

分类号 C8/355
U D C

密级 公开
编号 10741



硕士学位论文
(专业学位)

论文题目 基于多源数据的城市功能区识别与医疗设施布局优化
——以兰州市主城区为例

研究生姓名: 闫新宇

指导教师姓名、职称: 牛成英 教授

学科、专业名称: 应用统计硕士

研究方向: 大数据分析及应用

提交日期: 2023年5月30日

独创性声明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名： 闫新宇 签字日期： 2023年5月30日

导师签名： 牛咏英 签字日期： 2023年5月30日

导师(校外)签名： 张亚东 签字日期： 2023年5月30日

关于论文使用授权的说明

本人完全了解学校关于保留、使用学位论文的各项规定， 同意（选择“同意”/“不同意”）以下事项：

- 1.学校有权保留本论文的复印件和磁盘，允许论文被查阅和借阅，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文；
- 2.学校有权将本人的学位论文提交至清华大学“中国学术期刊（光盘版）电子杂志社”用于出版和编入CNKI《中国知识资源总库》或其他同类数据库，传播本学位论文的全部或部分內容。

学位论文作者签名： 闫新宇 签字日期： 2023年5月30日

导师签名： 牛咏英 签字日期： 2023年5月30日

导师(校外)签名： 张亚东 签字日期： 2023年5月30日

**Urban spatial feature recognition and
medical facility layout optimization based
on multi-source data——a case study of
Lanzhou urban area**

Candidate : Yan Xin Yu

Supervisor: Niu Cheng Ying

摘要

进入“十四五”规划时期，面对我国城市发展不均衡、不充分的现状，如何优化公共服务设施布局，助力解决相关问题成为当前城市规划研究的热点。近年来得益于智能移动设备的普及，海量的人类活动数据得以保存，形成结构多样的多源大数据。通过机器学习与空间统计分析，挖掘多源大数据中的信息，并以此为基础进行公共服务设施布局的优化设计是一种有效手段。将城市遥感影像数据、POI (Point of interest) 数据、人口统计数据整合为多源数据集，从城市空间特征识别的角度出发，对城市功能区、设施服务可达性测度、设施选址推荐三个角度进行分析研究，旨在对城市公共服务设施的空间布局进行更为精准的认识和综合优化。主要工作包括以下三个方面：

首先，基于 Mask R-CNN 算法和样方密度法，以兰州市主城区为例，通过挖掘遥感影像数据的自然特征和 POI 数据中的人文特征，识别建筑物轮廓和相关功能信息，对城市功能区进行划分并识别其功能类型。相较于单一视角和数据进行的功能区识别，结果更加精准、客观，更加贴合城市现状。

其次，基于改进的两步移动搜索法，对兰州市主城区医疗服务设施的空间可达性进行测度。利用第七次人口普查数据、POI 数据、统计年鉴数据等，划分三级医疗服务设施，分级构建服务半径，从供需角度出发测算服务距离，旨在更加精准的反映当前研究区内医疗设施的空间可达性情况，也更加贴合医疗服务的方式。结果显示当前研究区内的医疗设施空间可达性总体较好，但仍然存在部分区域设施空间分布不均衡的情况。

最后，利用随机森林算法的分类功能，分析兰州市主城区医疗服务设施的分布情况并进行选址推荐。通过引入机器学习算法，挖掘 POI 数据中医疗设施的空间分布特征，并以此为基础预测不同区域医疗设施分布的推荐度，降低了选址推荐模型的主观性。结果显示选址推荐度较高的区域大多分布在城市郊区，与实际情况基本相符。对兰州市主城区医疗设施空间分布特征进行综合考虑后，设计兰州市主城区医疗设施布局优化方案。

关键字：多源数据、功能区识别、空间可达性测度、医疗设施布局优化

Abstract

As we enter the "14th Five-Year Plan" period, optimizing the layout of public service facilities to help address the issue of unbalanced and inadequate urban development in China has become a hot topic in urban planning research. In recent years, the widespread use of intelligent mobile devices has enabled the storage of massive amounts of human activity data, resulting in structurally diverse multi-source big data. By using machine learning and spatial statistical analysis to extract information from multi-source big data, optimizing the design of public service facility layout becomes an effective means. This thesis combines multiple data sources, including urban remote sensing image data, Point of Interest (POI) data, and demographic data, to analyze and study urban functional areas, facility service accessibility, and facility location recommendations from the perspective of urban spatial feature recognition, achieving more precise recognition and comprehensive optimization of the spatial layout of urban public service facilities. The main work includes the following three aspects:

Firstly, based on the Mask R-CNN algorithm and sample density method, taking Lanzhou's main urban area as an example, this article mines the natural features of remote sensing image data and the humanistic features of POI data to identify building outlines and relevant functional information, dividing the urban functional areas and identifying their functional types. Compared with the identification of functional areas using a single perspective and data, this method achieves more accurate and objective results, better fitting the urban reality.

Secondly, using an improved two-step Floating Catchment method, this article measures the spatial accessibility of medical service facilities in Lanzhou's main urban area. By using data from the seventh national census, POI data, and statistical yearbooks, three levels of medical service facilities are defined, and service radii are constructed to calculate service distances from the supply and demand perspectives. This method aims to more accurately reflect the current spatial accessibility of medical facilities in the research area, and it better fits the service mode of medical facilities. The results show that the overall spatial accessibility of medical facilities in the research area is good, but there are still some areas where the distribution of facilities is uneven.

Finally, by using the classification function of the random forest algorithm, this

thesis analyzes the distribution of medical service facilities in Lanzhou's main urban area and makes location recommendations. By introducing machine learning algorithms and extracting the spatial distribution characteristics of medical facilities from POI data, this method predicts the recommendation degree of medical facility distribution in different regions, reducing the subjectivity of the location recommendation model. The results show that the regions with high location recommendation degrees are mostly located in the suburbs of the city, which is consistent with the current situation. After considering the comprehensive spatial distribution characteristics of medical facilities in Lanzhou's main urban area, an optimized layout plan for medical facilities in Lanzhou's main urban area is proposed.

Keywords: Multi-source data ; Spatial feature recognition ; Accessibility ; Optimization of medical facility layout

目录

摘要.....	I
Abstract.....	II
1. 绪论.....	1
1.1 选题依据与研究目的.....	1
1.1.1 选题依据.....	1
1.1.2 研究目的.....	1
1.2 国内外研究现状.....	2
1.2.1 国外研究现状.....	2
1.2.2 国内研究现状.....	5
1.3 主要研究内容.....	7
1.3.1 主要研究内容.....	7
1.3.2 重点解决的问题.....	9
1.4 主要研究方法.....	9
1.5 论文组织结构.....	10
2. 研究数据来源.....	11
2.1 研究区域.....	11
2.2 遥感影像数据.....	11
2.3 POI 数据.....	12
2.4 人口栅格数据.....	13
2.5 兰州市第七次人口普查数据.....	13
3. 基于 Mask R-CNN 算法和样方密度法的城市功能区识别.....	14
3.1 引言.....	14
3.2 研究方法.....	15
3.2.1 Mask R-CNN 算法.....	15
3.2.2 POI 点样方密度法.....	16
3.2.3 基于 Mask R-CNN 和样方密度法的城市功能区识别方法.....	17
3.3 实证分析.....	18
3.3.1 基于 POI 点样方密度法的功能区识别.....	18

3.3.2 自然特征与人文特征提取.....	19
3.3.3 基于自然特征与人文特征校验的城市功能区识别结果.....	22
3.3.4 精度评价.....	26
3.4 小结.....	27
4. 基于改进两步移动搜索法的医疗设施空间可达性评价.....	30
4.1 引言.....	30
4.2 两步移动搜索法.....	31
4.3 数据处理.....	31
4.3.1 人口栅格数据.....	31
4.3.2 医疗服务设施数据.....	32
4.3.3 空间投影与网格化.....	33
4.4 基于改进两步移动搜索法的可达性测度.....	33
4.4.1 改进的两步移动搜索法.....	33
4.4.2 医疗设施空间可达性测度.....	34
4.5 兰州市主城区医疗设施空间可达性结果分析.....	34
4.6 小结.....	36
5. 基于随机森林方法的医疗设施选址推荐.....	37
5.1 引言.....	37
5.2 随机森林算法.....	38
5.3 POI 数据的整合与网格化处理.....	39
5.3.1 POI 数据的整合.....	39
5.3.2 研究区域网格化处理.....	41
5.4 选址推荐模型的构建与训练.....	41
5.5 兰州市主城区医疗设施选址推荐结果分析.....	42
5.6 小结.....	44
6. 基于多源数据的医疗设施布局优化方案.....	45
6.1 推荐度较低且可达性较差区域优化方案.....	46
6.2 推荐度较低且可达性较好区域优化方案.....	46
6.3 推荐度较高且可达性较好区域优化方案.....	47

6.4 推荐度较高且可达性较差区域优化方案.....	47
7. 结论与展望.....	51
7.1 结论.....	51
7.2 展望.....	51
参考文献.....	52
致谢.....	57

1. 绪论

1.1 选题依据与研究目的

1.1.1 选题依据

改革开放以来，我国经济高速发展，与此同时城镇化呈现出快速发展趋势。40 多年过去，2021 年中国常住人口城镇化率已达 64.72%。城市的高速扩张与经济的快速发展带来了就业红利，吸引更多外来人口涌入城市，继而不可避免的催生出城市内各类公共服务设施需求量不足、设施空间分配不均等诸多问题。

医疗设施作为城市公共设施的重要组成部分，不仅担负着居民的医疗职能，其数量和质量更是评价一个城市发展状况的重要指标。“十三五”以来，我国城镇医疗机构数量和床位数都保持着较高的增长速度，各级医疗设施的服务能力也有了显著的提高（2022，国家卫生健康委）。特别是 2020 年新冠疫情爆发以来各类突发事件中的医疗挤兑现象，也无疑暴露出当下我国部分地区城镇人均医疗资源不足，难以满足居民医疗需求的现状。

2021 年党的“十四五”规划提出，要推进新型城市建设，打造智慧城市、数字城市，优化城市内各类公用服务设施布局，提升服务能力，使更多人民群众享有更高品质的城市生活。这些规划为解决当前存在的诸多“城市病”问题指明了方向。

信息技术的发展和移动互联网的广泛普及，使得与城市居民日常生产生活相关的各类数据得以收集与保存，这些数据与地理信息数据、政府统计数据等共同形成了庞大且冗杂的多源大数据库。当前多源数据已被广泛的使用于经济、金融等相关领域进行研究，如何挖掘多源数据中蕴含的城市功能区空间特征和人文活动信息，并将其用于城市治理和规划中，优化设施空间布局，助力城市优化升级，已成为当前相关领域学者研究和关注的重点。

1.1.2 研究目的

基于多源数据的城市空间特征识别，目的就在于通过挖掘卫星遥感数据、POI（Point of Interest）数据、人口数据等多源数据，尽可能准确的描述城市特有

设施分布规律，在此基础上实现更为科学的规划布局。城市公共服务设施分布的空间特征有着多种表征形式，主要选择基于城市功能区识别与划分、设施空间可达性测度和设施选址推荐三方面进行研究。

结合多源数据，从三个方面挖掘公共服务设施分布的空间特征。综合考虑影响分布的各类因素，利用空间统计分析和机器学习方法，一方面可以对公共服务设施分布情况进行精准识别，为优化设施布局方案提供依据。另一方面通过机器学习方法挖掘多源数据信息，从多个角度对城市设施的分布特点进行描述，可以降低定量分析研究主观性。因此基于多源数据的城市空间特征识别与设施布局优化具有一定的理论意义。

由于不同城市具有各自独特的地理条件和人文特征，准确把握城市独有的空间特征，并充分考虑人文特征，才能更好的实现城市公共服务设施的布局优化。兰州市作为甘肃省省会城市，位于我国大陆陆域版图的几何中心，是西北地区的主要城市之一，也是新亚欧大陆桥中国段五大中心城市之一，是西部重要的区域商贸中心和现代物流基地。以兰州市为例研究医疗设施的空间分布情况，能够助力当地政府的城市治理水平和能力提高，具有一定的现实意义。

1.2 国内外研究现状

1.2.1 国外研究现状

基于研究目的与研究的主要内容，主要从以下几个方面分析相关研究的现状，为论文研究提供理论指导和参考依据。

(1) 城市功能区识别

城市功能区是描述一定区域内人类活动的抽象概念，是区域内不同类型公共服务设施聚集的结果，也被视为城市规划的基本单元。通过梳理文献发现对城市功能区识别与划分问题一般采用三种方法。

部分学者选择通过统计聚类的方法进行功能区的划分。Yang 等（2016）首先基于道路数据进行区域的分割，再对潜在移动轨迹（LAT）进行聚类并以此推测区域的功能。Zhang 等（2018）则是基于 Kmean 算法对城市公共自行车的租赁记录和 POI 数据进行聚类；Yuan 等（2018）通过 GIS（Geographic Information

System 地理信息系统) 轨迹数据对区域进行划分, 并基于主题模型对区域内的 POI 数据进行处理以推断和总结区域功能; Du 等 (2019) 基于公共自行车流动数据、出租车流动数据, 通过聚类, 将其中的统计特征转为潜在语义特征, 并使用分类方法识别具有相似潜在主题的功能区。可以看出, 部分学者通过对人类活动数据进行聚类分析, 能够较为精准的概况相关区域的属性和功能, 但当前的人文数据大多是由城市居民主动上传或被动采集得到的, 与城市居民生活相关度越高则数据量越大, 因此城市内人类活动数据大多集中在城市中心的热点区域, 对于城市郊区和边缘区域存在数据缺失, 因而可能会造成对于城市边缘区域功能的误判

近年来随着深度学习等人工智能算法的兴起, 通过机器学习算法挖掘地理信息数据进行功能区识别已成为主流。Zhang 等 (2016) 提出一种利用遥感数据和卷积神经网络 (Convolutional Neural Networks, CNN) 算法的功能区识别方法, 并对沈阳和北京进行测试。Zhou 等 (2020) 通过提出超对象卷积神经网络 (Super Object -CNN) 算法进行城市功能区的识别。由此可见, 利用机器学习算法可以挖掘地理信息数据中的地类特征, 在区分建筑、绿地、空地等自然地理特征时候有着较高精度的识别效果。但城市与人类经济活动高度相关, 仅通过土地分类的角度无法详细区分商业、住宅等人文特征。

部分学者选择将遥感数据与各类人文社会数据相结合, 由此挖掘功能区的相关特征。Zhang 等 (2017) 通过高分遥感数据 (VHR) 对城市街区进行划分, 通过分层语义认知 (HSC) 对 POI 数据进行特征提取, 并以此作为城市功能区的分类标准。Liu 等 (2017) 将遥感数据和社交媒体数据结合; Tu 等 (2018) 将遥感数据与手机定位数据结合; Hong 等 (2019) 基于 Infomap 社区检测算法对道路数据和 POI 数据进行结合。在上述研究中, 大多是先通过遥感影像等自然数据进行功能区划分再基于人文数据进行功能定性, 在城市核心街区识别结果较为出色, 但人文数据在城市边缘区域数据量少的问题仍然存在。

分析相关文献可见, 在功能区识别中单一数据无法详尽描述城市特征, 仅通过自然地理数据虽然可以清晰识别用地特征较为明显的区域, 如, 操场、工业厂房、传统住宅楼等, 但这些具体的设施与城市功能区是从属关系而非因果关系, 如一些大型工业园区中的工业厂房与员工宿舍楼同时存在, 但这些设施无疑都是

为城市的工业功能服务的，而不是独立存在的功能区。仅通过人文数据可以做到对城市功能区的精准识别，但对于数据量较少的城市边缘地区，大多无法准确判断。将城市地理信息与人类活动数据结合后构建的多源数据集，能够更加全面和精准体现功能区的特征。基于机器学习算法和聚类方法进行功能区的划分一方面解决早期人工划分时的主观性，同时还能大幅度节省人力资源。

（2）空间可达性测度

可达性（accessibility）这一概念最早由 Hansen(1959)提出，旨在描述点到点的空间上的通达性。随着 GIS 技术的进步以及规划学科的发展，可达性的概念以及其测度方法也在发生着变化。

早期对于可达性的研究碍于地理测绘学和 GIS 技术的限制，多为简单且宏观的平均法。Stegman（1969）和 Ihlanfeldt 等（1990）都是基于住宅区到工作场所的平均通勤时间、平均距离等对可达性进行测度。Ingram(1971)将平均法中的物理距离抽象为空间阻隔，提出基于空间阻隔最小的可达性测度方法。Linneker 等(1992)使用基于重力模型的距离衰减公式对 M25 伦敦轨道高速公路的可达性进行测度。随着 GIS 软件的不断发展，也催生出服务区分析（WA O'Neill 等，1992）、路网可达性（Gutiérrez J 等，1996）等相关方法。

此外，Fujita 等(1982)和 Heikkila 等(1992)将可达性定义为一种“区位潜力”，并由其设施所在周边地区的累计机会进行测度，由此催生了基于机会累计假设的可达性评价方法，如移动搜索法（FCA）系列算法。Peng (1997)使用移动搜索法对通勤的可达性进行测算；Luo 等(2003)首次提出两步移动搜索法（2SFCA）并基于此测度了芝加哥地区初级卫生保健站的空间可达性；Wan 等（2012）首次提出三步移动搜索法（3SFCA）。由于 FCA 系列方法引入了需求这一要素，并且操作简单方便，因而当前被广泛的使用于各类设施的空间公平性评价上。

基于潜能模型的可达性评价方法是由万有引力定律延伸而来的，也是基于机会累计假设下的一种可达性测度方法。Geertman 等（1995）基于 GIS 系统对潜能模型进行实现；Salze 等（2011）基于改进的潜能模型，对城市内上班族的设施可达性进行测度。由于潜能模型假设对于出行者而言，同类别下的全部设施对其都具有一定的吸引力，且可用出行距离、设施规模、个人需求等相关因素进行定义，又具有很强的包容性，也成为当前使用最为广泛的可达性测度方法之一。

（3）公共设施布局优化

公共设施布局优化相关研究最早可以追溯至 1909 年 (Brandeau Chiu, 1989), Alfred Weber 为通过选址解决单个仓库与多个顾客间总距离最小的目的提出 Weber 问题, 通过构建空间距离模型寻找最优位置。早期的设施布局优化问题与 Weber 问题相类似, 大都是基于选址的角度, 通过构建最优化模型解决问题。为确定消防站等应急服务公共设施的最优选址 Roth (1969) 和 Toregas (1971) 提出集覆盖问题模型, 旨在解决满足覆盖需求的同时实现最小成本。在此基础上 Church (1974) 提出最大覆盖问题以通过优化选址实现设施服务面积的最大化。Hakimi (1964) 为了设施空间距离上的公平性, 构建中位问题模型以实现需求点到分配设施的平均距离最小。

随着计算机科学的不断进步和 GIS (Geographic Information System 地理信息系统) 技术的发展, 对公共设施布局优化的研究开始转向空间可达性视角下, 通过 GIS 技术进行相关问题的研究。Luo 等 (2003) 基于 GIS 技术对芝加哥的医疗设施分布情况进行可达性测度提出优化建议; Coulter (2008) 通过对城市公共设施进行测度, 阐释了公共设施可达性分布不公平的现象, 并提出解决的方法。Tompkins (2010) 的《Facilities planning》一书则对医疗设施规划布局进行详细的介绍。

为实现更加细致和人性化的服务, 大多针对细分人群和具体设施进行。Pourrezaie-Khaligh (2022) 提出一种修复和优化方法来优化选址, 以解决医疗设施空间分布公平性的问题。Amaya 等 (2022) 针对老年人进行街区步行可达性的测度和评价分析, 并提出改进建议。El-Ramady (2022) 对传染病医疗废物回收站点进行规划; Liu 等 (2022) 在新冠疫情视角下基于 SEIR 模型对急救医疗设施进行布局优化。

1.2.2 国内研究现状

（1）城市功能区识别

国内关于城市功能区的相关研究与国际主流基本一致, 不少学者也是基于机器学习挖掘遥感数据展开研究 (谷岩岩等, 2018; 王皓弘, 2021)。对 POI 数据构建评价体系后, 根据划分区域内的 POI 数据密度的大小进行功能区的判别, 也

是国内主要研究方向。

池娇等（2016）在对 POI 数据进行加权重分类后，对划分的每个功能区单元内的 POI 数据构建指标频数密度和类型比例，并依照其大小将功能区划分为单一功能区、混合功能区、无数据区，以此实现功能区的识别。康雨豪等（2018）在基于路网和街区进行功能区划分的基础上，将计算 POI 密度的过程抽象为 POI 点样方密度法。贾斐雪等（2021）引入面积因素更为科学的构建了 POI 数据的赋分指标体系，并对结果通过误差矩阵进行精度分析。周杭等（2022）通过特征向量分析法取代了 POI 点样方密度法，并基于重力模型和 K-means 聚类进一步实现城市功能区的热点识别。

由查阅的相关研究可以见，通过计算 POI 密度可以较为准确的反映出一个区域的功能属性，同时基于面积、印象等规则对 POI 数据进行赋权后，可减少因为 POI 数据采集方式引起的数据重复和大众偏好。

（2）空间可达性测度

国内学者对于可达性测度方法的研究基本都是在方法的改进上。杨育军（2004）将道路密度作为交通可达性的评价因子；上海交通所提出并命名的基于平均出行时耗的上海交通所法（陈声洪 1990）等，是对于早期平均法的改进和应用。之后周小平（2007）基于区位配置模型（LA 模型），以及最小距离和最大覆盖模型进行城市内医院的布局优化。

近年来，国内学者对于可达性测度方法的研究大多都是针对 FCA 系列算法和潜能模型的改进。田玲玲（2019）在对 2SFCA 进行改进后，对仙桃市的医疗设施进行可达性分析，并基于实现最大公平和效率的假设下设计了设施的重新选址和规模限制；张金萍（2022）基于改进的 3SFCA 算法测算并构建了乡村地区公共医疗设施可达性与空间公平性评估方法体系；胡舒云（2021）利用改进潜力模型对深圳市的医疗设施进行可达性测算，并基于多源大数据对社区进行了曼—惠特尼 U 检验，定义了城市中社区间的资源剥夺现象等。

随着 GIS 技术和分析软件的不断发展，可达性的定义和测度方法也在不断变化。当前国内外使用较多的可达性测度方法是 2SFCA 算法和潜能模型，这得益于二者较强的包容性，一方面能够结合地理信息进行空间可达性测度，另一方面又能根据设施信息、人口数据等进行非空间可达性评价。机器学习方法的引入

一方面为城市规划的研究提供便利工具、减少了时间成本，也降低了相关评价模型的主观性。

(3) 公共设施布局优化

早期国内学者对于公共设施布局的研究与外国学者相似，也是由设施选址最优化问题逐渐转向基于 GIS 技术的可达性测度上。王非等（2006）就离散设施选址问题，对外国学者的相关研究进行了综述和分类。万波（2012），针对公共服务设施选址问题进行文献研究后，通过重力模型基于 GIS 技术对以武汉市为例进行研究。

这一时期也与我国城市高速发展时间基本一致，周小平（2007）以实现当地居民道路可达性最佳为准则，将区位配置模型与 GIS 技术相结合，通过问卷调查了解当地的就医问题，对天门市主城区的医疗设施进行了空间优化研究。吴建军（2008）则根据改进的引力指标，引入人口密度构建了医院可达性模型，从医疗服务的供应和需求两个方面评估了兰考县医疗服务的状况。于珊珊（2012）基于 GIS 理论并结合两步移动搜寻法对长沙市医院的布局和优化进行了研究。此阶段对于公共设施的布局策略仍处在“保量”阶段，旨在通过新建设施满足日益增长的城市居民的医疗需求。

此后，刘璇（2017）将可达性与公平性分析结合起来通过构建泰森多边形模型分析了武汉市的部分居民点的医疗设施可达性；李漱洋（2021）在健康韧性视角下对社区医疗设施进行了布局分析，等。由此标志着我国对于公共设施的布局策略到了“保质”的阶段。

在经历过城市高速扩张期的今天，城市框架基本稳定，进行大批量横刀阔斧的规划已不现实。因而，国内外对于医疗设施规划的研究大都针对细分领域和特定人群展开，并且对城市医疗设施的公平性和服务能力提出了进一步的要求。

在大数据收集存储更方便的条件下，基于多源数据进行城市功能区的空间特征识别方法可以综合已有方法，挖掘更多更精准的空间特征。同时，利用机器学习算法和统计聚类方法，可以对多源数据进行融合和特征提取，可以极大地减少城市功能区空间特征识别过程中的主观性因素。

1.3 主要研究内容

1.3.1 主要研究内容

论文的研究内容主要包括三部分。

(1) 基于 Mask R-CNN 算法和样方密度法的城市功能区识别研究

基于自然特征与人文特征两个维度，以兰州市主城区为例将 POI 数据与高分遥感影像数据相结合，分别通过 POI 样方密度法和 Mask R-CNN 算法挖掘城市的人文特征和自然特征，并以此为基础综合判断城市不同区域的所属功能。

(2) 基于改进两步移动搜索法的医疗设施空间可达性评价

基于人口栅格数据和医疗设施 POI 数据，通过改进的两步移动搜索法对兰州市主城区医疗设施的空间可达性进行测度，为进一步的设施规划布局打好基础。

(3) 基于随机森林方法的医疗设施选址推荐

利用随机森林合成预测的优势，通过挖掘城市 POI 数据中城市医疗设施分布的空间特征进行选址推荐，并综合上述研究结果设计兰州市主城区医疗设施布局优化方案。

综合以上研究内容，研究思路与技术路线图 1.1 表示。

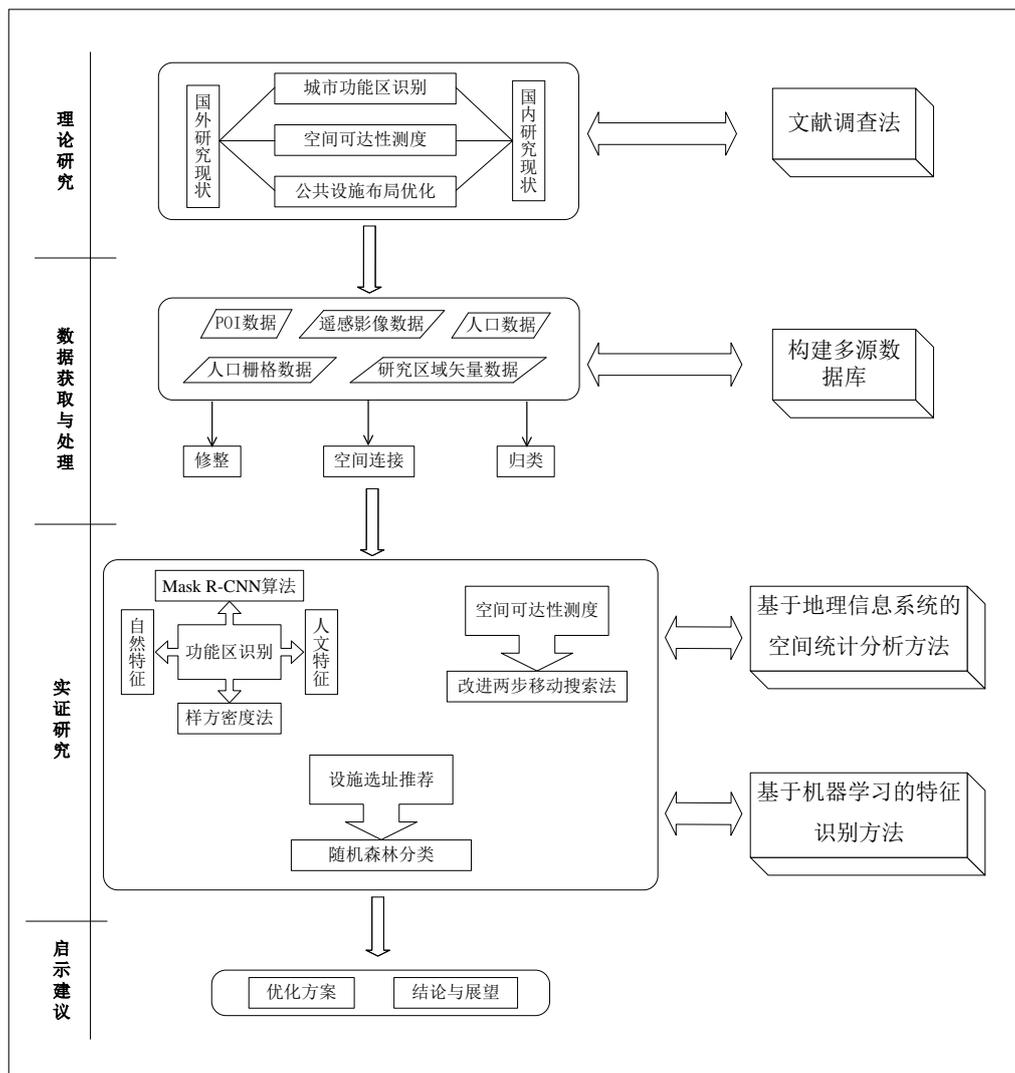


图 1.1 研究思路与技术路线

1.3.2 重点解决的问题

尝试解决城市高速扩张背景下，医疗设施分别不均匀、城市发展不充分的问题，并提出布局优化方案。通过融合城市地理数据、人口数据、道路数据、POI数据、卫星高分遥感数据等，构建城市多源数据集，将城市内的人文因素和地理因素结合再进行城市特征的挖掘，解决了此前数据来源单一导致的特征描述不完全、不详尽的问题。

将功能区识别结果、可达性测度结果、随机森林选址推荐结果结合起来对城市医疗设施的分布进行优化，一方面减少了仅从可达性测度进行评价时的主观性，另一方面亦保证了优化方案的公平性。

基于多源数据的城市空间特征识别与医疗设施布局优化，旨在对城市框架基本成熟的今天，提供更加精准有效的设施布局优化方案，以更好的满足市民的医疗需求。

1.4 主要研究方法

（1）文献调查法

已有相关研究文献是研究基础，全面细致的文献研究可以了解当前相关研究的前沿理论和成果，掌握研究主题的进展和不足，为论文研究提供方向指引和方法指导。

（2）基于地理信息系统的空间统计分析方法

通过构建地理信息系统，将栅格数据、矢量数据、文本数据在二位空间坐标上相统一，并基于定量分析方法，计算距离、面积、需求量、设施密度等相关参数，以此展开相关分析，探寻城市设施分布的空间特征，是论文研究的主要内容。

（3）基于机器学习的特征识别方法

机器学习（Machine Learning）方法，即通过构建算法模型，通过挖掘数据、训练模型的方式，以实现从数据中自动分析获得规律，并利用规律对未知数据进行预测的方法。本研究中主要使用了随机森林算法和卷积神经网络算法，用于提取遥感数据和 POI 数据内的城市特征，以此进行设施的布局优化。

1.5 论文组织结构

首先从城市功能区识别出发，初步宏观挖掘城市功能分布情况，再通过可达性测度了解医疗设施的空间分布情况，然后通过随机森林选址推荐形成初步的设施优化布局方案，最后综合三个角度的空间识别结果形成布局优化方案。主要安排如下。

(1) 绪论部分主要概况了当前三种空间特征识别领域的研究背景以及研究意义，并介绍了近几年国内外学者的研究现状，提出研究内容和组织框架。

(2) 研究数据来源中主要介绍了研究使用的多源数据的获取方式与具体来源，以及数据的相关参数。

(3) 基于 Mask R-CNN 算法和样方密度法的城市功能区识别研究中主要从自然和人文两个维度进行了城市功能区的识别，得到了兰州市主城区各功能区的空间分布情况。

(4) 基于改进两步移动搜索法的医疗设施空间可达性评价部分，通过人口栅格和医疗设施 POI 数据，基于高斯衰减函数，对兰州市医疗设施的空间可达性进行了测度，得到了医疗设施空间可达性的分布情况。

(5) 基于随机森林方法的医疗设施选址推荐，通过随机森林学习 POI 数据的空间分布特征，形成预测模型，得到医疗设施选址推荐方案。

(6) 基于多源数据的医疗设施布局优化方案是综合上述研究得到的三类空间识别结果，在分四类情况综合分析后得到的兰州市主城区医疗设施优化方案。

结论与展望部分对论文的研究进行了总结，并对现有研究的不足提出了进一步的探索方向。

2. 研究数据来源

2.1 研究区域

兰州市位于北纬 $36^{\circ}03'$ ，东经 $103^{\circ}40'$ ，现辖城关、七里河、西固、安宁、红古 5 区和永登、榆中、皋兰 3 县。为更好地研究城市内的医疗设施分布情况，选取城关区、七里河区、西固区和安宁区四个主城区作为研究区域，如图 2.1 所示。研究区的行政边界矢量数据来自全国地理信息资源目录服务系统中下载、分割得到，数据的最小行政区域单位是街道。

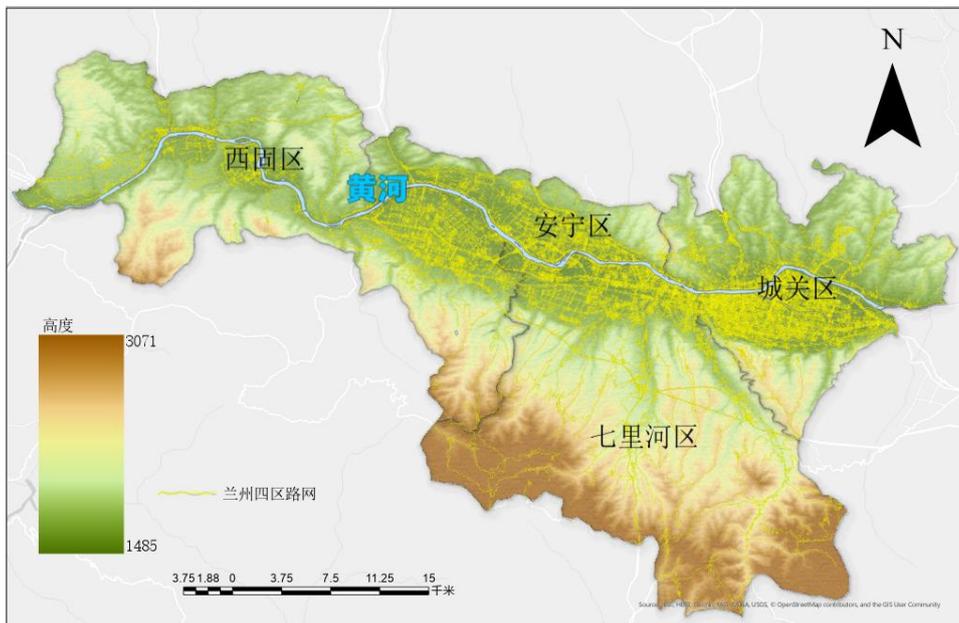


图 2.1 兰州市主城区高程图

由图 2.1 可以看出，兰州市主城区地理形态整体呈现依山傍水的态势，黄河自东向西穿城而过，呈现出南北两山夹一河的山谷地形，市区东西较为狭长，约 30 千米，南北较窄，最窄处仅为 5 千米左右，具有带状盆地城市的特征。

兰州市主城区依河而建，七里河区和西固区分别位于整个主城区南部和西部，路网密度相对较低，城关区地处研究区的东北部，地区道路密度高，经济相对较发达，是兰州市的政治、经济、文化中心。

2.2 遥感影像数据

遥感影像数据是基于 BIGEMAP 软件 (<http://www.bigemap.com/>) 经过行政

区域矢量数据裁剪后，进行爬取下载得到的。数据的地理坐标系是 WGS 1984，数据的行列栅格数分别为 5825 和 2883 个。

2.3 POI 数据

通过调用高德地图 API，使用 Python 以及后羿爬取器，基于爬虫技术得到了 2022 年 1 月兰州市城关区、七里河区、西固区、安宁区四个区内的全类别的 POI 数据，总共包含 125753 条，24 个大类，具体情况如表 2.1 所示。

表 2.1 兰州市主城区各类 POI 数据

一级类别	二级类别	POI 数量
餐饮服务	各类餐馆、甜品店、饮品店等	18987
道路附属设施	高速服务站、收费站等	11
地名地址信息	地名地址、交通地名、门牌信息等地图标注信息	14506
风景名胜	各类景点、公园、广场以及其周边相关设施	508
公共设施	报刊亭、电话亭、公共厕所等公共服务设施	932
公司企业	工厂、公司企业办公场所、农林渔牧基地	6801
购物服务	便民超市、各类大型商场、商业街等	29337
交通设施服务	地铁站、港口码头、公交车站、火车站等	4608
金融保险服务	保险公司、财务公司、银行、ATM 机等	1650
科教文化服务	博物馆、科研机构、培训机构、学校等	4510
摩托车服务	摩托车服务、维修、销售等	193
汽车服务	充电桩、加油站、汽车保养等	2581
汽车维修	各类品牌汽车维修点	1131
汽车销售	各类品牌汽车销售店	520
商务住宅	产业园区、楼宇、住宅区等	3542
生活服务	运营商营业厅、邮局、快递站、电力营业厅等	15455
室内设施	楼梯、电梯、扶梯等室内标识	787
体育休闲服务	疗养场所、电影院、运动场馆等休闲娱乐场所	2392
通行设施	建筑物门、交通通行设施的地图标注	5414
医疗保健服务	动物医疗场所、药房、诊所、综合医院等	4113
政府机构及社会团体	工商税务机构、公检法机构、政府机关等	4274
住宿服务	宾馆、大型酒店、招待所等	3500
总计		125752

在表 2.1 中的每条 POI 数据中，均包含此 POI 点基于 WGS1948 坐标系下的经、纬度坐标、点位名称以及具体地址等相关信息，且每条 POI 数据基于高德软

件公司的分类标准，按照大、中、小三个类别对其进行了归类。以 POI 数据“乌兰餐厅”为例，依照三级分类体系，依次归类为一级“餐饮服务”、二级“中餐厅”和三级“清真菜馆”。

2.4 人口栅格数据

来源于WorldPop网站 (<https://hub.worldpop.org/>) 下载得到的中国2020人口栅格数据，栅格分辨率为1000米，地理坐标系是WGS1984。

2.5 兰州市第七次人口普查数据

数据来自于 2021 年兰州市统计局发布的《兰州市第七次人口普查公报》(兰州市统计局, 2021)，选取城关区、安宁区、七里河区、西固区 62 个街道的人口数据。各区行政区划面积和人口数量，如表 2.2 所示。

表 2.2 兰州市主城区各区人口总量

区县	面积 (平方千米)	人口 (人)	人口密度 (人/平方千米)
城关区	220	1278745	5812.477273
七里河区	397	561020	1413.148615
西固区	385	364050	945.5844156
安宁区	86	288510	3354.767442

由表 2.2 可以看出，兰州市城关区人口密度最大，为 5812.477273 人/平方千米，说明城关区人口居住密集；西固区的人口密度最小，为 945.5844156 人/平方千米，这与西固区距离城市中心较远的地理位置有关。

3. 基于 Mask R-CNN 算法和样方密度法的城市功能区识别

3.1 引言

在经济发展过程中，由于物质资源的聚集效应和产业集群效应，城市中逐步形成居民提供特定服务的各类功能聚集区域。清晰把握城市的空间特征，尤其是各功能区的空间分布情况以及功能类型，对城市管理者进行去量保质的高水平发展和事半功倍的高效率治理具有重要意义。

近年来随着机器学习方法的兴起，出现了一些将遥感影像数据、光谱数据等自然地理数据（简称自然数据）与机器学习方法结合的功能区识别研究，这类研究多使用神经网络算法对遥感高分影像及其衍生数据进行挖掘，提取识别城市的用地特征、建筑分布特征、水体绿地特征等自然特征，旨在通过用地类型的分类对城市的功能区进行识别和划分（Wen Zhou, 2020）。在这类研究中，虽然可以清晰识别如操场、工业厂房、传统住宅楼等用地特征较为明显的区域（Z. Zhang, 2016），但这些具体的设施与城市功能区是从属关系而非因果关系，如一些大型工业园区中的工业厂房与员工宿舍楼同时存在，这些设施无疑都是为城市的工业功能而服务的，不是独立存在的功能区。近年来，各类工业厂房改造新闻屡见不鲜，其新用途也是五花八门。因此基于建筑的外观特征和用地特征判定其用途存在一定的缺陷，仅从自然特征的角度进行城市功能区识别，结果势必会与现实情况存在差距。

同时，利用 POI 数据、移动轨迹数据等人文地理数据（简称人文数据），进行城市功能区的划分已成为当前城市功能区识别与归类研究的主要方法之一。国内学者在构建评分体系对 POI 数据进行赋分后，针对人类活动轨迹数据等人文数据进行城市特征挖掘，通过样方密度法计算类别密度对城市功能区进行识别（池娇，2016）。基于 POI 数据进行分类赋分虽然可以做到对城市功能区的精准识别，但对于 POI 数据量较少的城市边缘地区，大多无法准确判断。由于人文数据与人类活动息息相关的特质，这类数据大多出自城市人口密集区域。对于城市开发的新区和边缘区域，人文数据量较少，仅通过单一维度无法体现真实现状。

因此，将人文数据与自然数据相结合，试图将两个维度结合实现城市的功能空间特征上的精准挖掘，利用 Mask R-CNN 算法和样方密度法对城市功能区进

行更为准确的识别分类。

3.2 研究方法

3.2.1 Mask R-CNN 算法

Mask R-CNN（掩膜卷积神经网络）算法是 2018 年由 Facebook AI 研究所的 Kaiming He（2017）发布的，在对 R-CNN（卷积神经网络）改进后的一种针对实例分割而开发的卷积神经网络。当前已被广泛利用在目标检测、目标实例分割、目标关键点检测等相关领域中。

Mask R-CNN 基于 Faster R-CNN 将图像分解为像素级别，提取目标特征空间分布，实现多任务输出。将图像输入骨干网络 ResNet50 得到相应的特征图；区域生成网络（Region Proposal Network, RPN）在特征图上使用滑动窗口，将特征映射成一维特征向量，在特征向量的基础上通过两个全连接层输出类别概率和边界框坐标。ROI Align 层使用双线性插值法通过调整边界框坐标来减少 RPN 造成的偏差，得到校正后的特征点位置；使用非极大值抑制（NMS）算法剔除重叠检测框，得到与真实检测框最接近的候选框。使用全连接层将特征向量映射成一维特征向量，得到预测结果。神经网络的具体搭建过程如算法 3.1 所示。

算法 3.1 Mask R-CNN

输入：遥感影像，感兴趣区域 ROIs，预训练的卷积神经网络模型 M

输出：每个 ROIs 的物体类别和掩码

- step1: 使用卷积神经网络 M 提取图像 I 的特征图
 - step2: 使用 Region Proposal Network (RPN) 提取感兴趣区域 ROIs
 - step3: 对于每个 ROI，计算其物体类别概率和边界框坐标
 - step4: 对于每个 ROI，使用 Mask 分支计算其掩码
 - step5: 使用非极大值抑制 (NMS) 方法过滤掉重叠的边界框
 - step6: 对于每个检测结果，将其掩码与图像进行融合，得到物体掩码
-

Mask R-CNN 在 Faster R-CNN 的基础上，通过添加每个感兴趣区域 (ROI) 上的掩膜分支，进而实现并行处理分类以及边界的回归，缩短检测时间的同时提高检测结果的准确度。基于遥感数据的建筑物识别问题存在精度要求高、数据量

大的问题，使用 Mask R-CNN 在目标检测和分割任务上具有较高的精度，可以识别建筑物的边缘和轮廓，生成高精度的建筑物掩码，有助于更精确地定位和测量建筑物。同时 Mask R-CNN 不依赖于人工标注的特征，可以自适应地学习目标特征表示，因此适用于各种类型、形状和大小的建筑物检测。此外，Mask R-CNN 可以生成高质量的建筑物掩膜，可用于可视化建筑物分布、密度等信息，有助于城市规划的进一步研究。

陈红顺等（2023）定义 Mask R-CNN 为一个多任务损失，即

$$L = L_{cls} + L_{box} + L_{mask} \quad (3.1)$$

其中， L_{cls} 为分类损失函数，用来区分前景和背景； L_{box} 为检测损失函数，用来修正前景锚点坐标，从而得到最佳检测框； L_{mask} 为掩膜分割结果损失，是平均二值交叉熵函数，利用 sigmoid 函数对每个像素进行特征映射，为不同类别 ROI 提供独立掩码预测。

3.2.2 POI 点样方密度法

不同类型的 POI 数据能反映城市功能分区的规律和信息。然而，POI 数据是点数据，不能表现其规模大小信息。若仅从 POI 数据频数的角度识别城市功能区，会使得区域主要功能被占地面积小且数量多的便利店、小商铺等所主导，导致计算结果与真实情况不符。因此，参考贾斐雪等（2021）面积赋分的思路，与样方密度法相结合，对 POI 数据赋值。

表 3.1 POI 面积与评分标准

面积评分	POI 类型
0-100 平方米（1 分）	便利店/小卖部、餐饮店、美容美发、快递点等日常生活服务设施
100-1000 平方米（10 分）	专卖店、各营业厅、宾馆、招待所、小型企业和社区服务中心等
1000-3000 平方米（30 分）	专科医院、度假村、小型商业街、综合市场、公共娱乐设施等
3000-5000 平方米（50 分）	办公场所、博物馆、图书馆、美术馆、艺术馆、住宅区、综合医院等
≥5000 平方米（100 分）	高级酒店、工厂园区、养殖基地、风景名胜、广场公园等

考虑每个 POI 点的规模对功能区影响程度不同, 参考我国的《零售业态分类》(中华人民共和国商务部, 2021) 对占地面积 1000 平方米以下的相关设施进行分类, 并依照《中国城市建设统计年鉴》对 1000 平方米以上的设施进行归类, 具体评价标准如表 3.1 所示。

依照占地面积对研究区各类设施进行综合考察, 将各类 POI 点进行面积赋分, 记 λ_i 为表示第 i 类 POI 的面积值。赋分后通过样方密度法判断区域的功能区类型。

样方密度法是指将研究区分割成一系列子区域, 即样方, 计算落入各个样方的 POI 点数与样方面积的比值。由于使用 POI 点频数计算样方密度会与真实情况有出入, 使用面积赋分代替频数, 即, 根据 POI 面积赋分结果计算各单元不同类型 POI 面积与网格单元面积的比值, 得出网格内占主导地位的 POI 类型, 其计算公式为:

$$P_i^l = \frac{\sum_{i=1}^k \lambda_i^l}{S}, \quad (3.2)$$

$$X_i^l = \frac{P_i^l}{\sum_{i=1}^n P_i^l}, \quad (3.3)$$

其中, l 表示第 l 个网格; P_i^l 表示第 l 个网格内第 i 类 POI 的密度分数; k 表示网格中 POI 数量; S 表示第 l 个网格面积; X_i^l 表示第 l 个网格内第 i 类 POI 密度占单元内所有类型 POI 密度的比例。

3.2.3 基于 Mask R-CNN 和样方密度法的城市功能区识别方法

Mask R-CNN 算法适用于对地观测的遥感影像数据进行建筑物检测识别。建筑是人类劳动的产物, 未检测到建筑物的区域大多数为山地、湖泊等不适宜人类居住的自然区域。因此, 将网格中是否检测到建筑物作为判断城市自然特征依据, 检测到建筑物的网格, 即不符合自然特征, 否则符合自然特征。样方密度法适用于反映人类活动的 POI 数据进行功能区识别。POI 数据是城市人文特征的直接体现, 因此将网格中是否存在 POI 数据作为判断城市人文特征依据, 存在 POI 数

据的网格，即设定有人文特征，否则不具备人文特征。

基于遥感影像提取的自然特征，仅表示城市外部自然特征，无法反映城市社会特征；POI 属于人类行为数据，其分布存在空间异质性，难以感知城市自然特征。因此，基于 Mask R-CNN 识别自然特征结果与 POI 数据分类结果进行相互补充校验得到最终结果。参考池娇等（2016）的研究，确定单一、混合功能区类型规则如表 3.2 所示。

表 3.2 功能区分类

功能区类型	功能区类别
单一功能区	商务区、风景区、公共区、工业区、居住区等
混合功能区	商务混合区、风景混合区、公共混合区、工业混合区、居住混合区等
无功能区	无数据区

(1) 当 $X_{MAX}^l = X_i^l \geq 0.5$ ，即当网格内某一类 POI 比例 $\geq 50\%$ 时，确定该网格为单一功能区，功能区性质由该网格 POI 类型确定。例如，商业设施服务类判读为商务区，风景名胜类判读为风景区，公共管理与企业设施类判读为公共区，工业类判读为工业区，居住类判读为居住区。

(2) 当 $X_{MAX}^l = X_i^l < 0.5$ ，即当网格内某一类 POI 比例 $< 50\%$ 时，确定该区域为混合功能区，混合功能区性质取决于网格内占比最大的 POI 类型。例如以商业设施服务类为主的混合区判读为商务混合区，以公共管理与企业设施类为主的混合区判读为公共混合区，以工业类为主的混合区判读为工业混合区，以居住类为主的混合区判读为居住混合区。

3.3 实证分析

3.3.1 基于 POI 点样方密度法的功能区识别

将研究区划分为 3883 个 $500m \times 500m$ 的网格后，通过 GIS 相交功能，确定研究区内重分类后 5 类 POI 数据所属网格，对存在 POI 数据的 1290 个网格，根据式(3.2)(3.3)及功能区判别规则进行功能区识别，识别结果及分布图如表 3.3、图 3.1 所示。

表 3.3 POI 点样方密度法功能区识别结果

功能	风景区	工业区	公共区	商务区	居住区	混合区	无功能区	总计
个数	85	63	320	304	69	449	2593	3883

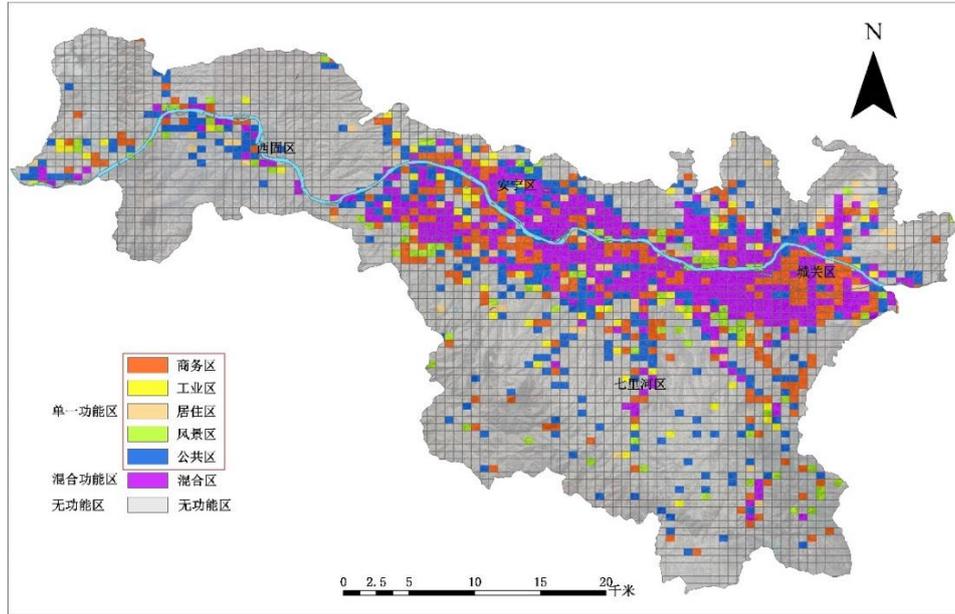


图 3.1 POI 点样方密度法功能区识别结果分布图

基于样方密度法功能区识别结果显示,单一功能区有 841 个网格,占 21.65%,其中公共区数量最多,大多分布在中心城区的周边,部分散落在乡镇与城郊,如七里河区南部和西固区中部,而公共区是以企业、政府等办公场所的定义相差较多。这主要是由于在城市边远地区 POI 数据较少引起的偏差。混合功能区共 449 个网格,占 11.56%,主要分布在城关区、七里河区和安宁区等繁华区域,其基础设施完善且生活便利度高,与混合功能区性质符合。无功能区共计 2593 个网格,占 66.78%,大多分布在研究区的边缘地区和南部山区以及河流附近。

通过分析发现基于 POI 点样方密度法的功能区识别结果在城市核心区域识别效果较好,而在城镇、村落等边缘区域识别结果与现实存在一定偏差。

3.3.2 自然特征与人文特征提取

基于遥感影像和 POI 数据使用 Mask R-CNN 算法和空间统计方法分别提取自然特征与人文特征,将研究区网格分为 4 类,进行校验后再使用样方密度法进行功能区识别。

遥感影像数据用于识别和提取建筑物，选取兰州市主城区用于训练和测试。将原始的遥感影像裁剪为 $5000\text{m}\times 5000\text{m}$ 固定大小的 117 个遥感栅格数据。为尽量降低分割遥感数据带来的精度损失，选择使用栅格三次卷积插值法进行分割。在 ArcGIS Pro 软件中进行 Mask R-CNN 的训练，随机选取 70%数据作为训练集，即随机选取 82 个遥感栅格数据作为训练集，每次训练从训练集中随机选取 10%作为验证集，剩余 30%数据作为测试集，生成相应的建筑物掩膜图像，如图 3.2 所示。

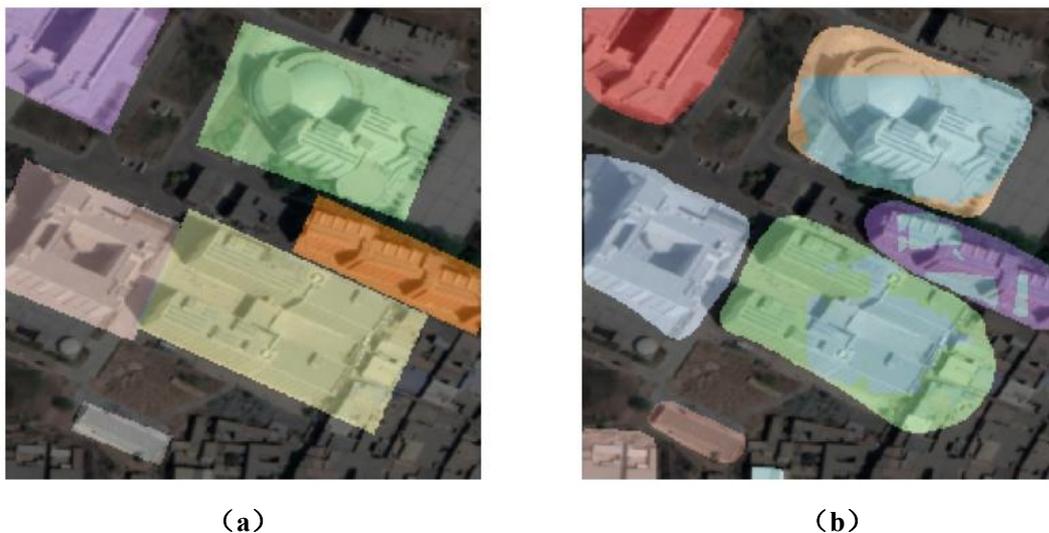


图 3.2 Mask R-CNN 提取遥感影像对应掩膜图像

图 3.2 (a) 是供深度学习训练使用的人工标注掩膜，图 3.2 (b) 是 Mask R-CNN 预测掩膜。由图 3.2 可知，经过训练后的 Mask R-CNN 模型可以在遥感影像中清晰识别建筑物并形成掩膜。对比可见，经过多轮识别后的结果与人工标注基本一致，且识别准度和精度都较高，同时还能识别出区域内人工标注遗漏的建筑。

模型的训练和测试在 NVIDIA 8.5.962 驱动版本。在 CUDA8.5 神经网络架构下，使用 NVIDIA 2060super 显卡进行训练。建筑物检测的训练和测试均在 GPU 加速下完成。

在 ArcGIS Pro 中进行 Mask R-CNN 的训练。导入 20 个制作好的训练集对神经网络模型进行训练，并根据硬件性能，将模型训练的最大纪元 (Epoch) 设定为 30，即默认循环训练 30 轮。训练时的切片大小 (Chip-size) 根据训练集所处的区域不同，与制作训练集时设定的分块大小保持一致，以实现最佳学习率；选

取 ResNet50 作为骨干网络模型进行特征提取；同时设定模型训练改进效果低于 0.01 时终止训练。

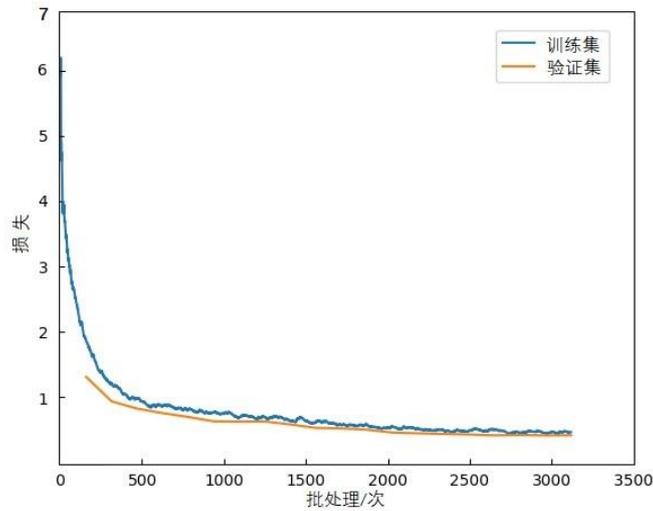


图 3.3 训练过程中损失变化曲线

训练过程中，模型的总损失变化如图 3.3 所示。由图 3.3 可以看出，当模型训练到 500 批次时损失达到较低水平。平均预测得分为 0.9019，模型训练结果较好，精度有了较大幅度的提升，可以对研究区内的建筑物进行检测和识别。

使用训练过的 Mask R-CNN 模型对研究区的遥感影像进行建筑物识别。为更好体现研究区内的功能区分布情况，方便后续研究，将研究区按照 $500m \times 500m$ 的面积分割成了 3883 个网格，通过 GIS 软件相交功能将建筑检测结果与网格进行关联，并通过几何计算功能测算不同建筑物的占地面积。建筑物识别结果如图 3.4 所示。

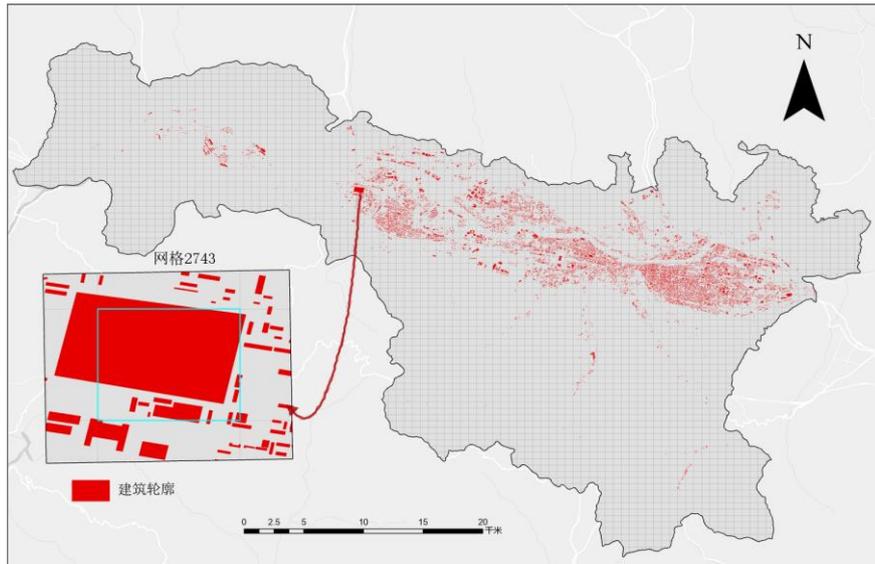


图 3.4 Mask R-CNN 模型检测建筑物结果

在 3883 个网格中，有 1108 个检测到建筑物，2743 号网格所在位置在兰州市郊区的工业园区内，厂房密集，建筑物面积最大，达到了约 272239 平方米；而 2646 号网格中建筑面积最小，仅为 0.03 平方米。按照城市自然特征判断的依据，将研究区中 2775 个未检测到建筑物的网格，认为其符合自然特征，1108 个检测到建筑物的网格，认为其不符合自然特征。

通过 GIS 相交功能，确定重分类后 5 类 POI 数据所属网格。根据 POI 数据的性质，按照城市人文特征判断依据，将研究区中 1290 个存在 POI 数据的网格，认为其存在人文特征，2593 个不存在 POI 数据的网格，认为其不存在人文特征。

3.3.3 基于自然特征与人文特征校验的城市功能区识别结果

将自然特征和人文特征提取结果进行补充校验，可以将网格数据分为 4 类。

第 1 类：自然特征和人文特征均存在。即未检测到建筑物但存在 POI 点，共 328 个。此类网格的一种可能是 POI 点与检测到的建筑物间存在偏移，通过 POI 样方密度法判断其功能区类型；另一种可能是部分工厂设施和村镇建筑未被检测到，且 POI 数量较小，分布在城市边缘，将其定义为村落。

第 2 类：存在自然特征，不存在人文特征。即未检测到建筑物且无 POI 点，共 2447 个。此类网格大多分布在研究区的边缘地区和南部山区以及河流附近，根据功能区识别方法将其定义为无功能区。

第 3 类：不存在自然特征，但存在人文特征。即检测到建筑物且存在 POI 点，

共 962 个。对此类网格直接通过 POI 样方密度判断其所属功能区类型。

第 4 类：自然特征和人文特征均不存在，即检测到建筑物但无 POI 点，共 146 个。此类网格的一种可能是建筑物占地面积较大，POI 数据与建筑检测数据间产生了偏移，采用近邻匹配方法确定其功能区类型；另一种可能是处于山地、河流的边缘地区，将其判断为无功能区。

对经过分类后的网格，按照功能区判别标准及样方密度计算结果得到兰州市主城区功能区识别结果如表 3.4 所示。

表 3.4 可见，单一功能区中商业娱乐区分布网格最多，为 202 个。混合区中网格为 444 个。由于兰州市主城区行政区划内山地占地面积较大，使得城市功能区识别中无功能区覆盖网格最多，达到 2866 个。

表 3.4 各类功能区数量

	功能区类型	数量（个）
单一功能区	风景区	67
	工业区	60
	公共办公区	175
	生活居住区	69
	商业娱乐区	202
混合区		444
无功能区		2866
总计		3883

研究区域内各类功能区空间分布如图 3.5 所示。

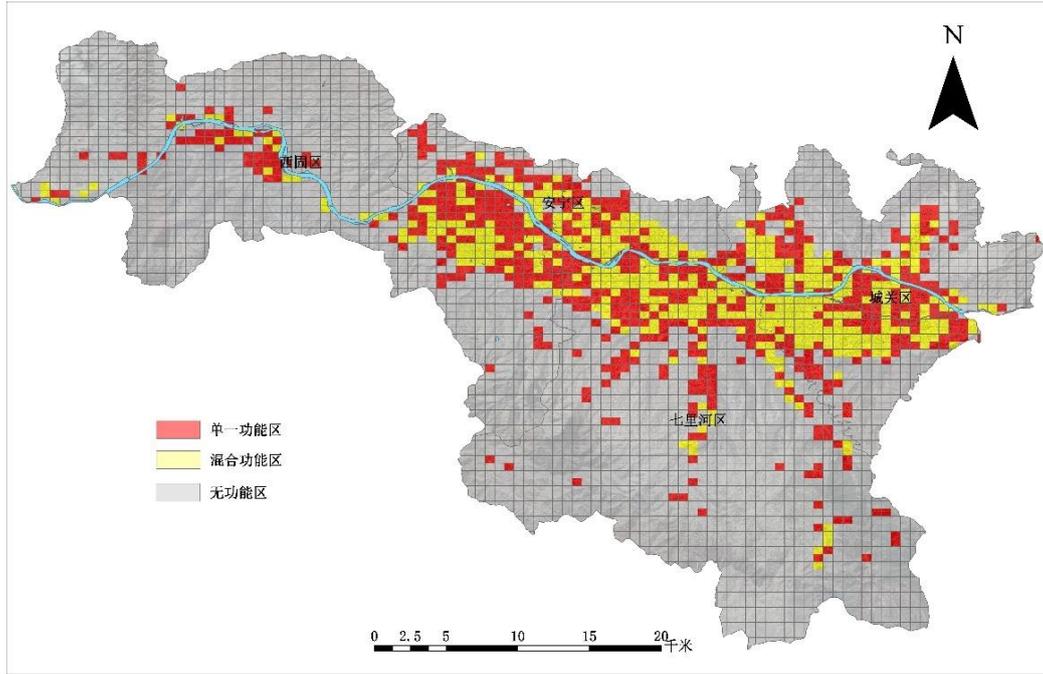


图 3.5 单一功能区、混合功能区和无功能区分布图

由图 3.5 可得，兰州市主城区三类功能区识别结果分布情况，单一功能区功能性质由网格单元中占主导地位的 POI 类别决定，共 573 个网格。

对单一功能区做进一步分析，其空间分布如图 3.6 所示。商务区共 202 个网格，其数量占比最多，主要集中在城关区核心街区和七里河区与安宁区的交界处，形成了兰州市主城区东西两个商业中心。东部城关区商业中心主要集中在西关什字等区域，西部商业中心则是呈现整体西扩的态势，主要为兰州中心、金牛街商业广场等区域。

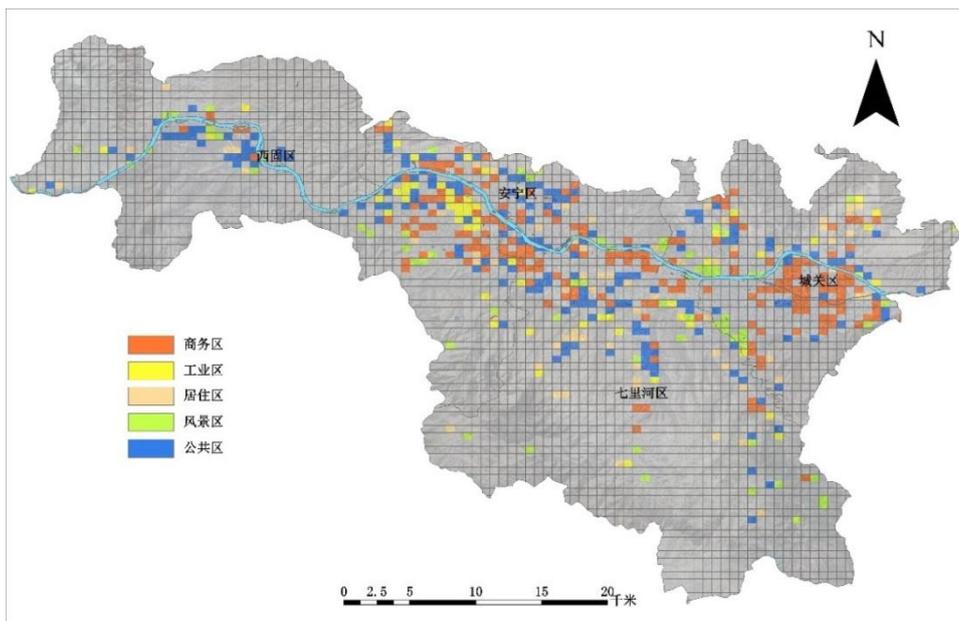


图 3.6 单一功能区分布图

图 3.6 显示，公共区共 175 个网格，主要集中分布在城市周边，距离城市核心街区存在一定距离，一般是大型企业以及高校的所在地。工业区和风景区也主要分布在远离城市的郊区。其中，工业区共 60 个网格，主要集中在安宁区西部和西固区东部，是各类工业产业的聚集区，风景区共 67 个网格，分布在南部山区和北部城郊。居住区共 48 个网格，主要分布在城郊和研究区北部，当前城市核心区域内的小区 and 住宅大多临近办公设施和商业设施，因而大多归属于混合区内。村落共 21 个网格，主要分布在研究区南部的山区中，与村镇的分布情况大致相同。

混合功能区主要集中在城市核心区域城关区和七里河区，此类区域是城市老城区。混合区内建筑面积大且 POI 数量多，很难用单一功能进行定义。混合功能区共 444 个网格，其混合分布如图 3.7 所示。

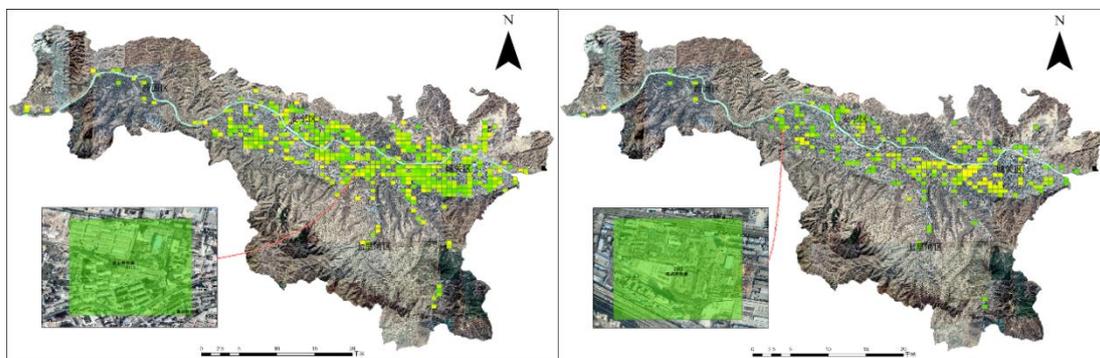


图 3.7 (a) 商务混合区

图 3.7 (b) 工业混合区

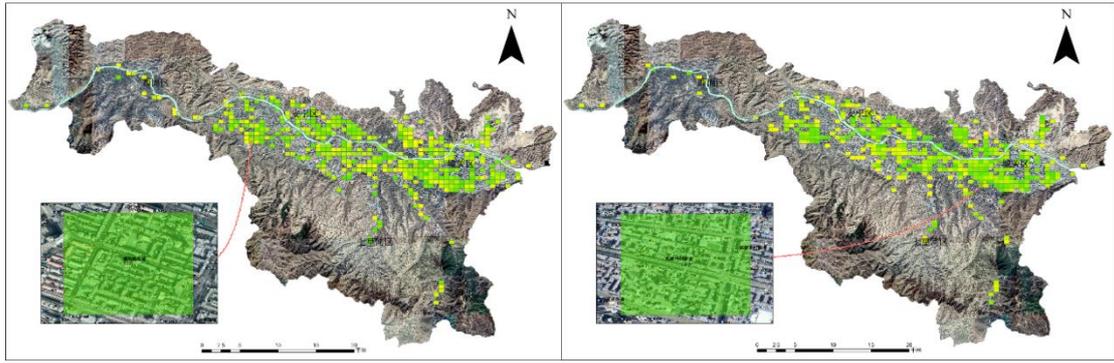


图 3.7 (c) 公共混合区

图 3.7 (d) 居住混合区

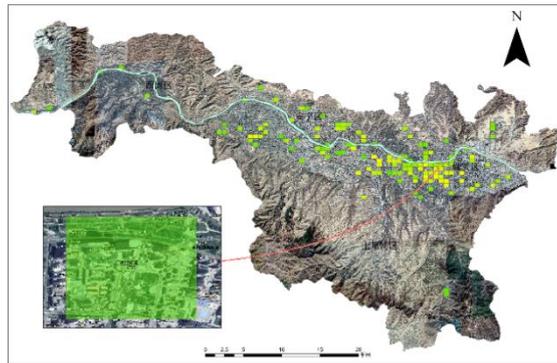


图 3.7 (e) 风景混合区

为分析混合功能区分布将商务区、工业区、公共区、居住区和风景区 5 类功能区识别结果运用信息熵，分析各类区域混合程度。颜色越深越绿的区域信息熵越高，如图 3.7 所示，商务混合区混合度较高的区域主要集中在七里河区龚家湾街道和晏家坪街道（图 3.7-a）；工业混合区混合度较高的区域主要集中在西固区临洮街街道（图 3.7-b）；公共混合区混合度较高的区域主要集中在西固区福利路街道（图 3.7-c）；居住混合区混合度较高的区域主要集中在城关区铁路西村街道（图 3.7-d）；风景混合区混合度较高的区域主要集中在城关区广武门街道（图 3.7-e）。

无功能区主要集中在分布在地、河流的边缘地区，体现了城市自然特征，共 2866 个网格。由于无功能区仅体现自然特征，不再对其进行分析。

3.3.4 精度评价

为验证城区功能区识别结果可靠性，通过构建混淆矩阵对识别结果进行精度验证。参考宋丽洁等（2023）精度评价指标总体精度和 Kappa 系数。其中，总体精度指随机样本分类结果与对应网格实际类型一致的概率。Kappa 系数是由混淆矩阵计算得到，取值范围为 $[0,1]$ ，Kappa 系数值越大，分类精度越高。具体计算

公式为：

$$\text{总体精度} = \frac{\sum_{i=1}^N x_{ii}}{M} \times 100\% , \quad (3.4)$$

$$Kappa = \frac{M \sum_{i=1}^N x_{ii} - \sum_{i=1}^N \left(\sum_{j=1}^N x_{ij} \sum_{j=1}^N x_{ji} \right)}{M^2 - \sum_{i=1}^N \left(\sum_{j=1}^N x_{ij} \sum_{j=1}^N x_{ji} \right)} , \quad (3.5)$$

其中， x_{ii} 表示误差矩阵中第 i 行第 i 列的元素； $\sum_{j=1}^N x_{ij}$ 表示第 i 行的和； $\sum_{j=1}^N x_{ji}$ 表示第 j 列的和； M 表示总样本数； N 表示行列数。

随机选取 20 个网格，依据高德地图数据选取真值，随机抽 5 次进行目视解译工作，构建混淆矩阵并计算总体精度和 Kappa 系数。5 次随机抽样分类精度计算结果如表 3.5 所示。

表 3.5 5 次随机抽样分类精度验证结果

精度验证	1 次抽样	2 次抽样	3 次抽样	4 次抽样	5 次抽样
总体精度	0.95	0.9	0.9	0.85	0.9
Kappa 系数	0.897	0.833	0.718	0.745	0.815

表 3.5 中计算得到 5 次随机抽样总体精度和 Kappa 系数，计算其平均值得到平均总体精度为 90%，平均 Kappa 为 0.802。根据评价精度标准，抽样后平均分类精度较高，可以认为利用遥感影像和 POI 数据，基于 Mask R-CNN 和样方密度法对城市功能区识别具有一定的合理性和可行性。

3.4 小结

(1) 首先通过 Mask R-CNN 算法进行建筑检测；其次利用 POI 样方密度法进行了功能区识别，将城市自然特征与人文特征相结合，较为详细和综合地对研究区下的功能区进行了识别研究。

功能区中数量最多的是混合区，有 444 个，一部分自然混合区多处是在黄河周边以及南部山区，大多是人类聚集区与自然环境的混合。另一份混合区处在城市的核心区域，这类混合区内功能复杂，建筑面积大，POI 数量多，很难用单一功

能进行定义，是城市的老城区，可以对此类区域进行针对性的规划和变迁，以扩大城市功能性的集群效应。

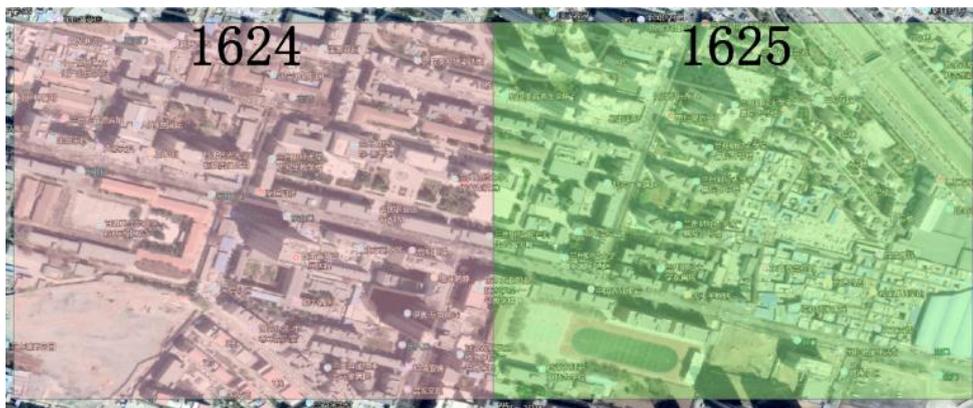


图 3.8 网格 1624 和 1625

如图 3.4 所示，以所属区域位于城关区的拱星墩街道的网格 1624 和 1625 为例，左图 1624 网格内设施种类繁多，既有学校、医院这类公共办公设施，也有像小区住宅这类的住宅设施，是城市功能复杂的混合区。右图 1625 网格覆盖了一个大型的农贸市场，虽然也有学校和住宅等设施存在，但其主要功能还是进行采购和消费的商业区，城市功能区识别情况与现实状况基本相符。

其次，商业娱乐区中大部分集中在城关区的核心街区，以及七里河区、安宁区的交界处，形成了兰州市主城区东西两个商业中心。东部的商业中心身处老城区，是城市长久发展的结果；西部的商业中心则是呈现整体西扩的态势。西固区也有本区内固定的商业中心，但面积和质量与其他区相比都存在一定差距。南部山区内的娱乐区大多为农家乐以及度假村养老院等休闲娱乐场所，也反映出兰州市民日常娱乐生活的倾向性。

排名第三的是公共办公区，大多分布在城市周边，距离城市核心街区存在一定距离，一般是大型企业以及高校的所在地。

工业区和风景区排第四和第五，大多分布在远离城市的市郊区。工业区主要集中在安宁区西部和西固区东部，是各类工业产业的聚集区。风景区则多分布在南部山区和北部城郊。

生活居住区有 48 个，大多分布在城郊和研究区北部。如图 3.5 所示，所属区域位于兰州市城关区青白石街道，位于研究区北部，网格 2703、2704、2614 覆盖的区域是两个新开的楼盘。当前城市核心区域内的小区 and 住宅大多临近办公设

施和商业设施，因而大多归属于混合区内。因此如图 5 所示的住宅区大多为新开的楼盘或城郊村镇交界处，此处设施数量和类型较少是下一步城市扩张的主要方向，也是相关设施选址的重要区域。网格 2615 也覆盖着部分的住宅区，但同时覆盖了部分的工业设施，因而被识别为混合区，与现实情况基本相符。



图 3.9 网格 2703、2704、2614、2615

数量最少的是城郊与村落，大多分布在研究区南部的山区中，与村镇的分布情况大致相同。相较于仅使用样方密度法单一维度进行功能区识别，在城市边缘区域以及设施混杂区域的识别精度也更高，结果显示城市功能识别情况与现实状况基本一致。

(2) 基于 Mask R-CNN 算法进行建筑物检测能够大幅降低人工识别标注的成本，同时可以做到对建筑物进行较为精准和快速的检测。

(3) 引入面积要素对 POI 数据赋分后再进行样方密度法的计算可以更好的反映各类设施的空间分布综合情况，从而更精准的体现所属区域的功能性。

(4) 本研究基于 Mask R-CNN 算法和 POI 样方密度法可以较好地区分单一功能区和混合功能区。特别对生活居住区而言，识别得到的仅提供单一生活居住功能的区域大多是人口数量较多，同时配套设施类型较少的区域，是城市扩张和相关配套产业发展的方向之一。

4. 基于改进两步移动搜索法的医疗设施空间可达性评价

4.1 引言

近年来经济高速发展，人均收入水平不断提高，居民对于医疗资源数量和质量的需求也水涨船高。城镇化推进使得各大中型城市不断对外扩张，为了突破旧城区土地供给对城市发展的局限性，越来越多居民区和商业区的规划建设已延伸至城市偏远地区。早期的城市规划基于人口、商业、交通等因素，医疗设施多集中在老旧城区，无法切合城市高速发展趋势下的医疗需求。从当前城市发展需求和城市更新规划布局的角度来看，城市边缘区域医疗设施相对不足，旧城区医疗设施配置则存在重复配置，医疗卫生资源分配不均衡成为城市发展中存在的普遍现象（国务院办公厅，2019）。

医疗设施作为公共设施的重要部分，一方面需要体现设施的公共属性，即尽量实现较大的服务面积以满足更多人的医疗需求。另一方面，医院较为集中的运营模式，使得不同等级医疗设施之间的服务供给能力差距较大。因此通过基于供给—需求角度展开的两步移动搜索法对医疗设施的空间可达性测度更加贴合医疗设施分布的内在逻辑，能够得出较为精准的结果。

两步移动搜索法（2SFCA）起源于重力模型（Joseph, 1982），并进一步受空间分解法的启发（Radke, 2000），旨在通过供求关系变动的角度对医疗服务设施的空间可达性进行评价（Luo, 2003）。近年来两步移动搜索法及其改进办法被广泛用于各领域设施的可达性测度研究中，如，针对城市绿地进行可达性评价（全德，2021）和养老设施的空间可达性测度（Liu L, 2022）。医疗设施作为一种特殊的供求关系，且其中的大部分需求会随着空间距离的增加而递减，因此两步移动搜索法也被广泛的适用于医疗设施的空间可达性测度中。为了进一步挖掘兰州市主城区内医疗设施的空间分布特征，并实现医疗设施的优化，基于第七次人口普查数据、人口栅格数据、医疗服务设施POI数据、街道矢量数据，选用改进两步移动搜索法，结合高斯衰减函数对兰州市主城区医疗设施的空间可达性进行研究。

4.2 两步移动搜索法

两步移动搜索方法的具体操作过程可以概况为进行两次搜索的过程,在同时考虑供给和需求两个因素下,较为全面、简便地对设施的空间可达性进行测度。

第一步:从供给端出发,搜索计算供需比 R_j ,体现了供应满足其服务范围内所有需求的能力,如式(4.1)所示。

$$R_j = \frac{S_j}{\sum_{k \in \{d_{kj} < d_0\}} P_k} \quad i=1,2,\dots,n; j=1,2,\dots,m \quad (4.1)$$

其中, j 为服务供给设施点, k 是需求单元点, d_{kj} 为二者间距离, d_0 是设施服务的极限距离, S_j 代表着设施的供给水平总量, P_k 代表着需求端的总量。即以供给为中心、服务极限距离为半径建立搜索域后汇总需求总量计算供需比 R_j 。

第二步:从需求端出发,搜索计算得出可达性指数 A_i ,如式(4.2)所示。

$$A_i = \sum_{j \in \{d_{ij} < d_0\}} R_j = \sum_{j \in \{d_{ij} < d_0\}} \frac{S_j}{\sum_{k \in \{d_{kj} < d_0\}} P_k} \quad k=1,2,\dots,o \quad (4.2)$$

其中 i 为任意需求点,即以需求为中心、极限距离为半径建立搜索域后,将域内供给设施的供求比加总求和得到基于距离关系的可达性水平 A_i 。

通过两步移动搜索,一方面对服务距离内的供给量进行了加总,另一方面也对需求端的人类活动空间进行了定义。

4.3 数据处理

4.3.1 人口栅格数据

为了进行更精确的可达性测算,基于第七次人口普查数据对人口分布栅格进行修正,制作兰州市主城区第七次人口普查人口栅格数据,以便进行两步移动搜索法中需求端的计算。其中第七次人口普查数据来源于兰州市第七次全国人口普查公报(兰州市统计局,2021),中国2020人口栅格数据来源于WorldPop。

以兰州市主城区63个街道为最小单位构建数表,第七次普查各街道人口数量记为 $G_i(i=1,2,\dots,63)$,并将人口数表、人口栅格数据、街道矢量数据导入ArcGIS

Pro中。通过区域分析工具，统计计算中国2020人口栅格数据中各街道的人口总数情况，记为 $W_i(i=1,2,\dots,63)$ ，由式(4.3)可得到街道人口修正系数 R_i 。

$$R_i = \frac{G_i}{W_i}, \quad i=1,2,\dots,63 \quad (4.3)$$

将人口修整系数与街道矢量图连接，转换为栅格后与中国2020人口栅格数据进行相乘，进而得到了兰州市主城区第七次人口普查人口栅格数据。

4.3.2 医疗服务设施数据

我国的医疗服务设施主要有以下四类：医院、基础医疗服务设施、特殊公共健康设施和其他设施。其中，医院分为三个等级，分别是三级医院、二级医院和一级医院。基础医疗系统包括城市和乡镇两个部分，乡镇系统由村卫生站和乡镇卫生院组成，城市系统由门诊诊所和社区卫生服务中心（站）组成。

医疗设施数据源于爬取得到的POI数据，对研究区内的“医疗保健服务”这一大类中的数据进行了修剪和降重，删去了动物医疗场所和医疗保健销售店等无法对居民提供医疗服务职能的POI数据，并对剩下的1221条数据按照《甘肃省医疗年鉴》进行了对比，将所有医疗设施分为三个等级。其中，一级包含大型的三甲医院和综合医院；二级为中小型的专科医院和大型的社区医疗服务站；三级为卫生院、诊所等小型的医疗服务设施。

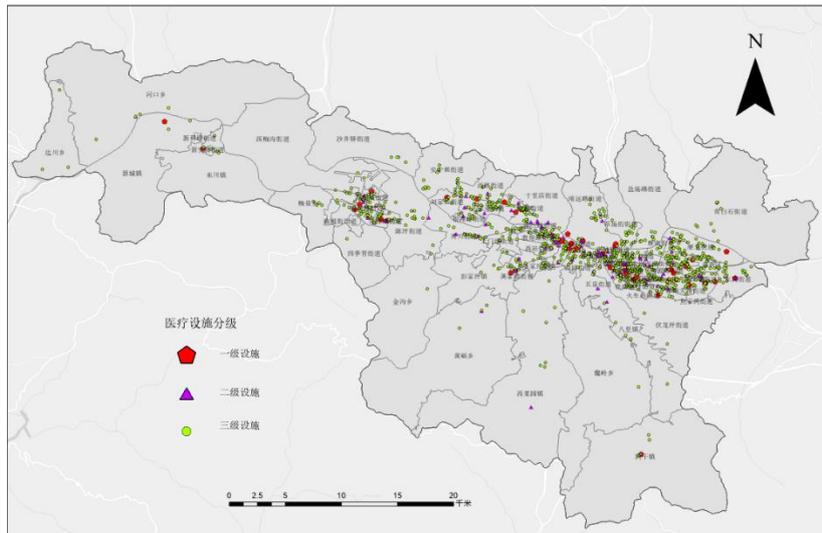


图 4.1 兰州市主城区医疗设施空间分布

AOI (Area of Interest, 兴趣面) 数据中包含了设施的面积数据, 但大多只收录了自成一栋的大型设施。由于三级医疗服务设施面积较小, 且大多为建筑物中的某一个店面, 在网络地图中爬取的面积数据并不准确, 因此在参照兰州市医疗设施面积和《中国城市建设统计年鉴》的基础上, 以POI数据为中心, 分别以10米、50米、100米为半径制作缓冲区来模拟三级医疗服务的服务面积。

4.3.3 空间投影与网格化

为了得到更精准的距离数据和面积数据, 尽可能的减少投影坐标系带来的精度偏差, 测算得到研究区的中央经度为东经103.660044, 并以WGS1984为基础, 设定中央经度为东经103.66, 构建了兰州市主城区投影坐标系。同时将研究区划定为距离为2000米的网格, 以便进行可达性测度。

4.4 基于改进两步移动搜索法的可达性测度

4.4.1 改进的两步移动搜索法

为了体现空间距离增加导致的需求量和供给能力的下降, 利用基于高斯混合函数的改进两步移动搜索法。两步移动搜索法中, 需求量和供给量与距离变化无关, 这与现实情况不尽相同。现实生活中, 随着空间距离的增大, 需求量往往会逐渐减小, 服务质量也会随之降低。最直接的例子就是外卖服务, 往往越远距离的外卖订单数越少, 送达时顾客的用餐满意度也相对较低。因此基于高斯函数建立空间衰减规则, 能够较好的模拟需求和共计在空间距离增大时发生的衰减现象, 因此也成为两步移动搜索法诸多扩展形式中最为常用的方法之一 (Xiang Chen, 2019)。根据式 (4.1) 有

$$P_k = G(d_{ij})D_k, \quad i = 1, 2, \dots, o; j = 1, 2, \dots, n; k = 1, 2, \dots, m \quad (4.4)$$

$$G(d_{ij}) = \frac{e^{-0.5 \times \left(\frac{d_{ij}}{d_0}\right)^2} - e^{-0.5}}{1 - e^{-0.5}} \quad (4.5)$$

其中 D_k 是每个人口栅格中的人口数量, d_{ij} 是搜索点与搜索域之间的距离, d_0 是医疗设施服务的极限距离。 $G(d_{ij})$ 为考虑空间摩擦问题的高斯衰减函数。

由此可得：

$$R_j = \frac{S_j}{\sum_{k \in \{d_{kj} < d_0\}} G(d_{ij}) D_k} \quad (4.6)$$

$$A_i = \sum_{j \in \{d_{ij} < d_0\}} G(d_{ij}) R_j \quad (4.7)$$

参考兰州市交通情况和国家提出的城区15分钟步行健康圈目标，分别将一至三级别医疗设施的常规最大服务半径设定为40千米、20千米、5千米。将处理过的医疗设施缓冲区数据和兰州市主城区人口栅格数据导入ArcGIS Pro中，设定坐标系为中央经线为东经103.66的WGS1984坐标系，并计算缓冲区面积。

4.4.2 医疗设施空间可达性测度

首先以服务设施 j 为中心， d_0 为搜索半径，通过ArcGIS Pro中的生成近邻表功能，搜索计算搜索域内的人口数量 D_k 和距离 d_{ij} ，并按照式 (4.5)、(4.6) 计算人均供需比 R_j 。其次再次以任一人口位置 i 为中心， d_0 为搜索半径，再次使用生成近邻表功能，计算搜索域内服务设施的供需比 R_j 且在高斯衰减函数基础上汇总求和，根据式 (4.5)、(4.7) 可得到居民点 i 基于距离成本下的医疗设施空间可达性 A_i 。将数据计算汇总后可以得到兰州市主城区各网格中医疗设施空间可达性情况。

4.5 兰州市主城区医疗设施空间可达性结果分析

为了更好的体现医疗设施在兰州市不同区域之间医疗设施的可达性的相对情况，按照几何间隔对医疗设施空间可达性情况按照“好”、“较好”、“中”、“较差”、“差”五个级别进行划分，并将数据进行空间连接后制成图像，以便直观体现兰州市各区域内的空间可达性情况，具体如表4.1所示。

表4.1说明，兰州市主城区医疗设施可达性水平最好级别的网格数占绝大多数，为961个，在这些网格中一级医疗设施最多，为45个。较好级别网格数为204个，总体来看，兰州市主城区医疗设施可达性水平级别较好。

表 4.1 兰州主城区医疗设施可达性水平分级

可达性水平	指数区间	网格个数	医疗设施个数	一级医疗设施个数
最好	0.89-1	737	961	45
较好	0.81-0.89	703	204	11
中	0.68-0.81	821	33	0
较差	0.49-0.68	968	15	2
差	0-0.49	654	8	1

兰州主城区医疗设施各级可达性水平空间分布如图4.2所示。

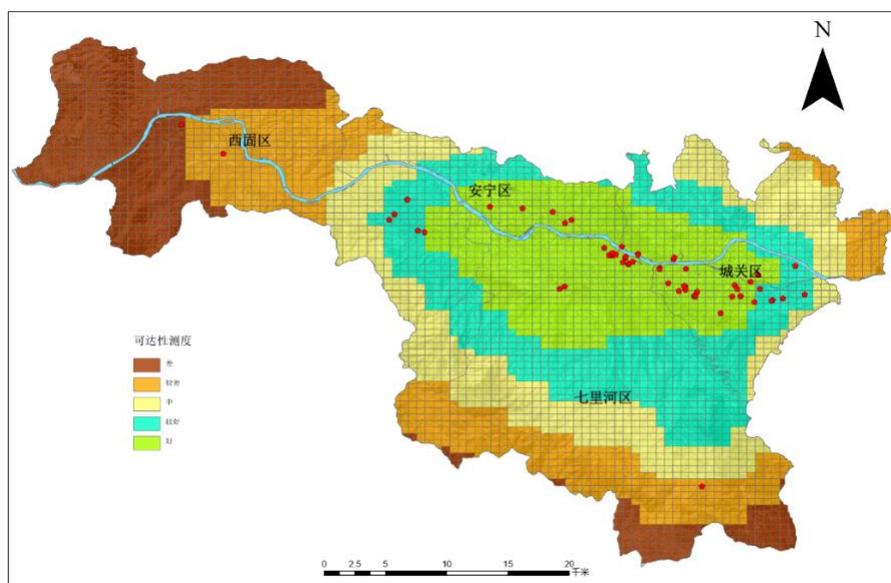


图 4.2 兰州市主城区医疗设施空间可达性

由图4.2可见，城关区西部、七里河北部以及安宁区东部核心街区可达性最好，是兰州市的经济文化中心，人口密度较大，一级医疗设施个数最多。以该类区域为中心越向外扩散，可达性逐渐降低。

由此可见，兰州医疗设施可达性总体较好，城关区的西部、七里河北部以及安宁区东部的核心街区医疗设施空间可达性最好，可达性指数在0.89-0.1之间，此类区域是兰州市经济中心和政治中心，人口密度大，一级医疗设施个数最多，是研究区内医疗设施空间通达性最佳的区域。

可达性较好的区域，其可达性指数区间在0.81-0.88内，空间上分布在可达性最好区域的周边，大部分是城市的郊区，人口密度较高，区域内有八个一级医疗设施，同时由于兰州市交通路网较为发达，因而此类区域虽然无法与核心街区相比较，但医疗设施可达性情况依然较好，能够较为便利的实现医疗需求。

可达性中等的区域基本上是可达性最好的区域向外扩散的区域，可达性指数在0.68-0.80之间，空间上大致分布在城乡结合处，主要位于城关区的东北部、七里河区的中部和西固区的东部。人口密度和医疗设施个数与核心街区相比均存在较大差距，但得益于地理空间距离的优势，医疗设施可达性适中。

医疗设施较差的区域位于西固区的中部、城关区东北部和七里河区的中南部，可达性指数在0.49-0.67内，区域大多是距离城市较近的乡镇，虽然区域内有两个一级医疗设施，但人口密度相对较大且距离核心城区距离较远，因此医疗设施的可达性较差。

医疗设施可达性最差的区域位于七里河区的南部和西固区的西部。仅有一个二级医疗设施和七个三级医疗设施，医疗资源较为匮乏，人口密度相对较低。此类区域位于研究区的边界位置与核心街区的距离最远，医疗设施的可达性情况最差。

可以看出，兰州市主城区医疗设施可达性情况整体较好，城乡结合处和乡村地区医疗设施可达性差，医疗资源较少的情况虽然符合供需关系，但也存在一定的医疗设施空间分布不均衡的现象。

4.6 小结

以兰州市主城区为例，针对三级医疗设施设置不同的搜索半径，并通过构建兰州市主城区第七次人口普查栅格作为两步移动搜索法的需求端进行了医疗设施空间可达性情况的测度。

由于医疗服务设施是一类相对注重供给能力，且需要兼具服务公平性的服务设施，因而使用两步搜索法能够较好的表现医疗设施的空间可达性情况。同时通过高斯函数按照距离进行衰减也较为符合现实中距离引发的供给减弱和需求量减小情况。基于第七次人口普查数据进行的人口栅格修整，平均误差由修正前的27.3%，减低为0.01%，较为准确的反映了兰州市人口的空间分布情况。

研究结果说明，基于改进两步移动搜索法的兰州市医疗设施空间可达性测度情况较好，较为准确的反映出了医疗设施的空间分布情况，为进一步优化医疗设施布局方案提供了参考。

5. 基于随机森林方法的医疗设施选址推荐

5.1 引言

由于医疗服务的特殊性,国内外诸多学者多以创新和改进各类可达性测度模型进行医疗设施选址和布局的相关研究,宋正娜(2009)通过构建潜能模型对江苏省如东县医院的可达性进行了评价;Yang N 等(2016)基于平衡分析对武汉市医疗设施的可达性进行了评价。尤其是在新冠病毒全球大流行的背景下,医疗资源的紧缺和医疗设施可达性的不均衡问题尤为突出,近年来也出现了针对医疗设施分布公平性的研究,胡舒云等(2021)基于多源大数据对深圳市医疗设施的公平性进行测算和评价,张金萍等(2022)通过改进 3SFCA 算法对于乡村地区的以设施公平性进行了评价。

当前基于空间可达性的医疗设施选址和布局优化,大多数情况下只考虑了医疗设施与需求点间的空间距离和供需关系。但在可达性评价因子的选取以及定权的过程中存在一定的主观性,且多依赖于前期研究的经验,可能无法符合城市高速发展的现状。同时医疗设施作为城市各类设施中的必要部分,其分布一方面与城市的早期规划有关,另一方面也与所处的区位和周边设施有关。因而通过多角度进行医疗设施布局优化研究才能更为精准的满足居民需求。

近年来大数据和机器学习算法广泛应用于最优化问题中,国内外学者也将其引入了设施选址布局研究中,如,戈好雨(2020)基于设施的人流量和地理信息数据构建了指标体系,利用随机森林算法对无人零售店的位置和销量进行了预测。张嘉琪等(2022)基于 POI 数据利用随机森林方法对线下广告牌进行了选址推荐。汪晓春等(2021)以人口老龄化数据为基础使用决策树对武汉市养老设施进行了选址布局。引入机器学习算法,一方面通过挖掘大数据中的设施空间分布特征,综合各方面因素进行了设施的选址布局研究;另一方面也减少了构建评价模型时的主观性。

基于城市 POI 数据使用随机森林算法,以兰州市主城区为例,对城市医疗设施规划布局进行选址推荐研究。首先利用 ArcGIS Pro 软件对研究区域进行网格化处理与分区统计分析,利用随机森林算法提取网格内医疗设施的分布特征,并以此为依据进行布局推荐。

5.2 随机森林算法

由于设施的选址过程可以近似看成是一个不断搜寻点位以逐个满足设定条件的二叉树问题，因而将研究区进行网格化处理后，利用随机森林算法预测推荐网格。

设有集合 $U = \{U_1, U_2, \dots, U_n\}$ ， U 中包含了研究区域内的全部设施。其中研究设定的目标设施为 T ，研究区内此类设施依次为 $t_i (i=1, 2, \dots, n)$ ；研究区中目标设施 T 外的其他类的设施记为 $O^j (j=1, 2, \dots, m)$ ，研究区内其他各类设施点就依次为 $o_i^j (i=1, 2, \dots, n)$ ，进而研究区内含有 $j+1$ 类设施。并将研究区分为边长为 E 的网格，共有 e 个。其中存在目标设施 T 的网格记为 R ，依次为 $r_k (k=1, 2, \dots, n)$ ；不存在 T 的记为 Q ，依次为 $q_l (l=1, 2, \dots, n)$ 。

随机森林方法是一种基于决策树方法构建而得到的集成学习方法，利用 Bootstrap 重抽样方法从原始样本中抽取多个样本，对每个 Bootstrap 样本进行决策树建模，然后组合多棵决策树的预测，通过投票得出最终预测结果。

设定样本集为 $U = \{U_1, U_2, \dots, U_n\}$ ，其对应的标签集 $X = \{T, O^1, O^2, \dots, O^j\}$ 。其中 T 类标签下的样本为 $t_i (i=1, 2, \dots, n)$ ；同理，第 $j+1$ 类标签 O^j 下的样本为 $o_i^j (i=1, 2, \dots, n)$ ，则有 X 的信息熵为：

$$h(X) = -\sum_{j=1}^m p_j \cdot \log_2 p_j \quad (5.1)$$

式中，第 $j+1$ 类标签的概率 p_j 用该类样本的个数与总样本数的比值进行估计。

设有 A 属性 $A = \{A_1, A_2, \dots, A_k\}$ ，其中第 i 个条件属性的取值为 $A_i (i=1, 2, \dots, k)$ ， $|A_j|$ 为该取值的样本数， $|A_{ij}|$ 为 A_i 中的第 j 类样本数，则属性 A 进行划分的信息熵为：

$$h(X|A) = -\sum_{i=1}^k \frac{|A_i|}{|S|} \cdot \sum_{j=1}^m p_{ij} \log_2 p_{ij} \quad (5.2)$$

最终可以得到属性 A 划分前后的信息增益为：

$$g(X, A) = h(X) - h(X | A) \quad (5.3)$$

在此之后经过遍历样本的信息熵和各类标签的信息增益实现信息增益最大化，通过递归和迭代，构建决策树模型。随后利用 Bootstrap 重抽样方法从样本中抽取多个样本，通过 k 轮训练，得到一个分类模型序列 $\{h_1(X), h_2(X), \dots, h_k(X)\}$ ，再用其构成一个多分类模型系统，该系统的最终分类结果采用简单多数投票法。最终的分类决策为：

$$H(X) = \arg \max_T \sum_{i=1}^k I(h_i(X) = T) \quad (5.4)$$

其中， $H(X)$ 表示组合分类模型， $h_i(X)$ 是单个决策树分类模型， T 即为输出的标签。

进而可以得到网格的推荐度 S ：

$$S = \frac{C(H_{h(X)=T}(X))}{C(H(X))} \quad (5.5)$$

其中， $C(H(X))$ 为全部分类模型的个数， $C(H_{h(X)=T}(X))$ 为归类在 T 标签下分类器的个数（方匡南，2011；李航，2021）。

5.3 POI 数据的整合与网格化处理

5.3.1 POI 数据的整合

医疗设施为目标设施 T ，在经过修剪后共 1221 个样本。在爬取的除医疗设施外的兰州市主城区 POI 数据中，还包含了部分地图标注信息，为了避免信息重叠，将地名地址信息、室内设施、通行设施这三个大类剔除。随后将研究区内的 POI 数据归为六类，具体归类情况如下表 5.1 所示。

政府及办公场所标签内包含了研究区内的日常办公场所，涵盖了政府机构、公司、企业、金融等行业，这类设施工作日时的人口流动量大、人口密度大，非工作日时人口密度较小，基于这一特性，将其单独归为一类，记为 O^1 。

道路交通与相关设施标签内包含了研究区内的道路交通设施与交通工具服

务场所。这一类设施大多建立在主干道、路边等交通可达性优良的地区，因而可以直观的反映当前区块的交通便利程度，将其归为一类，记为 O^2 。

表 5.1 兰州主城区 POI 数据特征

特征	一级类别	POI 数据	个数
政府及企业办公场所	政府机构及社会团体	工商税务机构、公检法机构、政府机关等	4274
	公司企业	工厂、公司企业办公场所、农林渔牧基地等	6801
	金融保险服务	保险公司、财务公司、银行、ATM 机等	1650
	道路附属设施	高速服务站、收费站等	11
道路交通与相关设施	交通设施服务	地铁站、港口码头、公交车站、火车站等	4608
	摩托车服务	摩托车服务、维修、销售等	193
	汽车服务	充电桩、加油站、汽车保养等	2581
日常服务与娱乐设施	汽车维修	各类品牌汽车维修点	1131
	生活服务	运营商营业厅、邮局、快递站、电力营业厅等	15455
	风景名胜	各类景点、公园、广场及其周边相关设施	508
	动物医疗场所	动物医疗场所、药房、诊所、综合医院等	91
	医疗保健服务场所	养生馆、按摩馆、保健护理室等	680
	体育休闲服务	疗养场所、电影院、运动场馆等休闲娱乐场所	2392
	公共设施	报刊亭、电话亭、公共厕所等公共服务设施	932
销售与购物场所	购物服务	便民超市、各类大型商场、商业街等	29337
	汽车销售	各类品牌汽车销售店	520
	医药保健销售店	药店、保健品销售店等	1646
住宅与科教设施	商务住宅	产业园区、楼宇、住宅区等	3542
	科教文化服务	大学、中学、小学、科研机构、培训机构等	4510
餐饮服务场所	住宿服务	宾馆、大型酒店、招待所等	3500
	餐饮服务场所	各类餐馆、甜品店、饮品店等	18987

日常服务与娱乐设施标签内包含了日常生活中必须的便民设施，如，快递点、理发店、运营商营业厅、公共厕所等，以及公园、景点、电影院、酒吧等休闲娱乐设施。此类设施主要分布在商圈以及住宅区附近，将其归为一类，记为 O^3 。

购物服务设施中包含的 POI 数据数量有 29337 条，是所有标签中最的一项。其中包含了超市、商场、综合商店等物品销售场所，将汽车销售与医疗保健销售场所与其合并，定义为销售与购物场所标签，记为 O^4 。

住宅与科教设施标签内的设施大多为常住人口密集地，主要包括小区、住宅、学校、培训机构以及酒店住宿等，记为 O^5 。餐饮服务场所标签下主要包含了各类餐馆、饭店等提供餐饮服务的设施，由于其设施的独立性，将其单独列为一类，记为 O^6 。

5.3.2 研究区域网格化处理

为方便后续研究,对所有的 POI 数据按照上述所属的类别进行标识。按照经纬度信息以点的形式将 POI 数据导入 ArcGIS Pro 软件中,在研究区范围内以 $E = 500$ 米构建网格,将研究区域分成了 3883 个网格,并通过构建掩膜的方式排除了黄河等无法建设设施的水体,需要说明的是在此参考的地理坐标系为 WGS1984。

将 POI 数据与网格进行空间连接,进而确定每个 POI 数据所处的网格及网格内所包含的各类 POI 数据的个数,如在 205 号网格中共有 POI 点 12 个,其中 O^1 类有 4 个设施、 O^2 类有 2 个设施、 O^3 类有 5 个设施、 O^5 类有 1 个设施。

进行汇总得到,在研究区的 3883 个网格中,空白网格有 2559 个,包含医疗设施 (T 类) 的网格 R 有 384 个。不存在目标设施的网格 Q 有 940 个。并将包含 T 类设施的网格 R 标记为正标签,即 True,将不包含的标记为负标签,即 False。同时对其余标签下的数据也进行离散化处理,以便后续构建随机森林模型。

5.4 选址推荐模型的构建与训练

基于 Python 软件中的 Scikit-Learn 库进行随机森林选址推荐模型的构建,选择生成的每个决策树随机抽取 30% 的样本进行训练,并且设定生成树的上限为 400 个。

训练生成的决策树为 321 棵,此时得到分类评价指标 ROC 得分(即测试得到数据和标签分类的平均精确度)为 0.8461。为了进一步提高分类精度和算法效率,通过固定变量交叉验证的方法,实现选址推荐模型参数优化。

由计算得出本研究构建的医疗设施随机森林选址推荐模型的最佳分类计算方法为基尼指数,决策树的最佳个数为 215 棵。为防止过拟合,得到随机森林的最大深度为 3 层、剪枝处最小样本为 10 个、分支节点最小样本数为 120 个。最终优化参数后的 ROC 得分为 0.8977。

构建的随机森林模型得到的各类标签对模型的特征贡献度如表 5.2 所示。

表 5.2 兰州主城区 POI 数据特征

设施类型	标签编号	特征贡献度	排名
住宅与科教设施	5	0.266262	1
日常服务与娱乐设施	3	0.233887	2
销售与购物场所	4	0.194357	3
餐饮服务场所	6	0.162546	4
政府及企业办公场所	1	0.099074	5
道路交通与相关设施	2	0.043875	6

由表 5.2 可得贡献度排名第一和第二的分别是“住宅与科教设施”与“日常服务与娱乐设施”，从正面和侧面分别体现了城市中的人口密集区域，这符合医疗设施分布与人口密集度呈正相关的客观规律。

特征贡献度排名三、四的“销售与购物场所”与“餐饮服务场所”大多建立在商圈和交通可达性较好的地区，代表着交通要素在医疗设施选址中的重要地位。

由参数优化得到的 ROC 得分可以看出本模型具有较好的模型分类精度，根据模型分类标签的贡献度可以看出分类结果与现实客观规律基本一致，具有较好的预测能力，可以使用此模型进行医疗设施的选址推荐。

5.5 兰州市主城区医疗设施选址推荐结果分析

将全部数据再次导入上述训练好的随机森林医疗设施选址推荐模型中，就可以得到预测结果。设施点位的推荐基于所处的网格进行展开，不同网格的推荐程度（简称推荐度）以概率的形式呈现，概率越大说明构建的随机森林中有较多个数的分类器将这一网格预测为“True”标签，推荐度越高。同理，推荐概率越小则说明构建的随机森林中较多个数的分类器将这一网格预测为“False”标签，推荐度越低。

在使用随机森林算法对问题进行分类研究时，推荐度低于 50%一般被认为是不被推荐。研究发现，在进行公共设施布局规划时，通过扩大推荐度范围，可以逐渐扩大设施的区域覆盖率，进而提高公共设施的公平性。不同推荐度范围下的网格推荐情况如下表 5.3 所示。

表 5.3 网格推荐结果汇总

推荐度范围	50%-100%		40%-50%		30%-40%		20%-30%		10%--20%	
网格总数	331		25		42		25		57	
当前网格内有	<i>R</i>	<i>Q</i>								
无医疗设施	288	43	11	14	13	29	6	9	9	8

注：*R* 表示已有医疗设施的网格数；*Q* 表示无医疗设施的网格数

为直观表示以上分布情况，基于兰州市主城区的遥感影像数据制作了图 5.1 进行可视化展示。

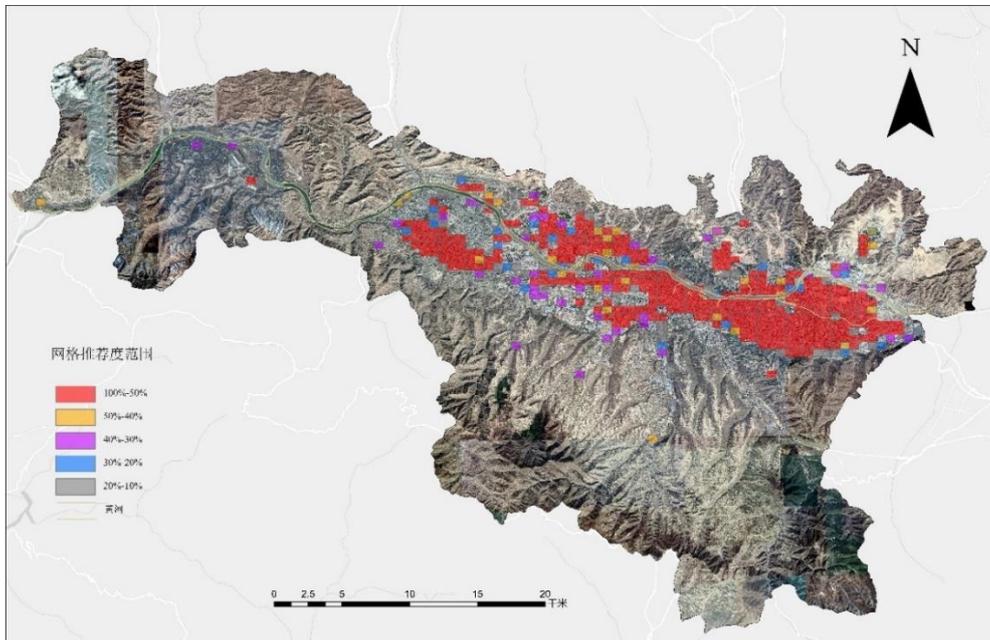


图 5.1 网格推荐结果空间分布

图 5.1 中红色为推荐度在 100%-50%时推荐布置医疗设施的网格，共有 331 个。由此可以看出，此类网格多密集分布在市区繁华且人口密度高的地区，这与已有医疗服务设施的分布网络高度重合。

黄色为 50%-40%推荐度内的网格，共有 25 个，其中已含有医疗设施的网格占 44%。空间上此类网格多分布在红色网格的周边，为城区内相对偏远的区域，以及乡镇的中心。

紫色为 40%-30%推荐度内的网格，共有 42 个，其中已含有医疗设施的网格约占 31%。空间上此类网格多分布在各区的分界处，距离城市中心较远，但人口密度相对较高。

蓝色为 30%-20%推荐度内的网格，共有 25 个，其中已含有医疗设施的网格约占 24%、地图空间上此类网格中除分布在黄河上外，大多分布在城市边缘的工业区中，以及乡镇的周边区域。

灰色为 20%-10%推荐度内的网格，共有 57 个，大多分布在城乡的结合处或乡镇的中心，还有部分位于研究区内海拔相对较高的山区。

由此可见，当推荐度在 50%以上时，网格大多分布在现在已有医疗设施的网格下，及其周边，大多都是城市的核心区域，医疗资源丰富。在逐步扩大推荐度范围时，网格也就由城市核心区域向边缘区域以及城镇中心偏移，逐步扩大了医疗设施的覆盖范围，提高了公平性。

5.6 小结

(1) 分析路网数据和 POI 数据的空间分布情况可以看出，兰州市主城区的医疗资源主要集中在城关区和安宁区的城市核心街区，且分布十分密集，交通路网密集，表明研究区内医疗设施交通可达性整体较好，但城市边缘地区与北部山区距离城关区、安宁区的核心街道较远，医疗设施相对较少。

(2) 随机森林算法特征提取结果显示，住宅与科教设施类标签和日常服务与娱乐设施类标签在随机森林分类过程中起到了决定性的作用，是影响现有兰州市主城区医疗设施分布的重要因素，符合医疗设施与人口密度具有强相关性的客观规律。

6. 基于多源数据的医疗设施布局优化方案

为了实现兰州市主城区医疗设施布局情况的综合优化,将基于前文研究得到的兰州市主城区功能区分布情况、医疗设施空间可达性情况、随机森林选址推荐度三个维度,分四类进行优化方案的设计。

为了方便优化方案设计,将兰州市医疗设施空间可达性分布情况中的“较好”和“好”归为一类,即当前研究区内医疗设施空间可达性较好的区域。将“中”、“较差”、“差”三个情况归为一类,即当前研究区内医疗设施空间可达性较差的区域。同时将随机森林选址推荐度在 100%-40%的网格归为一类,即当前研究区内较为推荐设置医疗设施的网格。将选址推荐度在 40%-0%的网格归为一类,即研究区内相对不推荐设置医疗设施的网格。具体分类情况如表 6.1 所示

表 6.1 兰州市主城区选址推荐与可达性综合情况

		可达性情况		总计
		较差	较好	
选址推荐度	较低	2366 (A)	1161 (B)	3527
	较高	17 (D)	339 (C)	356
总计		2383	1500	3883

将上述结果进行空间相交处理后,与兰州市主城区 500 米*500 米网格相关联,再在将功能区也叠加后的兰州市主城区城市空间特征识别情况如图 6.1 所示。

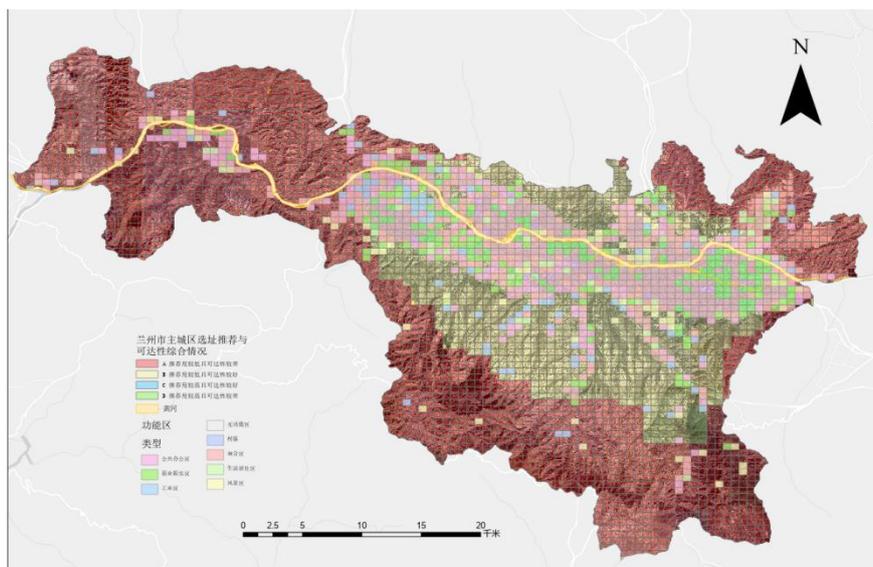


图 6.1 兰州市主城区城市空间特征识别情况

由图 6.1 可见，通过三种空间识别结果叠加制成的城市空间特征识别图，能够清晰直观的展现设施分布特征，进而针对不同分类进行布局优化方案设计。

6.1 推荐度较低且可达性较差区域优化方案

图 6.1 可以看出，A 类网格是研究区域内医疗设施选址推荐度较低，且空间可达性较差的区域，在图中以红色标识，此类区域大致位于七里河区南部、城关区的东北北部、以及西固区的西部和中部，共 2366 个。A 类网格中七里河区南部为海拔高度接近 3000 米的山区，大多为乡镇，人口密度较低，医疗需求相较于核心城区较低。城关区盐场路街道和青白石街道的北部区域也属于 A 类网格中，虽然海拔在研究区内相对始终，但依旧属于山地，人口较低。A 区覆盖了西固区的大部分区域，区域内虽然有两所一级医疗设施，但由于医疗设施数量整体较小，因而人均可达性水平较低。

由于 A 类网格中大部分功能区是村落、风景区以及部分工厂所在的公共办公区和工厂区，且多分布在村镇中心。因而对于 A 类网格，一方面要通过驻村医生等方式增加基层医疗设施投入，提高人均医疗设施可达性水平，另一方面要在较大的村镇中心和工厂院区内等人口密度相对较大的区域设立二级医疗设施，以实现基础性病情的有效处理和及时转运患者的能力。

6.2 推荐度较低且可达性较好区域优化方案

B 类网格是选址推荐度较低且空间可达性较好的区域，在图中以黄色进行标识，大致分布在安宁区核心街区的周边、西固区核心街道的周边、七里河区的中部北部以及城关区核心街区的周边区域，共 1161 个。可以看出 B 类网格在空间上大多分布在兰州市四个区的核心街区周围，大多是乡镇和城市边缘区域，因为距离各区域内城市中心较近，交通通达性较好，且人口密度相对较低医疗设施需求量较少，因此人均医疗设施可达性较好，医疗设施选址推荐度较低。

结合功能区分布情况可以看出，除了西固区与七里河区交界处刘家堡街道和陈萍街道内以及和城关区中北部的一些区域的 B 类网格是商业娱乐区以外，大部分都是混合区、工业区和公共办公区，且多沿交通主干道分布。因此对于 B 类网格在交通主干道和人口密度较大的工业办公区应该增设或强化现有三级医疗

设施的服务能力，使其能够满足和处理基本的医疗需求。

6.3 推荐度较高且可达性较好区域优化方案

C类网格是医疗设施选址推荐度高同时空间可达性较好的区域，在图中以蓝色进行标示。C类网格空间上分布在兰州市城关区、安宁区、七里河区、西固区的核心街区内，且大致分布在黄河的两岸，是兰州市的经济和文化中心，医疗设施丰富，经济发达，人口密度大。

C类网格大多是综合了商业娱乐、公共办公、生活居住等多种功能的混合区，其医疗资源密集。因此对于C类网格一方面要呼吁和引导城市核心街区居民进行分级治疗，对于小症、轻症可以就近去居住区附近的二三级医疗设施进行救治。另一方面也要加强二三级医疗设施的服务能力，扩大投入二三级医疗设施药品数量、床位、医生数量等基础建设，减少一级医疗设施的挤兑现象。

6.4 推荐度较高且可达性较差区域优化方案

D类网格是医疗设施选址推荐度高且空间可达性较差的区域，是当前兰州市主城区医疗设施布局中的缺漏点，排除黄河经过的2844号网格和2871号网格以外，共有15个，下面将逐个进行分析并提出布局优化方案。



图 6.2 2838 号网格和 2920 号网格

如图 6.2 所示，2838 号网格和 2920 号网格分别位于西固区的达州乡和东川镇，是村镇的中心区域，分布属于公共办公区和混合区，居住人口相对较多，除 2920 网格内有一个三级医疗诊所以外无其他医疗设施。虽然临近网格内有卫生院等基础医疗设施，但由于人口密度较大，医疗需求较大，因此医疗设施的可达性水平较差。因此对于这两个网格，通过扩大周边医疗设施基础建设投入，增强

其服务能力，以此满足此区域内较大的医疗需求。



图 6.3 2847 号网格

2847 号网格是兰州石化公司门口的商业街,属于混合区,在其正南部的 2813 网格中有两个三级医疗设施。对于此网格,一方面可以出台相关政策,在必要允许大型工厂企业部分周边人员到工厂内部医院进行治疗,另一方面也应加强周边医疗设施的服务能力,满足到此务工的流动人员的医疗需求。



图 6.4 2327、2370、2371、2411 号网格

如图 6.4 所示四个绿色网格连在一起构成了一个 T 字形。此区域位于城关区青白石街道中部,属于混合区。区域内有数个兰州市新开发在建的楼盘,是未来兰州市主城区的一大发展方向,因此应该提前布局建立一二级医疗设施,未雨绸缪,以满足未来此区域内可能出现的大量医疗需求。



图 6.5 1450 和 1493 号网格

1450 和 1493 号网格位于城关区东岗街道，是城市的郊区，由于覆盖了部分五金建材市场和居住区，因此属于混合区。区域内人口密度较大，且有五个三级医疗设施，因此需要提高现有医疗设施的服务能力，并选择合适区位将其升级为二级医疗设施，以满足附近较大人群的医疗需求。



图 6.6 2421 号网格及周边的六个网格

此片区域在空间上呈反“F”状，属于西固区西柳沟街道和临洮街街道，周边有数个工厂居民区和学校。也覆盖了部分工厂厂房，属于混合区和办公区，是西固区的西部郊区。人口密度相对较大，网格内仅存在八个三级医疗设施，无法满足附近居民的医疗需求。可以选择将 2421 号网格北部临街的三级医疗设施升级为二级、一级医疗设施。其次，也可在 2380 南部临街区域选择新建二级、一级医疗设施，以此提高此区域内的人均医疗设施可达性水平。

7. 结论与展望

7.1 结论

(1) 通过城市功能区识别、公共设施空间可达性测度、随机森林选址推荐三个角度,基于多源大数据对兰州市主城区的医疗设施的空间分布特征进行了识别和挖掘,并以此为基础分四类设计了兰州市主城区医疗设施布局优化方案。

(2) 基于 Mask R-CNN 算法和样方密度法的功能识别研究,从人文和自然两个角度对兰州市主城区的功能区空间分布情况进行了综合性的识别,弥补了单一数据来源和分析角度造成的结果偏差和数据缺失问题,同时基于神经网络算法大幅度减少了在建筑标记时的人工时间成本。

(3) 基于改进两步移动搜索法的医疗设施空间可达性评价研究中,通过利用人口栅格和第七次人口普查数据修正得到兰州市主城区第七次人口栅格,使一维的人口数据空间化,根据贴近现实人口分布情况,进而实现兰州市主城区医疗设施可达性的精准测度。同时通过改进两步移动搜索法基于供求关系视角构建模型对医疗设施可达性进行测度,更加贴近医疗设施的服务模式。

(4) 通过随机森林算法对研究区内的 POI 数据进行分布特征的挖掘,并以此为基础进行设施分布的推荐预测,降低了人工构建可达性测度模型的主观性,同时将城市内部的全部设施均纳入挖掘范围内,提高了纳入模型影响因素的数量。

(5) 通过针对三种空间特征识别结果提出布局优化方案,与空间特征识别方法进行的布局优化方案相比更具有综合性,相关方案为城市管理者进行进一步城市规划优化提供了参考方案,具有一定的现实意义。

7.2 展望

(1) Mask R-CNN 算法进行建筑物提取结果坐标与 POI 数据存在偏差,未来可以在如何消除坐标偏差上展开。

(2) 基于改进两步移动搜索法进行可达性测度研究是围绕供给和需求点直接的直接物理距离展开的,没有考虑道路路网因素,与现实情况存在一定的差距,在后续研究中可以展开进行优化。

(3) 通过随机森林算法挖掘 POI 数据进行选址推荐,在 POI 数据较为丰富的网格中表现较好,但存在 POI 数据缺失的网格中表现较差,在后续研究中可研究针对性的解决方法。

参考文献

- [1] Amaya V, Moulaert T, Gwiazdzinski L, *et al.* Assessing and qualifying neighborhood walkability for older adults: construction and initial testing of a multivariate spatial accessibility model[J]. *International journal of environmental research and public health*, 2022, 19(3): 1808-1826.
- [2] Brandeau M L, Chiu S S. An overview of representative problems in location research[J]. *Management science*, 1989, 35(6): 645-674.
- [3] Chen X, Jia P. A comparative analysis of accessibility measures by the two-step floating catchment area (2SFCA) method[J]. *International Journal of Geographical Information Science*, 2019, 33(9): 1739-1758.
- [4] Church R. ReVelle C. The maximal covering location problem[J]. *Papers in Regional Science*.1974,32(1):101-118
- [5] Coulter P. Measuring the inequity of urban public services: A methodological discussion with applications[J]. *Policy Studies Journal*, 2008, 8(5): 683-698.
- [6] Du Z, Zhang X, Li W, *et al.* A multi-modal transportation data-driven approach to identify urban functional zones: An exploration based on Hangzhou City, China[J]. *Transactions in GIS*, 2020, 24(1): 123-141.
- [7] El-Ramady H, Brevik E C, Elbasiouny H, *et al.* Planning for disposal of COVID-19 pandemic wastes in developing countries: a review of current challenges[J]. *Environmental Monitoring and Assessment*, 2021, 193(9): 1-15.
- [8] Geertman S C M, Ritsema Van Eck J R. GIS and models of accessibility potential: an application in planning[J]. *International journal of geographical information systems*, 1995, 9(1): 67-80.
- [9] Gutiérrez J, Urbano P. Accessibility in the European Union: the impact of the trans-European road network[J]. *Journal of transport Geography*, 1996, 4(1): 15-25.
- [10] Hakimi S L. Optimum locations of switching centers and the absolute centers and medians of a graph[J]. *Operations research*, 1964, 12(3): 450-459.
- [11] Hansen W G. How accessibility shapes land use[J]. *Journal of the American Institute of planners*, 1959, 25(2): 73-76.
- [12] He K, Gkioxari G, Dollár P, *et al.* Mask R-CNN[C]//*Proceedings of the IEEE international conference on computer vision*. 2017: 2961-2969.
- [13] Hong Y, Yao Y. Hierarchical community detection and functional area identification with OSM roads and complex graph theory[J]. *International Journal of Geographical Information Science*, 2019, 33(8): 1569-1587.
- [14] Ihlanfeldt K R, Raper M D. The intrametropolitan location of new office firms[J]. *Land Economics*, 1990, 66(2): 182-198.

- [15] Ingram D R. The concept of accessibility: a search for an operational form[J]. *Regional studies*, 1971, 5(2): 101-107.
- [16] Joseph A E, Bantock P R. Measuring potential physical accessibility to general practitioners in rural areas: a method and case study[J]. *Social science & medicine*, 1982, 16(1): 85-90.
- [17] Linneker B J, Spence N A. Accessibility measures compared in an analysis of the impact of the M25 London Orbital Motorway on Britain[J]. *Environment and Planning A*, 1992, 24(8): 1137-1154.
- [18] Liu J, Cao L, Zhang D, *et al.* Optimization of Site Selection for Emergency Medical Facilities considering the SEIR Model[J]. *Computational Intelligence and Neuroscience*, 2022, 1-17.
- [19] Liu L, Lyu H, Zhao Y, *et al.* An Improved Two-Step Floating Catchment Area (2SFCA) Method for Measuring Spatial Accessibility to Elderly Care Facilities in Xi'an, China[J]. *International Journal of Environmental Research and Public Health*, 2022, 19(18): 11465.
- [20] Liu X, He J, Yao Y, *et al.* Classifying urban land use by integrating remote sensing and social media data[J]. *International Journal of Geographical Information Science*, 2017, 31(8): 1675-1696.
- [21] Luo W, Wang F. Measures of spatial accessibility to health care in a GIS environment: synthesis and a case study in the Chicago region[J]. *Environment and planning B: planning and design*, 2003, 30(6): 865-884.
- [22] O'Neill W A, Ramsey R D, Chou J C. Analysis of transit service areas using geographic information systems[J]. *Transportation Research Record*, 1992, 1364: 131-138.
- [23] Peng Z R. The jobs-housing balance and urban commuting[J]. *Urban studies*, 1997, 34(8): 1215-1235.
- [24] Pourrezaie-Khaligh P, Bozorgi-Amiri A, Yousefi-Babadi A, *et al.* Fix-and-optimize approach for a healthcare facility location/network design problem considering equity and accessibility: A case study[J]. *Applied Mathematical Modelling*, 2022, 102: 243-267.
- [25] Radke J, Mu L. Spatial decompositions, modeling and mapping service regions to predict access to social programs[J]. *Geographic Information Sciences*, 2000, 6(2): 105-112.
- [26] Roth R. Computer solutions to minimum cover problems[J]. *Operations Research*, 1969, 17: 455-465
- [27] Salze P, Banos A, Oppert J M, *et al.* Estimating spatial accessibility to facilities on the regional scale: an extended commuting-based interaction potential model[J]. *International journal of health geographics*, 2011, 10(1): 1-16.
- [28] Stegman M A. Accessibility models and residential location[J]. *Journal of the American Institute of planners*, 1969, 35(1): 22-29.
- [29] Tompkins J A, White J A, Bozer Y A, *et al.* *Facilities planning*[M]. John Wiley & Sons, 2010.
- [30] Toregas C, Swaim R, ReVelle C, *et al.* The location of emergency service facilities[J].

- Operations Research,1971,19: 1363-1373
- [31] Tu W, Hu Z, Li L, *et al.* Portraying urban functional zones by coupling remote sensing imagery and human sensing data[J]. Remote sensing, 2018, 10(1): 141.
- [32] Wan N, Zou B, Sternberg T. A three-step floating catchment area method for analyzing spatial access to health services[J]. International Journal of Geographical Information Science, 2012, 26(6): 1073-1089.
- [33] Yang J, Cao J, He R, *et al.* A unified clustering approach for identifying functional zones in suburban and urban areas[C]//IEEE INFOCOM 2018-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). IEEE, 2018: 94-99.
- [34] Yang N, Chen S Y, Hu W L, *et al.* Spatial distribution balance analysis of hospitals in Wuhan[J]. International Journal of Environmental Research and Public Health, 2016, 13(10) : 1-16.
- [35] Yuan N J, Zheng Y, Xie X, *et al.* Discovering urban functional zones using latent activity trajectories[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 27(3): 712-725.
- [36] Yuan N J, Zheng Y, Xie X. Discovering functional zones in a city using human movements and points of interest[J]. Spatial Analysis and Location Modeling in Urban and Regional Systems, 2018: 33-62.
- [37] Zhang X, Du S, Wang Q. Hierarchical semantic cognition for urban functional zones with VHR satellite images and POI data[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2017, 132: 170-184.
- [38] Zhang X, Li W, Zhang F, *et al.* Identifying urban functional zones using public bicycle rental records and point-of-interest data[J]. ISPRS international journal of geo-information, 2018, 7(12): 1-16.
- [39] Zhang Z, Wang Y, Liu Q, *et al.* A CNN based functional zone classification method for aerial images[C]//2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, 2016: 5449-5452.
- [40] Zhou W, Ming D, Lv X, *et al.* SO-CNN based urban functional zone fine division with VHR remote sensing image[J]. Remote Sensing of Environment, 2020, 236: 111458.
- [41] 陈红顺,郑荣升.基于 Mask R-CNN 遥感图像建筑物自动提取[J].电子设计工程, 2023, 31(02):94-97+102.
- [42] 陈声洪.上海城市交通战略之我见[J].城市规划,1990(03):23.
- [43] 池娇,焦利民,董婷等.基于 POI 数据的城市功能区定量识别及其可视化[J].测绘地理信息, 2016,41(02):68-73.
- [44] 方匡南,吴见彬,朱建平等. 随机森林方法研究综述[J].统计与信息论坛, 2011,26(03): 32-38.
- [45] 戈好雨. 基于机器学习的无人零售商店选址及其销量预测研究[D]. 南京大学, 2020.
- [46] 谷岩岩,焦利民,董婷等.基于多源数据的城市功能区识别及相互作用分析[J].武汉大学学

- 报(信息科学版),2018,43(07):1113-1121.
- [47] 国务院办公厅关于印发《深化医药卫生体制改革 2019 年重点工作任务的通知》[R]. 中华人民共和国国务院公报, 2019(17):18-22.
- [48] 胡舒云,陆玉麒,胡国建等.基于多源大数据的深圳市医疗设施可达性与公平性测算[J].经济地理, 2021,41(11):87-96.
- [49] 贾斐雪,闫金凤,王甜.大数据构建的赋分评价模型与功能区识别研究[J].测绘科学, 2021,46(08):172-178.
- [50] 康雨豪,王玥瑶,夏竹君等.利用 POI 数据的武汉城市功能区划分与识别[J].测绘地理信息, 2018,43(01):81-85.
- [51] 兰州市统计局.兰州市第七次全国人口普查公报[M],兰州.2021.
- [52] 李航. 统计学习方法[M]. 北京: 清华大学出版社, 2021.
- [53] 李漱洋,蔡志昶,唐寄翁.健康韧性视角下社区医疗设施空间布局分析—以南京市中心城区为例[J].现代城市研究,2021(07):45-52+59.
- [54] 宋丽洁,吴政,戴昭鑫.基于多源数据的城市功能区的识别与分析[J].测绘工程, 2023,32(02):63-68+80.
- [55] 宋正娜,陈雯.基于潜能模型的医疗设施空间可达性评价方法[J].地理科学进展, 2009, 28(06):848-854.
- [56] 田玲玲,张晋,王法辉等.公平与效率导向下农村公共医疗资源的空间优化研究—以湖北省仙桃市为例[J].地理科学, 2019, 39(09):1455-1463.
- [57] 仝德,孙裔煜,谢苗苗.基于改进高斯两步移动搜索法的深圳市公园绿地可性评价 [J].地理科学进展, 2021, 40(7): 1113-1126.
- [58] 万波. 公共服务设施选址问题研究[D].华中科技大学,2012.
- [59] 王非,徐渝,李毅学. 离散设施选址问题研究综述[J].运筹与管理, 2006(05):64-69.
- [60] 王皓弘. 基于多模态数据的城市功能区分类集成学习方法研究与实现[D].吉林大学, 2021.
- [61] 吴建军. 基于 GIS 的农村医疗设施空间可达性分析[D]. 河南大学, 2008.
- [62] 杨育军,宋小冬.基于 GIS 的可达性评价方法比较[J].建筑科学与工程学报, 2004(04): 27-32.
- [63] 于珊珊,彭鹏,田晓琴等.基于 GIS 的长沙市医院空间布局及优化研究[J].长沙大学学报, 2012, 26(02):90-94.
- [64] 张嘉琪,杜开虎,任书良等. 多源空间大数据场景下的家装品牌线下广告选址布局研究[J].武汉大学学报(信息科学版), 2022-03-30:1-14.
- [65] 张金萍,林丹,周向丽等. 基于改进 3SFCA 的乡村地区公共医疗设施空间可达性与公平性评价——以海口市为例[J].地理科学进展, 2022,41(04):636-647.

-
- [66] 中国卫生健康委. 中国卫生统计年鉴[M]. 北京: 中国统协和医科大学出版社, 2013.
- [67] 中华人民共和国商务部. 零售业态分类[M]. 北京.2021.
- [68] 中华人民共和国住房和城乡建设部. 中国城市建设统计年鉴[M].北京: 中国统计出版社, 2021.
- [69] 周小平. GIS 支持下的城市医院空间布局优化研究[D].西南交通大学,2007.

致谢

光阴似箭、岁月如梭。一眨眼的功夫，三年的研究生之旅即将画上句号，在这个特殊的时间点，在此占用一些篇幅回忆这段难忘的路程。

犹记得最初踏上这段旅途的时候，在 2020 年的 9 月 2 日。拖着两个行李箱、背着背包，独自乘坐高铁来到了兰州。这虽不是第一次离家，但却是旅途最远的一次。在旅途过程中，看着植被从阔叶到针叶，再到灌木，看着景色由深绿到浅绿，再转黄。1400 公里的路程、10 个小时的高铁，让我第一次对“幅员辽阔”四个字有了深刻地认识。

或许是考研失利，理想与现实之间的差距过大，亦或是离家距离过远，来学校报到之后的一段时间里，我状态很差，整个人非常的不自信，浑浑噩噩不知未来发展方向。对现实不满，但又不知如何排解，对命运不忿，但又不知如何宽恕。所谓“天无绝人之路”，虽然因为种种客观原因和理想的学校失之交臂，但正因如此，才有了研究生三年难忘的历程。

说到这里，我要感谢我的导师——牛成英老师，是您在我状态最差的时候，给了我自信。第一次跟您见面，是在图书馆一楼电子借阅区旁边的小沙发上，当时您问我的最后一个问题是对未来有何打算，我丝毫不掩饰自己堕落的想法，跟您说打算混到毕业考公就业，当时您很勉强的接受了我这个错误的想法。在几天之后的一次和任潇潇师姐一起讨论中，您对我说，“新宇，谁都不知道未来会发生什么，三年后也有可能就业更好、也有可能读博更好，所以我不能对你放低要求，该学的还是得学，该做的还是得做，我得对你负责。”也正是这一句话，让我逐渐开始建立自信，让我明白正因为身处深渊，才更应该努力。

在之后的日子里，我积极参加学生活动，最后成为了校研会主席；积极参与党团活动，从入党积极分子逐渐发展称为正式党员。认真学习专业知识，钻研大数据地理统计方向，获得全国硕士研究生应用统计专业案例大赛校赛第一、国赛二等奖的成绩。科研方面，也尝试投过一些核心刊物，受限于专业知识储备较少和个人能力不足，没能成功录用，但也写出三篇学术小论文，参加两项科研项目，工作量较为饱满，在 2022 年获得了校级二等学习奖学金。

一路走来，还要感谢研究生学生会的各位同僚们，上一届关心照顾我的师兄

师姐们，以及我的两位室友，在与你们的相处中，让我感受到了快乐，让我逐渐真正的认识自己，让我逐渐成长。最后，我还要感谢我的父母，是你们给与的物质、精神支持，才让我能够顺利完成学业，感谢你们多年来为我的付出。

经此一程，学生生涯算是告一段落。站在学校与社会的分界线上，我的心情即激动又忐忑。激动是感觉自己终于学有所成，能够用自己 20 年学生生涯习得的知识和本领服务社会，奉献社会，能够真正的创造价值服务民众。忐忑是担心自己无法承担起这份责任，未来还要面临很多新的挑战。

毕业快乐，感谢自己多年的坚持与奋斗！祝愿我们未来可期！