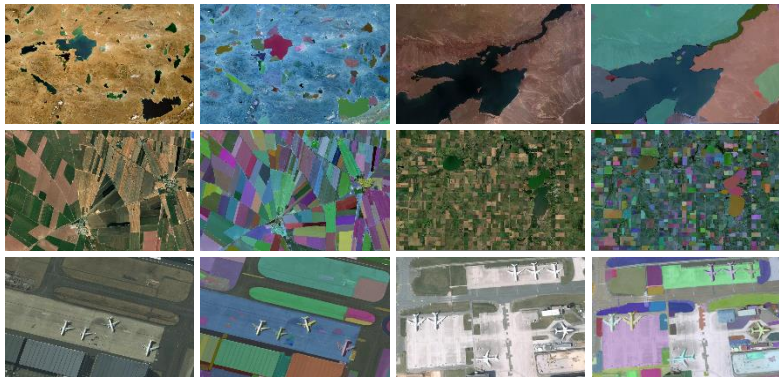
## CONTENTS

- 01 研究意义与挑战
- 02 研究现状与问题
- 03 研究框架与路线
- 04 实验模型与数据
- 05 结果与未来展望

## • Segment Anything?



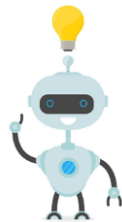
SAM-图片实例分割<sup>[1]</sup>



SAM-影像实例分割(湖泊/农田/建筑物/...)

## • 挑战：城市土地利用

- 土地覆盖(✓,自然属性): 耕地/林地/草地/...
- 城市土地利用(✗,自然-人文属性耦合): 工业/商业/住宅/...



## • 意义：城市规划/交通地理

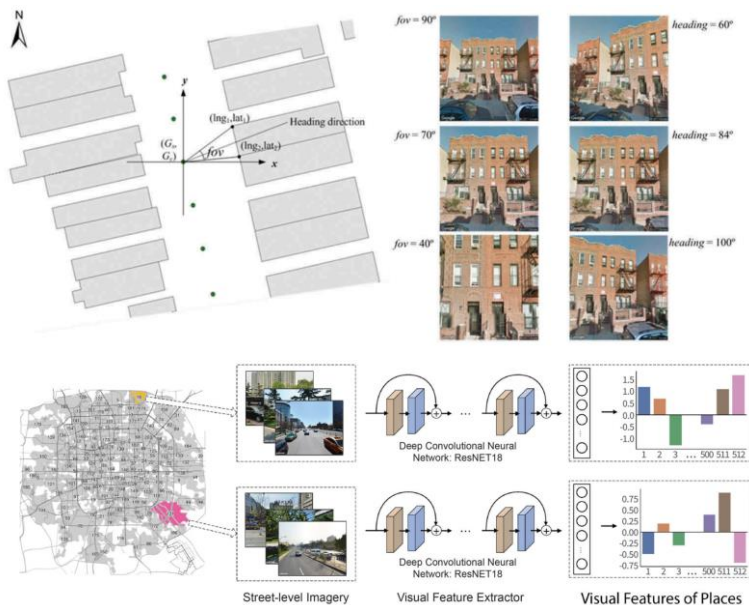
- 动态监测：原始规划  $\neq$  使用现状
- 规划建设：混合土地利用  $\propto$  城市活力、低交通压力



死气沉沉 - 交通拥堵

### 研究现状

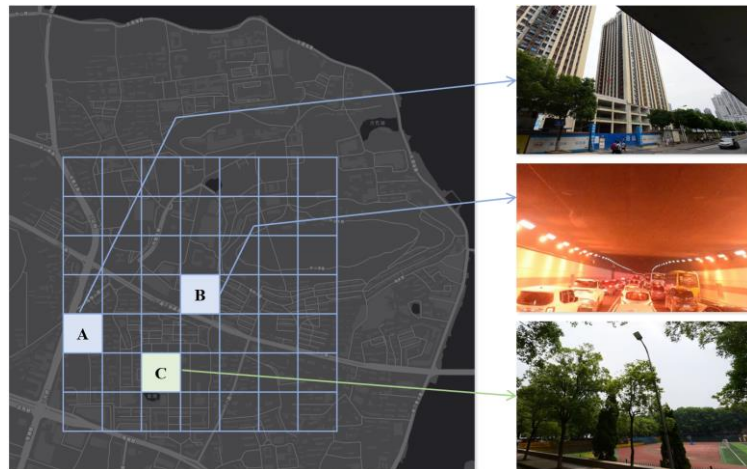
- 基于社会感知数据(街景/兴趣点)的方法
- 融合遥感影像数据的方法



基于街景影像的城市用地识别<sup>[2-3]</sup>

### 当前问题

- 社会感知数据的不完备性 (空间依赖性)

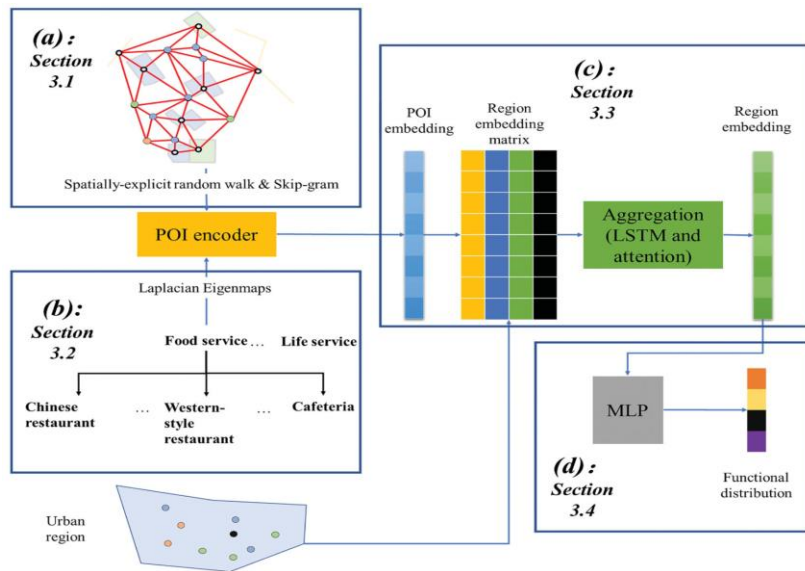


- A: 珞狮路 (预测: 商业主导)
- B: 八一路隧道 (预测: 交通主导)
- C: 星湖二路 (预测: 教育主导)



### 研究现状

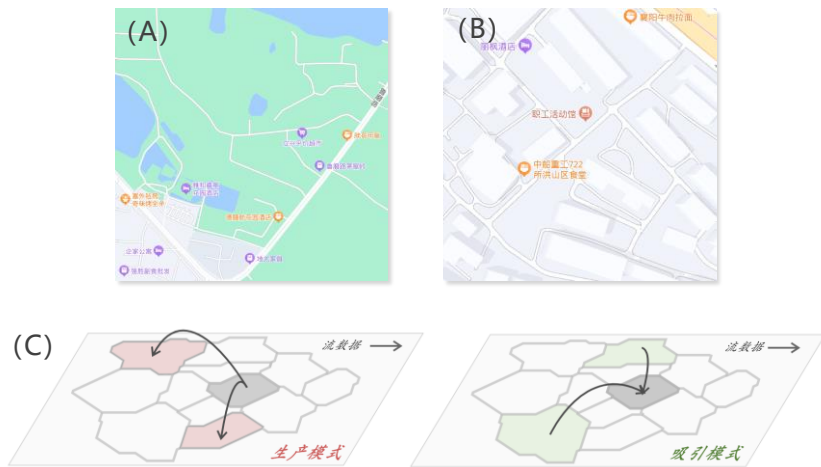
- 基于社会感知数据(街景/兴趣点)的方法
- 融合遥感影像数据的方法



基于兴趣点的城市用地识别<sup>[4-5]</sup>

### 当前问题

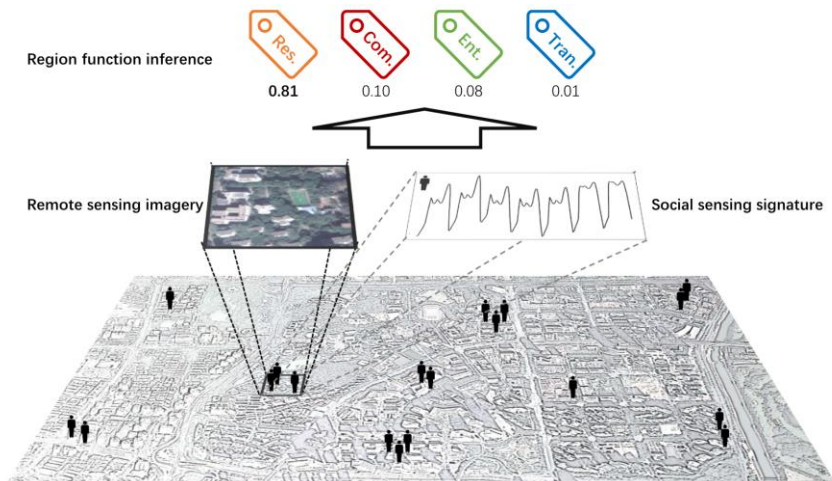
- 社会感知数据的不完备性 (空间依赖性)
- 社会感知数据的非代表性 (空间交互性)
- 区域的“自然-人文”二元耦合属性



- A: 绿地为主的区域兴趣点全为商业 (自然-人文耦合)
- B: 居住为主的区域缺少必要兴趣点 (空间交互)

### 研究现状

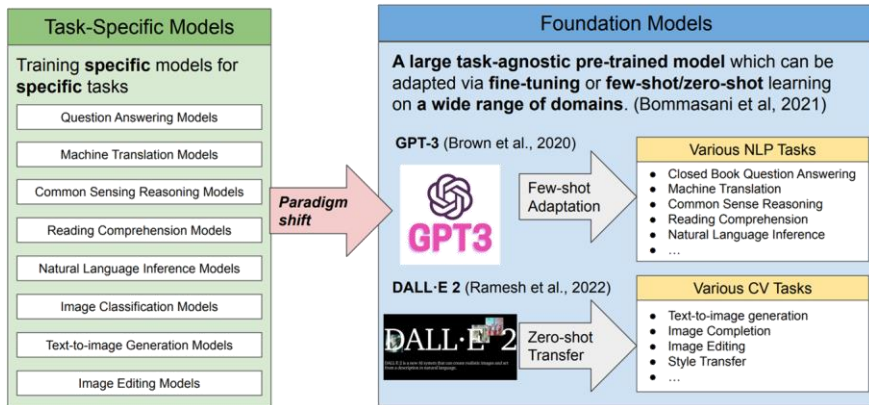
- 基于社会感知数据(街景/兴趣点)的方法
- 融合遥感影像数据的方法



融合遥感影像的城市用地识别<sup>[6]</sup>

### 当前问题

- 社会感知数据的不完备性 (空间依赖性)
- 社会感知数据的非代表性 (空间交互性)
- 区域的“自然-人文”二元耦合属性
- 区域的“空间-属性”二元耦合特征
- 研究范式的转移与视觉模型的训练难度 (数据规模)



自监督预训练基础模型兴起<sup>[7]</sup>



## 03 研究框架与路线

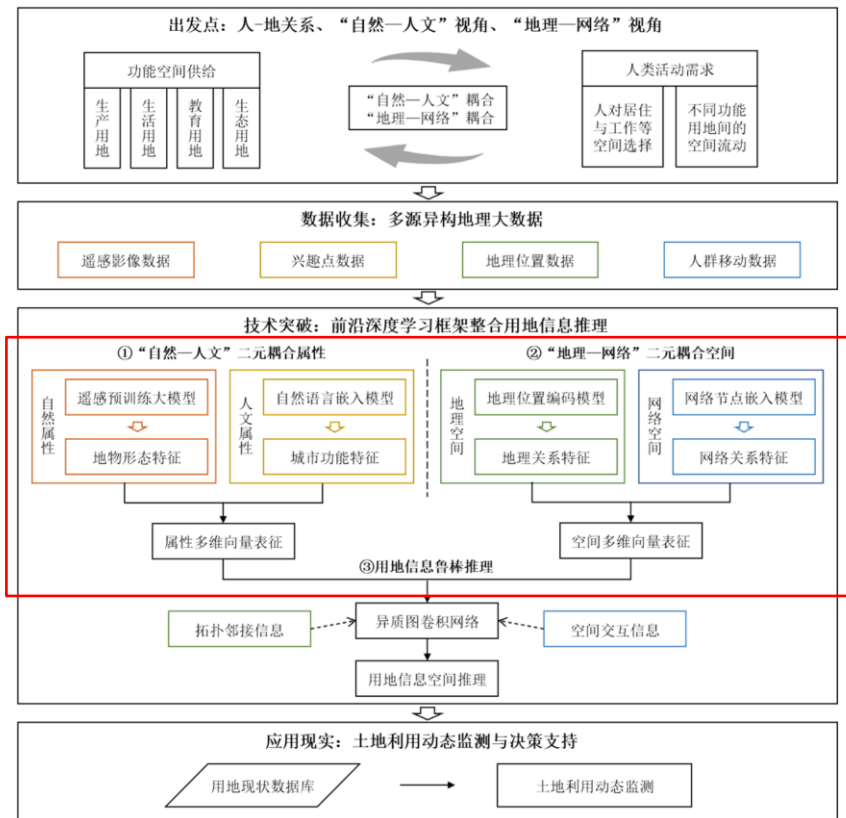
### • 当前问题

- 社会感知数据的不完备性 (空间依赖性)
- 社会感知数据的非代表性 (空间交互性)
- 区域的“自然-人文”二元耦合属性
- 区域的“空间-属性”二元耦合特征
- 研究范式的转移与视觉模型的训练难度 (数据规模)



### • 创新点

- 区域的多模态表征
  - 属性 (自然[遥感影像] - 人文[兴趣点])
  - 空间 (地理[经纬度] - 网络[节点嵌入])
- 人文属性的鲁棒更新
  - 创新性提出用于土地利用推理的异质图卷积神经网络
- 自然属性的任务无关抽取
  - 创新性运用遥感预训练大模型抽取影像特征



研究框架与技术路线图



## 03 研究框架与路线

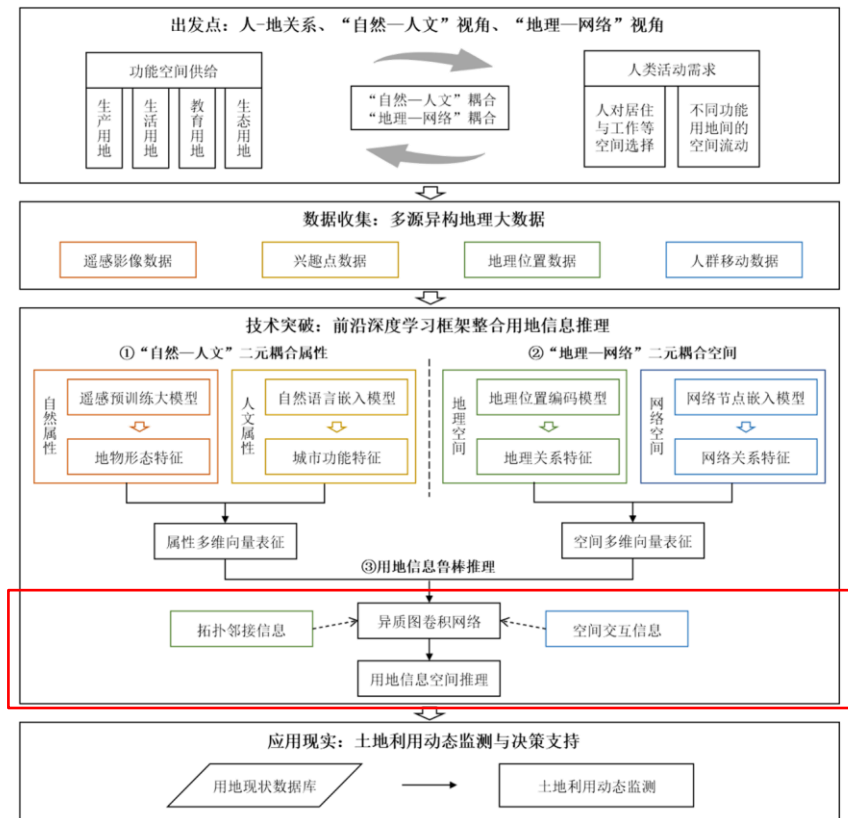
### • 当前问题

- 社会感知数据的不完备性 (空间依赖性)
- 社会感知数据的非代表性 (空间交互性)
- 区域的“自然-人文”二元耦合属性
- 区域的“空间-属性”二元耦合特征
- 研究范式的转移与视觉模型的训练难度 (数据规模)



### • 创新点

- 区域的多模态表征
  - 属性 (自然[遥感影像] – 人文[兴趣点])
  - 空间 (地理[经纬度] – 网络[节点嵌入])
- 人文属性的鲁棒更新
  - 创新性提出用于土地利用推理的异质图卷积神经网络
- 自然属性的任务无关抽取
  - 创新性运用遥感预训练大模型抽取影像特征



研究框架与技术路线图





## 03 研究框架与路线

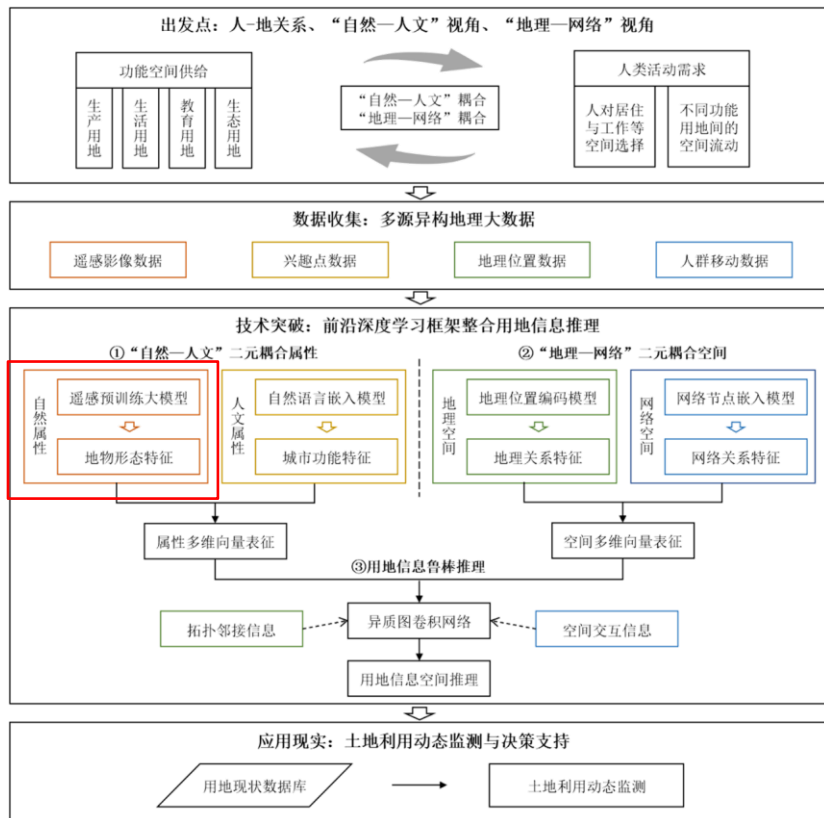
### • 当前问题

- 社会感知数据的不完备性 (空间依赖性)
- 社会感知数据的非代表性 (空间交互性)
- 区域的“自然-人文”二元耦合属性
- 区域的“空间-属性”二元耦合特征
- 研究范式的转移与视觉模型的训练难度 (数据规模)



### • 创新点

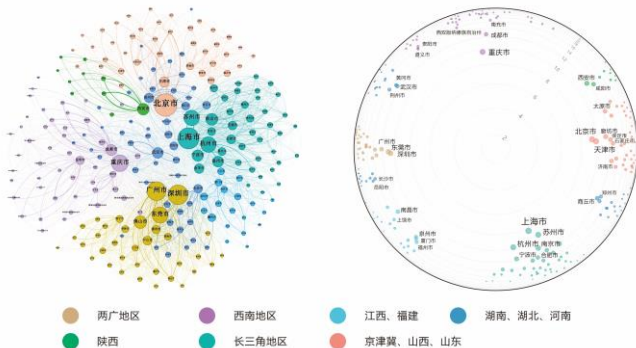
- 区域的多模态表征
  - 属性 (自然[遥感影像] – 人文[兴趣点])
  - 空间 (地理[经纬度] – 网络[节点嵌入])
- 人文属性的鲁棒更新
  - 创新性提出用于土地利用推理的异质图卷积神经网络
- 自然属性的任务无关抽取
  - 创新性运用遥感预训练大模型抽取影像特征



研究框架与技术路线图

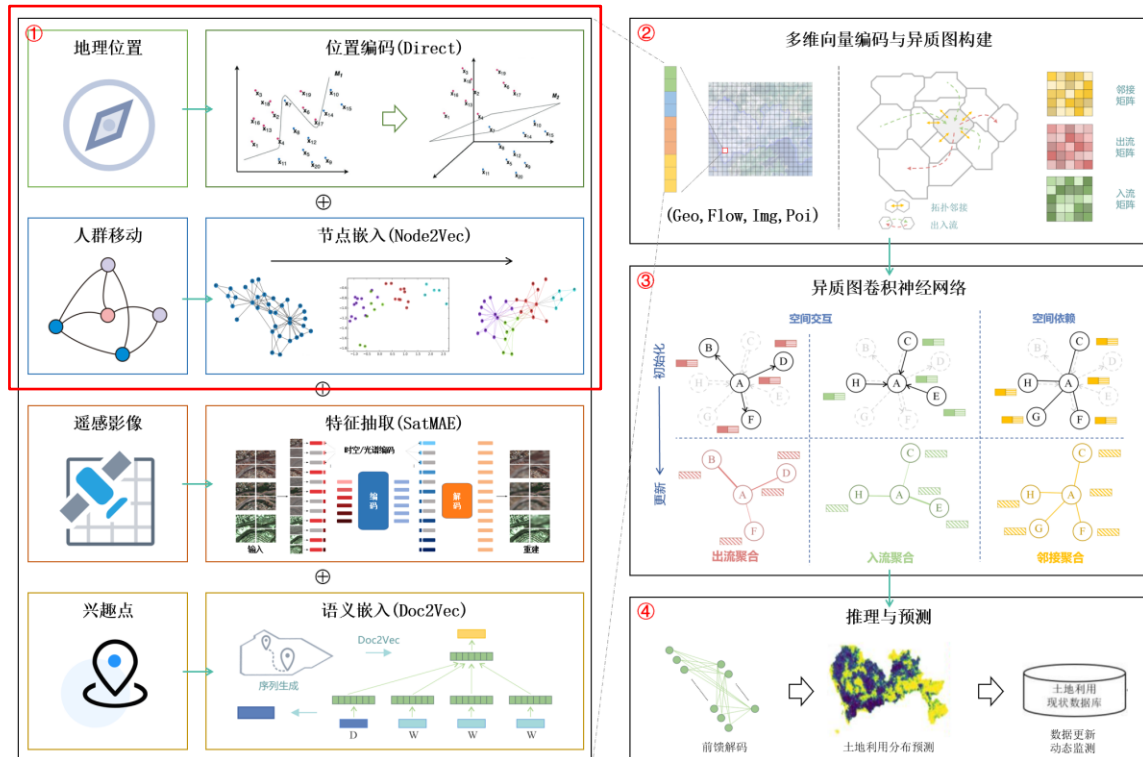
## 区域表征

- 地理位置编码
- 节点嵌入表征
- 影像特征抽取
- POI 语义挖掘
- 多维向量编码
- 异质图卷积神经网络 (理论依据)
- 城市土地利用信息推理



$$H_{geo} = Enc\_geo(X_{geo}) = NN(PE_{direct}([\lambda, \phi]))$$

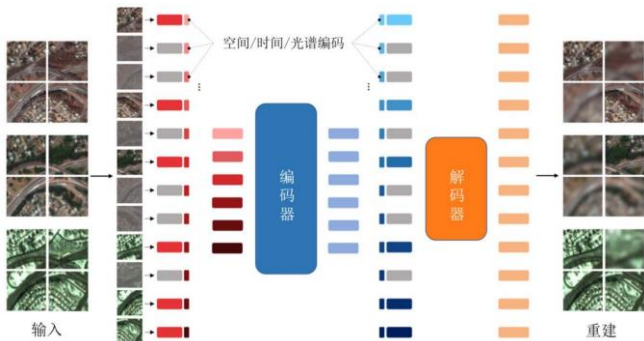
$$H_{flow} = Enc\_flow(X_{flow}) = Word2Vec(Generate\_seq(X_{flow}))$$



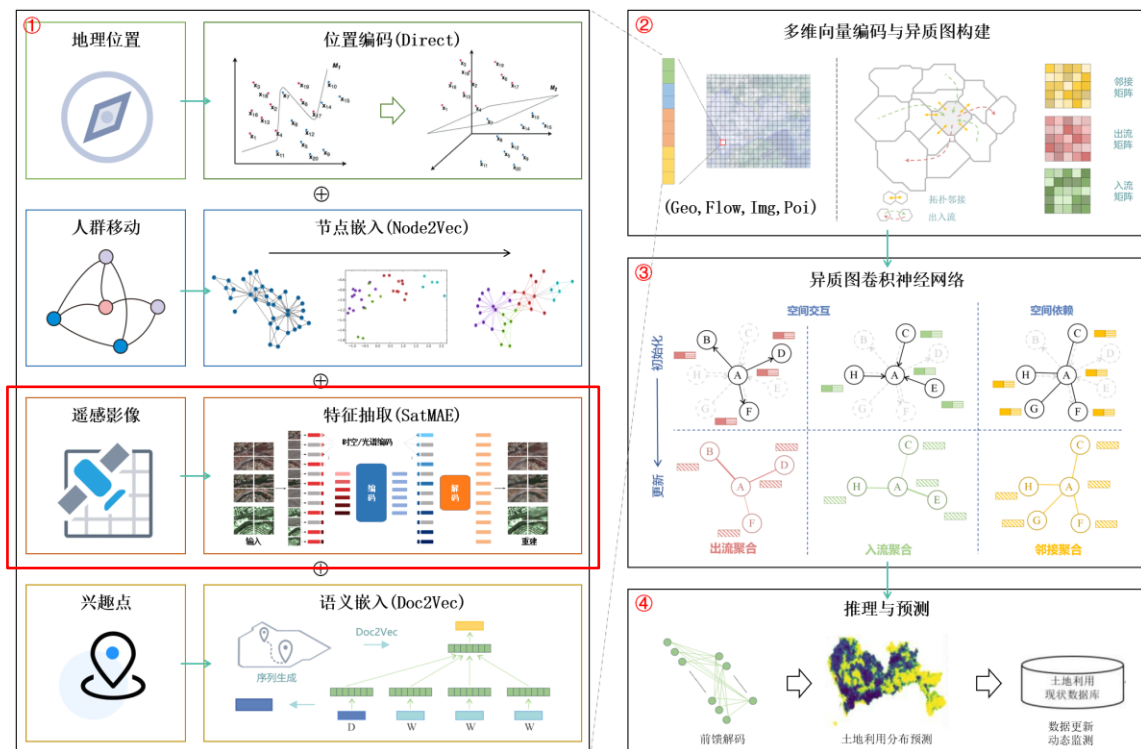
模型框架图

## 区域表征

- 地理位置编码
- 节点嵌入表征
- **影像特征抽取**
- POI 语义挖掘
- 多维向量编码
- 异质图卷积神经网络 (理论依据)
- 城市土地利用信息推理



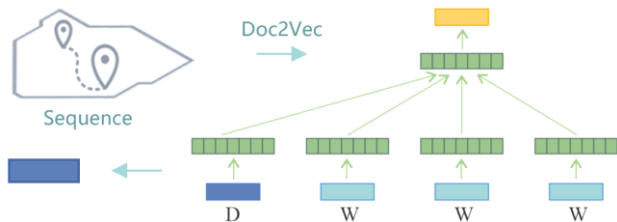
$$H_{img} = Enc\_img(X_{img}) = Pooling(Overlay(SatMAE(X_{img})))$$



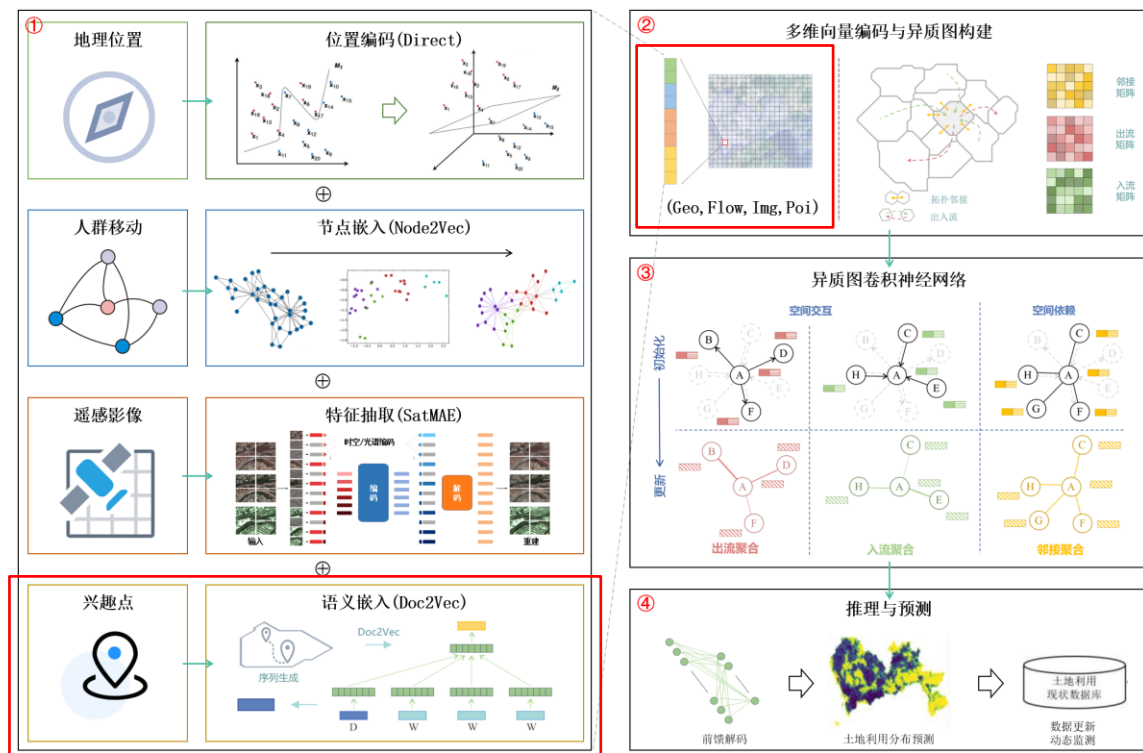
模型框架图

## 区域表征

- 地理位置编码
- 节点嵌入表征
- 影像特征抽取
- **POI 语义挖掘**
- **多维向量编码**
- 异质图卷积神经网络 (理论依据)
- 城市土地利用信息推理

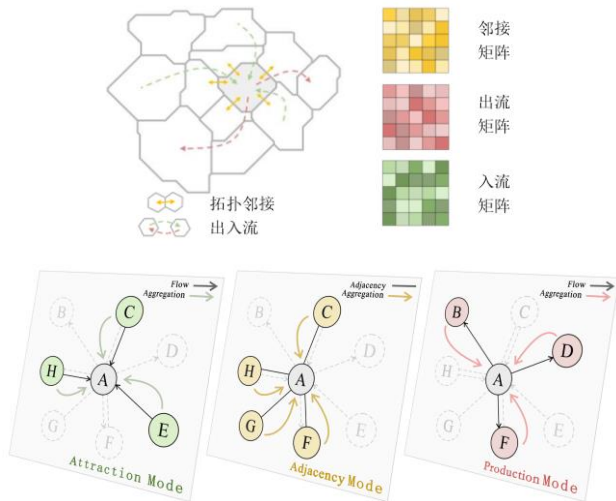


$$H_{poi} = Enc\_poi(X_{poi}) = Doc2Vec(Generate\_seq(X_{poi}))$$



模型框架图

- 区域表征
- 异质图卷积神经网络 (理论依据)
- 城市土地利用信息推理

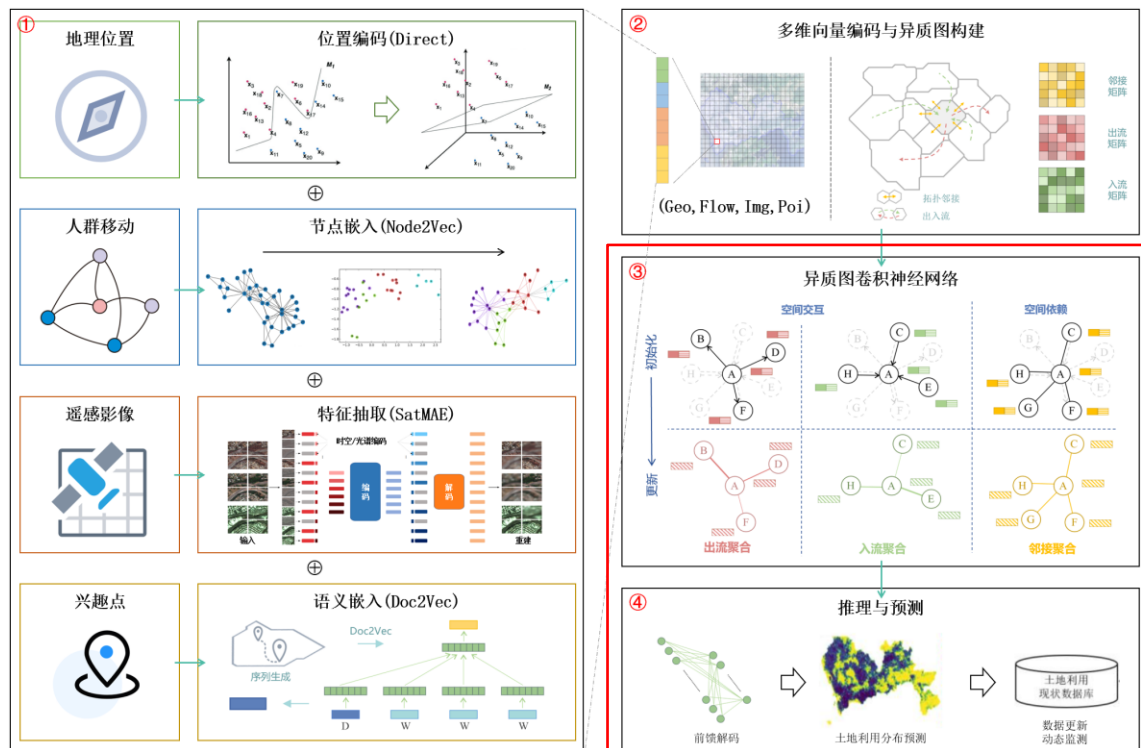


$$H' = HGCN(H^{t-1}, A, I) = \text{Concat}(\text{head}_{adj}^t, \text{head}_{in}^t, \text{head}_{out}^t)$$

$$\text{head}_{adj}^t = GCN(H^{t-1}W_{adj}^t, A)$$

$$\text{head}_{in}^t = GCN(H^{t-1}W_{in}^t, I^T)$$

$$\text{head}_{out}^t = GCN(H^{t-1}W_{out}^t, I)$$



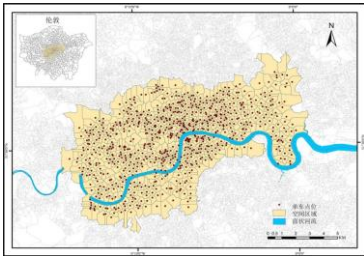
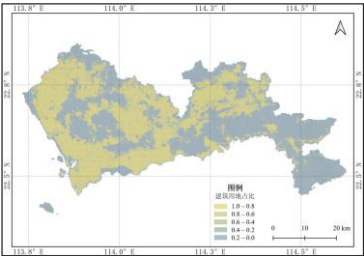
模型框架图





研究区域与数据

- 区域：伦敦/深圳
- 数据：流/兴趣点/遥感影像/土地利用
- 特点：不同空间单元划分/流类型/兴趣点分级



研究区域示意图

数据内容	深圳来源	伦敦来源
流数据	联通智慧足迹手机信令数据	共享单车租赁数据
兴趣点数据	高德地图数据	英国军械测量局数据
遥感影像数据	Landsat8 OLI	Landsat8 OLI
土地利用数据	第三次全国国土调查数据	Geomni UKMap 土地利用数据

研究数据来源

土地利用类型	深圳		伦敦	
	分类编码	土地利用类型	分类编码	土地利用类型
商业用地	05H1	商业和服务设施	18	办公区域
	0508	仓储	20	零售分销和服务
			21	仓储
工业用地	0601	制造业	16	制造业
	0602	矿业	17	矿业
住宅用地	0701	城市住宅	19	住宅
	0702	农村居民点		
行政和公共服务用地	08H1	新闻出版	12	社区和卫生服务
	08H2	教育医疗	13	军事
	08H2A	高等教育	14	教育
市政公用设施用地	0809	市政公用设施	24	市政公用设施
绿地公园用地	0810	公园绿地		
	0810A	广场	1501	娱乐休闲
街道和交通用地	1001	铁路		
	1002	轨道交通		
	1003	公路		
	1004	乡道		
	1005	交通服务场	22	运输
	1007	机场		
	1008	港口和码头		
	1009	管道运输		
	00	湿地		
非建设用地	01	耕地	11	农业和渔业
	02	种植园		
	03	森林		
	04	草地	23	未使用的土地、水和建筑物
	11	水域		

土地利用类型映射表



- 异质图卷积有效性测试
- 异质图卷积泛化性测试
- 多模态表征效果测试



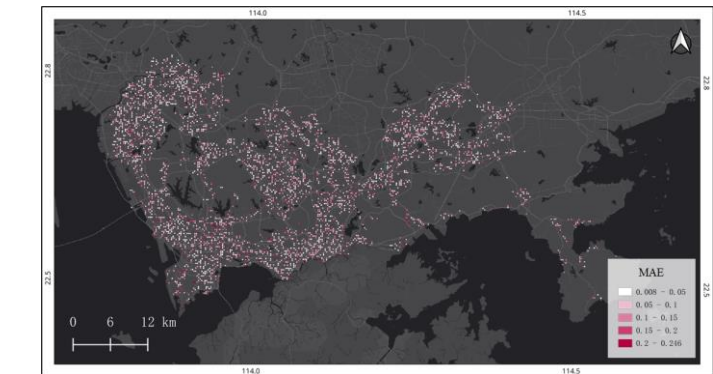
组别	? 方法	空间依赖			空间交互			依赖与交互特征处理
		自身属性	拓扑邻接	距离衰减	无向流	入流	出流	
A	None	✓						~
	GCN <sub>adj</sub>		✓					~
	GCN <sub>dd</sub>			✓				~
B	GCN <sub>avg</sub>	✓			✓			+
	GCN <sub>in</sub>	✓				✓		+
	GCN <sub>out</sub>	✓					✓	+
C	Bi-GCN <sub>avg</sub>	✓			✓			⊕
	Bi-GCN <sub>in</sub>	✓				✓		⊕
	Bi-GCN <sub>out</sub>	✓					✓	⊕
D	Tri-GCN	✓				✓	✓	⊕
E	Adj-Bi-GCN <sub>avg</sub>		✓		✓			⊕
	Adj-Bi-GCN <sub>in</sub>		✓			✓		⊕
	Adj-Bi-GCN <sub>out</sub>		✓				✓	⊕
	Adj-Tri-GCN(HGCN)		✓			✓	✓	⊕

有效性测试设计

组别	方法	评价指标			排名
		KL 散度 ↓	绝对误差 ↓	余弦相似 ↑	
A	None	0.5434 <sub>±0.0057</sub>	0.0926 <sub>±0.0006</sub>	0.7985 <sub>±0.0024</sub>	10
	GCN <sub>adj</sub>	0.5259 <sub>±0.0042</sub>	0.0912 <sub>±0.0005</sub>	0.7987 <sub>±0.0020</sub>	9
	GCN <sub>dd</sub>	0.7075 <sub>±0.0065</sub>	0.1139 <sub>±0.0006</sub>	0.7106 <sub>±0.0028</sub>	14
B	GCN <sub>avg</sub>	0.6714 <sub>±0.0067</sub>	0.1074 <sub>±0.0006</sub>	0.7331 <sub>±0.0028</sub>	11
	GCN <sub>in</sub>	0.6820 <sub>±0.0073</sub>	0.1087 <sub>±0.0007</sub>	0.7278 <sub>±0.0029</sub>	13
	GCN <sub>out</sub>	0.6795 <sub>±0.0072</sub>	0.1086 <sub>±0.0006</sub>	0.7287 <sub>±0.0029</sub>	12
C	Bi-GCN <sub>avg</sub>	0.5203 <sub>±0.0061</sub>	0.0895 <sub>±0.0006</sub>	0.8082 <sub>±0.0025</sub>	6
	Bi-GCN <sub>in</sub>	0.5241 <sub>±0.0063</sub>	0.0900 <sub>±0.0006</sub>	0.8065 <sub>±0.0027</sub>	8
	Bi-GCN <sub>out</sub>	0.5208 <sub>±0.0060</sub>	0.0896 <sub>±0.0006</sub>	0.8080 <sub>±0.0024</sub>	7
D	Tri-GCN	0.5144 <sub>±0.0057</sub>	0.0890 <sub>±0.0005</sub>	0.8103 <sub>±0.0024</sub>	5
E	Adj-Bi-GCN <sub>avg</sub>	0.5005 <sub>±0.0053</sub>	0.0877 <sub>±0.0006</sub>	0.8112 <sub>±0.0022</sub>	2
	Adj-Bi-GCN <sub>in</sub>	0.5025 <sub>±0.0052</sub>	0.0880 <sub>±0.0005</sub>	0.8103 <sub>±0.0022</sub>	4
	Adj-Bi-GCN <sub>out</sub>	0.5012 <sub>±0.0053</sub>	0.0878 <sub>±0.0005</sub>	0.8108 <sub>±0.0021</sub>	3
	Adj-Tri-GCN(HGCN)	0.4974 <sub>±0.0053</sub>	0.0876 <sub>±0.0005</sub>	0.8119 <sub>±0.0023</sub>	1

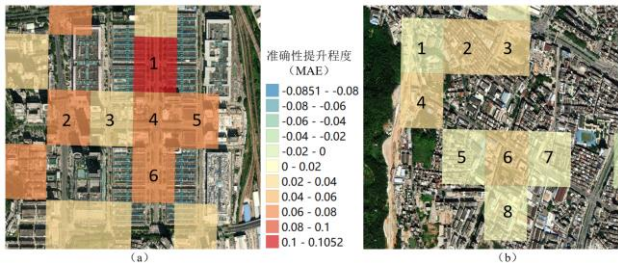
深圳有效性测试结果 (伦敦见论文)

- 异质图卷积有效性测试
- 异质图卷积泛化性测试
- 多模态表征效果测试



土地利用	评价指标					
	None			GCNadj		
	绝对误差	相对误差	绝对误差	相对误差	绝对误差	相对误差
行政和公共服务用地	<b>0.060</b>	136.1%	0.062	137.2%	<b>0.060</b>	133.2%
市政公用设施用地	<b>0.021</b>	169.2%	<b>0.021</b>	170.4%	<b>0.021</b>	168.2%
商业用地	0.088	125.5%	0.087	122.6%	<b>0.084</b>	122.3%
住宅用地	0.145	81.0%	0.149	82.5%	<b>0.138</b>	80.7%
工业用地	0.141	117.3%	0.130	109.7%	<b>0.122</b>	108.3%
绿地公园用地	0.062	163.8%	<b>0.058</b>	153.0%	<b>0.058</b>	160.7%
街道与交通用地	0.107	59.6%	0.110	60.8%	<b>0.106</b>	55.9%
非建筑用地	0.109	128.9%	0.105	126.5%	<b>0.100</b>	119.3%

绝对误差分析



相对误差分析: (a)GCNadj-None; (b)HGCN-GCNadj

ID	土地利用类型							
	行政公共 服务用地	市政公用 设施用地	商业 用地	住宅 用地	工业 用地	绿地公园 用地	街道与 交通用地	非建筑 用地
1	0	0	<b>0.694</b>	0	0	0	0.306	0
2	0	0	<b>0.559</b>	0.054	0.022	0	0.364	0
3	0	0	<b>0.718</b>	0	0	0	0.282	0
4	0	0	<b>0.579</b>	0	0	0	0.421	0
5	0	0	<b>0.770</b>	0	0	0	0.230	0
6	0	0	<b>0.695</b>	0	0	0	0.305	0

图(a)区域特征

ID	土地利用类型							
	行政公共 服务用地	市政公用 设施用地	商业 用地	住宅 用地	工业 用地	绿地公园 用地	街道与 交通用地	非建筑 用地
1	0	0.255	0.005	<b>0.446</b>	0.201	0.024	0.049	0.020
2	0.082	0.029	0.132	<b>0.702</b>	0.020	0	0.035	0
3	0	0	<b>0.643</b>	0.221	0.033	0	0.102	0.001
4	0	0.120	0.018	0.099	<b>0.383</b>	0.072	0.145	0.163
5	0	0	0.019	<b>0.590</b>	0.310	0	0.045	0.036
6	0.158	0	0.041	0.130	<b>0.482</b>	0	0.132	0.057
7	0.046	0	<b>0.330</b>	0.147	0.264	0	0.213	0
8	0.036	0	0.050	0.295	<b>0.565</b>	0	0.040	0.014

图(b)区域特征

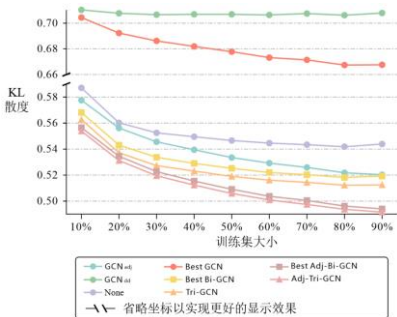


## 05 结果与未来展望

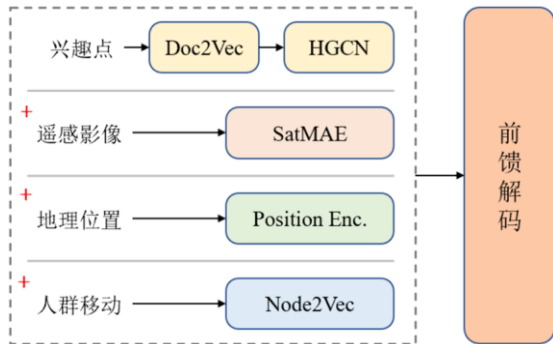
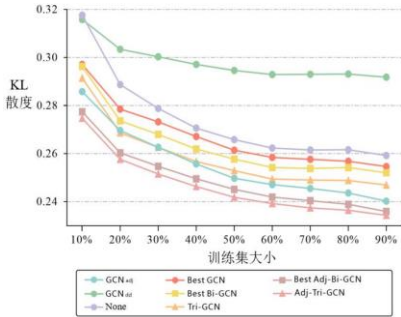
- 异质图卷积有效性测试
- 异质图卷积泛化性测试
- 多模态表征效果测试

训练集大小	KL 散度 ↓	
	迁移微调模型	从头训练模型
10%	0.2714	0.2747
20%	0.2544	0.2576
30%	0.2416	0.2515
40%	0.2332	0.2463
50%	0.2298	0.2418
60%	0.2238	0.2392
70%	0.2135	0.2374
80%	0.2143	0.2364
90%	0.2114	0.2343

不同地区的迁移效应



相同地区的规模效应



方法	评价指标			排名
	KL 散度 ↓	绝对误差 ↓	余弦相似 ↑	
M <sub>P</sub>	0.4974 $\pm$ 0.0053	0.0876 $\pm$ 0.0005	0.8119 $\pm$ 0.0023	4
M <sub>PI</sub>	0.4631 $\pm$ 0.0067	0.0847 $\pm$ 0.0008	0.8187 $\pm$ 0.0031	3
M <sub>PIG</sub>	0.4619 $\pm$ 0.0065	0.0845 $\pm$ 0.0007	0.8187 $\pm$ 0.0028	2
<b>M<sub>PIGF</sub></b>	<b>0.4607<math>\pm</math>0.0058</b>	<b>0.0839<math>\pm</math>0.0006</b>	<b>0.8193<math>\pm</math>0.0027</b>	<b>1</b>

多模态表征消融实验



通用人工智能的应用  
地理空间多模态大模型

[Instruction] There are six land use types: (1) residential, (2) commercial, (3) industrial, (4) education, health care, civic, governmental and cultural, (5) transportation facilities, and (6) outdoors and natural.

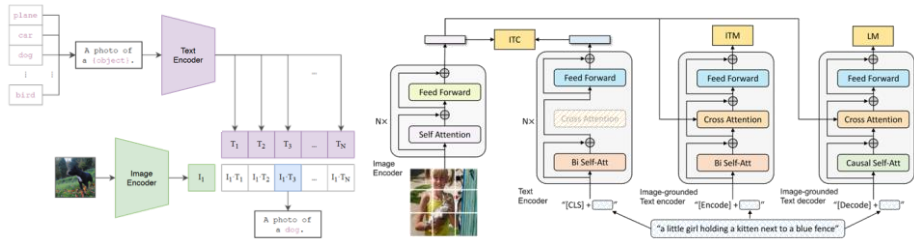
Paragraph: In this urban region, there are 128 points of interest, including 2 Chinese restaurant, 1 food restaurant, 2 hotel, 2 apartment hotel, 1 daily life service, 1 mobile communication shop, 24 company, 1 logistics company, 1 real estate agency, 1 lottery retailer, 3 beauty shop, 1 manicure, 2 barber shop, 4 Internet cafe, 3 bath massage, 2 stadium, 4 training institutions, 1 pharmacy, 4 automotive sale, 6 car service, 2 car repair, 1 car rental, 1 Automobile parts, 3 shopping, 5 shop, 5 parking lot, 5 Parking lot entrance, 2 transportation facility, 1 port harbor, 1 road intersection, 1 atm machine, 2 office building, 2 residential area, 7 building, 1 real estate, 1 park, 1 factory, 7 administrative agency, 1 entrance and exit, 3 gate door, 6 convenience store, 4 home building materials.

Q: What is the primary land use category of this urban region?  
A: outdoors and natural

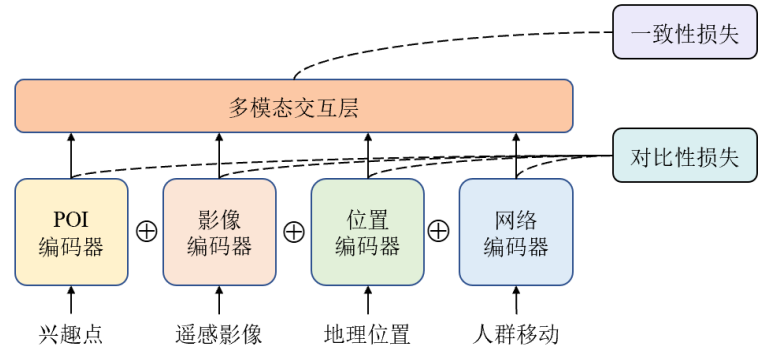
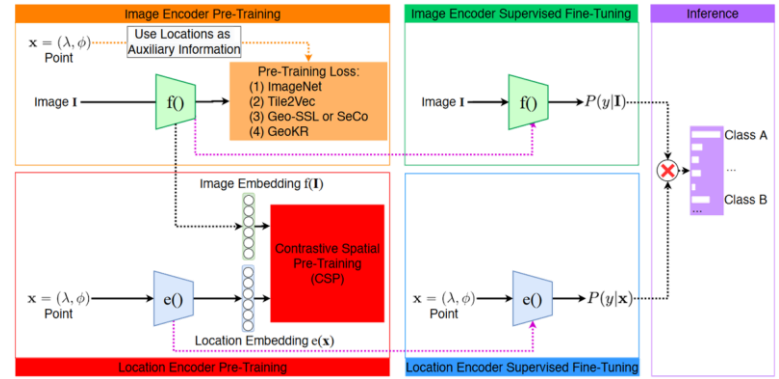
Paragraph: In this urban region, there are 17 points of interest, including 1 food restaurant, 3 public toilet, 3 funeral service, 2 road station for walking and cycling, 1 beach, 2 parking lot, 2 road intersection, 1 corporate company enterprise, 2 administrative agency.

Q: What is the primary land use category of this urban region?  
A: outdoors and natural

模型	提示
CLIP	我们首先视觉编码器编码遥感影像，文本编码器端编码构建的 8 种提示“这是一个以[土地利用类型]为主导的地区”，然后计算视觉特征与文本特征相似性并取最大值作为目标类型。
BLIP	视觉编码器编码遥感影像，我们使用提示“在这个地区中有 17 个 POI，其中 1 个是餐厅，...，那么这个区域的土地利用类型是什么？是住宅区、...还是工业区？”让模型生成答案。



通用人工智能的应用<sup>[8]</sup>



地理空间多模态大模型<sup>[9]</sup>





- [1] Kirillov A, Mintun E, Ravi N, et al. Segment anything[J]. arXiv preprint arXiv:2304.02643, 2023.
- [2] Li X, Zhang C, Li W. Building block level urban land-use information retrieval based on Google Street View images[J]. GIScience & Remote Sensing, 2017, 54(6): 819-835.
- [3] Zhu D, Zhang F, Wang S, et al. Understanding place characteristics in geographic contexts through graph convolutional neural networks[J]. Annals of the American Association of Geographers, 2020, 110(2): 408-420.
- [4] Huang W, Cui L, Chen M, et al. Estimating urban functional distributions with semantics preserved POI embedding[J]. International Journal of Geographical Information Science, 2022, 36(10): 1905-1930.
- [5] Huang W, Zhang D, Mai G, et al. Learning urban region representations with POIs and hierarchical graph infomax[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2023, 196: 134-145.
- [6] Cao R, Tu W, Yang C, et al. Deep learning-based remote and social sensing data fusion for urban region function recognition[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 163: 82-97.
- [7] Mai G, Cundy C, Choi K, et al. Towards a foundation model for geospatial artificial intelligence (vision paper)[C]//Proceedings of the 30th International Conference on Advances in Geographic Information Systems. 2022: 1-4.
- [8] Mai G, Huang W, Sun J, et al. On the opportunities and challenges of foundation models for geospatial artificial intelligence[J]. arXiv preprint arXiv:2304.06798, 2023.
- [9] Mai G, Lao N, He Y, et al. CSP: Self-Supervised Contrastive Spatial Pre-Training for Geospatial-Visual Representations[J]. arXiv preprint arXiv:2305.01118, 2023.



# 面向多源异构空间大数据的城市用地 识别与监测：一种多模态深度学习框架

答辩人：王成龙

指导老师：焦利民教授、宫兆亚助理教授