Learning to Reason over Multi-Granularity Knowledge Graph for Zero-Shot Urban Land-Use Mapping

Yansheng Li[a], Yu Wang[a,*], Lei Yu[b], Bo Dang[a], Gang Xu[c], Zhenyu Zhong[a], Yuning Wu[a], Xin Guo[d], Kang Wu[a], Zheng Li[a], Linlin Wang[a], Jian Wang[b], Jingdong Chen[b], Ming Yang[b], Yongjun Zhang[a,*]

[a]*School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, 430079, Hubei, China*
[b]*Ant Group, Hangzhou, 310013, Zhejiang, China*
[c]*School of Resource and Environmental Sciences, Wuhan University, Wuhan, 430079, Hubei, China*
[d]*Shanghai Academy of Artificial Intelligence for Science, Shanghai, 200003, Shanghai, China*

Abstract

Accurate urban land-use mapping is an essential undertaking for various urban issues, such as urban planning, disease transmission, and climate change. Recently, learning-based method has emerged as a prevalent approach for urban land-use mapping, although it relies heavily on abundant labeled data. However, since land-use categories are jointly determined by physical and social attributes, obtaining such labels is challenging. This scarcity of labeled data often leads existing learning-based methods to overfit, resulting in models that struggle to recognize diverse land-use categories. To bypass these limitations, this paper for the first time advocates knowledge graph to leverage indirect supervision from related tasks for zero-shot land-use mapping. Toward this goal, this paper introduces a multi-granularity knowledge graph reasoning (mKGR) framework.R Only with indirect supervision from other tasks, mKGR can automatically integrate multimodal geospatial data as varying granularity entities and rich spatial-semantic interaction relationships. Subsequently, mKGR incorporates a fault-tolerant knowledge graph embedding method to establish relationships between geographic units and land-use categories, thereby reasoning land-use mapping outcomes. Extensive experiments demonstrate that mKGR not only outperforms existing zero-shot approaches, but also exceeds those with direct supervision, achieving improvements from 0.08 to 0.20 across several performance metrics. Furthermore, this paper reveals the superiority of mKGR in large-scale holistic reasoning, an essential aspect of land-use mapping. Benefiting from mKGR's zero-shot classification and large-scale holistic reasoning capabilities, a comprehensive urban land-use map of China is generated with low-cost. In addition, a nationwide assessment of 15-minute city walkability over the land-use map provides insights for urban planning and sustainable development. The code and data are available at https://github.com/vvangfaye/mKGR.

Keywords: Urban land-use mapping, Knowledge graph, Geospatial big data, Holistic reasoning

*Corresponding author
Email addresses: yansheng.li@whu.edu.cn (Yansheng Li), wangfaye@whu.edu.cn (Yu Wang), leiy@whu.edu.cn (Lei Yu), bodang@whu.edu.cn (Bo Dang), xugang@whu.edu.cn (Gang Xu), zhongzy@whu.edu.cn (Zhenyu Zhong), yuning.wu@whu.edu.cn (Yuning Wu), guoxin@sais.com.cn (Xin Guo), kangwu@whu.edu.cn (Kang Wu), lizheng.s.whu@gmail.com (Zheng Li), wangll@whu.edu.cn (Linlin Wang), bobblair.wj@antgroup.com (Jian Wang), jingdongchen.cjd@antgroup.com (Jingdong Chen), m.yang@antgroup.com (Ming Yang), zhangyj@whu.edu.cn (Yongjun Zhang)

1. Introduction

Urbanization, characterized by increases in urban land coverage and land-use diversity, has farreaching effects on urban dynamics (Chen et al., 2020; Li et al., 2023). The interplay of factors, particularly urban land-use patterns and intensity, is intricately linked to a range of urban opportunities and constraints, including urban planning (Zheng et al., 2023), disease transmission (Alidadi and Sharifi, 2022), as well as climate change (Deng et al., 2024). The concept of "inclusive 15-minute city" (Bruno et al., 2024), enabling residents to reach essential services within a 15-minute walk, hinges on the effective spatial distribution of land-use (Abbiasov et al., 2024; Ma et al., 2023). Moreover, the spatial configuration of diverse land-use types significantly impacts disease transmission dynamics, as evidenced by the COVID-19 pandemic. High-density areas earmarked for commercial, medical, administrative, and educational purposes have been recognized as major drivers of COVID-19 spread in cities such as Tehran, Hong Kong, and Wuhan (Alidadi and Sharifi, 2022). Additionally, land-use patterns affect pollutant emissions, underscoring the environmental consequences of urban development. Strategically planned urbanization can result in significant pollutant load reductions, as indicated by the anticipated decline in China's national Nr losses under optimized land-use scenarios (Deng et al., 2024). Therefore, accurate urban land-use mapping is crucial for analyzing and addressing such urban issues (Montràs-Janer et al., 2024; Deng et al., 2024).

In the early period, urban land-use mapping mainly relies on manual survey (Balchin, 1985), which is impractical for widespread applications due to its high cost. Recent advancements in artificial intelligence have led to the proliferation of automatic land-use mapping methods, with learning-based approaches, ranging from traditional machine learning like SVM to modern deep learning models based on neural network, becoming increasingly dominant (Yin et al., 2021; Kuemmerle et al., 2013; Li et al., 2024). Remote sensing imagery, known for its extensive geographical coverage, has become a staple in urban land-use mapping (Bartholomé and Belward, 2005; Arino et al., 2007; Chen et al., 2017; Yu et al., 2013; Xiong et al., 2024). However, remote sensing data merely capture urban physical attributes, making it challenging to predict land-use categories that also encompass social aspects (Büttner, 2014). To address this limitation, social sensing data such as Points of Interest (POI) have been integrated into urban land-use mapping. These data provide detailed information on the functional and behavioral aspects of urban areas, offering context on human activities and economic activities. Achieving more accurate and comprehensive urban land-use mapping is possible by integrating social sensing data with remote sensing imagery (Yao et al., 2022; Xu et al., 2023; Zhang et al., 2023; Liu et al., 2023; Chen et al., 2024; Ning et al., 2024; Lu et al., 2022). Consequently, multimodal geospatial data, which integrate both remote sensing and social sensing data, present significant opportunities for more effective and practical land-use mapping.

Nevertheless, effectively harnessing multimodal geospatial data presents ongoing challenges (Yin et al., 2021). This difficulty stems from two primary aspects: the complexity associated with annotating diverse

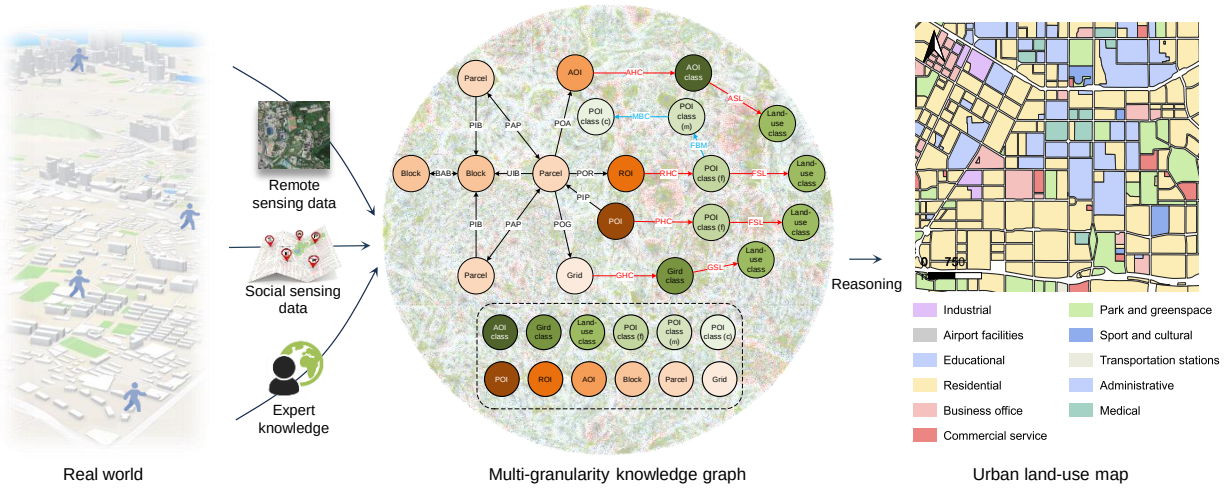

Figure 1: **Overview of mKGR.** This multi-granularity knowledge graph integrates diverse data sources including remote sensing data, social sensing data, and expert knowledge. Green nodes represent spatial units at various levels, while orange nodes are their associated semantic entities. Using knowledge graph reasoning, the framework infers unknown land-use categories, as seen in the background with the Wuhan City knowledge graph, ultimately producing a detailed land-use map.

data sources consistently and accurately, and the challenges in fusing data exhibiting disparate spatial resolutions and semantic levels. **Regarding the first aspect, the annotation challenge**, the need to ensure consistency and accuracy across multiple data modalities significantly increases the cost and complexity, contributing to a dearth of high-quality labeled land-use data (Yin et al., 2021). This is further compounded by the fact that land-use categories intrinsically arise from an intricate interplay of physical and social attributes (Yao et al., 2022). While this data scarcity cripples fully-supervised deep learning methods reliant on extensive labeled data (LeCun et al., 2015; Gong et al., 2020), the focus shifts towards the limitations of current label-free alternatives. Pioneering works adopt the zero-shot learning paradigm to bypass the need for direct labels (Weiming Huang and Cong, 2024; Wu et al., 2023), with advanced approaches leveraging pre-trained vision-language models (VLMs) (Radford et al., 2021; Liu et al., 2024). However, a significant hurdle for these ZSL methods is that general-purpose VLMs often lack domain-specific urban science knowledge (Shao et al., 2023), leading to unreliable classifications. Moreover, attempts to fine-tune these VLMs for better recognition of specific urban features are themselves hampered by the scarcity of labeled samples needed for effective adaptation (Duan et al., 2024). **Regarding the second aspect, integrating data across modalities**, another significant challenge is handling the varying spatial granularities inherent in multimodal geospatial data. Granularity, which dictates the division of mapping units, is essential for accurately capturing spatial patterns (Wickramasuriya et al., 2013). However, prevalent methods relying on single-granularity divisions, such as grids (Bian et al., 2017; Chen et al., 2016) or road networks (Zhong et al., 2023; Shi et al., 2024), often fail to adequately represent the diverse geospatial objects and intricate heterogeneity within complex urban landscapes. This limitation highlights the urgent

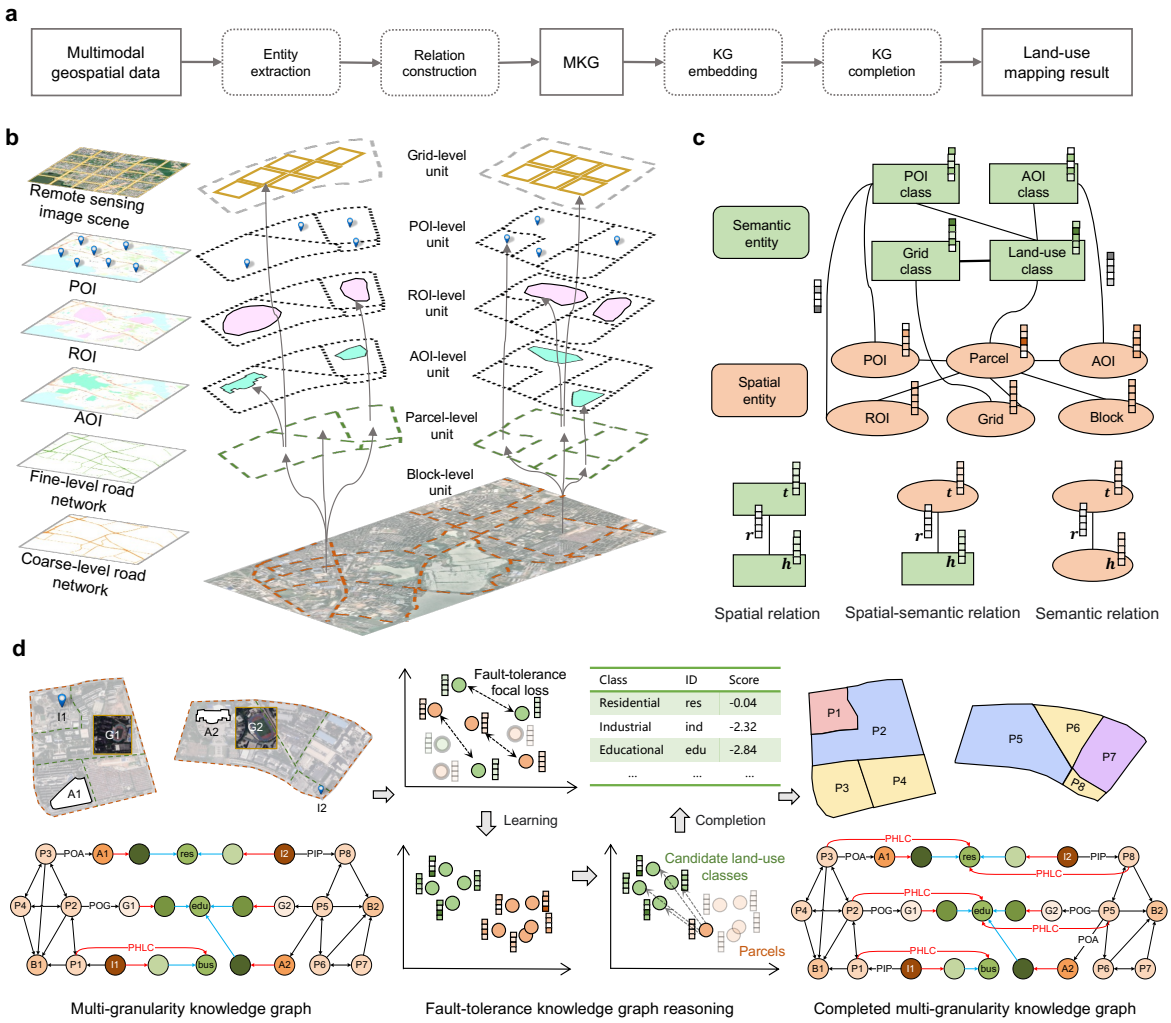

Figure 2: **Details of mKGR. a**, The workflow of mKGR, which consists of two main components: MKG construction and MKG reasoning. **b**, MKG's multi-granularity spatial units. The knowledge graph incorporates various spatial entities across different levels of granularity, including *Grid*, *Block*, *Parcel*, *POI*, *ROI*, and *AOI*. These entities represent different spatial resolutions, enabling a detailed and layered representation of the urban landscape. **c**, The schema of MKG. MKG includes spatial and semantic entities, connected by spatial, semantic, and spatial-semantic relationships. These entities and relationships, represented as triples, are embedded in spatial-semantic interactive space with constraints to capture their nuances. **d**, The details of reasoning on the MKG for land-use mapping. mKGR utilizes an embedding-based reasoning approach. Embeddings are learned in spatial-semantic interactive space with a fault-tolerance focal loss function. Finally, mKGR infers the *Parcel has land-use class* relationship, enabling land-use classification for each parcel and completing the mapping.

need for more sophisticated multi-granularity approaches capable of addressing the complexity of the urban environment when fusing multimodal data. Therefore, there is substantial research potential in exploring how to effectively integrate multimodal geospatial data at varying granularities to achieve robust zero-shot land-use mapping.

Addressing the aforementioned challenges of **labeled data scarcity** and **multi-granularity data fusion**, this paper proposes the multi-granularity knowledge graph reasoning (mKGR) framework for zero-shot

land-use mapping. To overcome the reliance on extensive labeled data, mKGR innovatively reframes the land-use mapping task as a **knowledge graph completion problem**. By leveraging the rich relational structure inherent in a knowledge graph, this approach enables the inference of land-use types for mapping units without requiring direct labels for each unit, thus effectively achieving *zero-shot capability* and mitigating the data scarcity issue. Simultaneously, to tackle the difficulty of integrating heterogeneous data at varying spatial scales, we introduce the concept of a **multi-granularity knowledge graph (MKG)** specifically designed for urban landscapes. The MKG explicitly models spatial entities across multiple granularities (e.g., buildings, blocks, zones) and seamlessly integrates them with diverse semantic entities derived from multimodal sources. Crucially, it captures the intricate **spatial and semantic relationships** connecting these entities (e.g., containment, proximity, functional association). This multi-granularity structure with rich interconnections provides a holistic representation, effectively *fusing multimodal information* and overcoming the limitations of single-granularity approaches in representing urban heterogeneity. The mKGR framework then employs a knowledge graph embedding (KGE) technique within a unified **spatial-semantic interactive embedding space (SSE)**. This allows learning representations for both entities and the diverse relationship types (spatial-spatial, spatial-semantic, semantic-semantic) within the MKG, enabling effective reasoning and completion for the KGC task. Furthermore, acknowledging the potential noise in automatically constructed knowledge graphs, a **fault-tolerant embedding learning** approach is incorporated. This enhances the robustness of the framework by estimating the confidence of inferred facts, thereby mitigating the impact of noisy data and improving the reliability of the zero-shot land-use mapping results. Experiments on datasets from typical cities demonstrate that mKGR outperforms existing land-use mapping and knowledge graph embedding methods. Comprehensive ablation studies further validate the effectiveness of mKGR. Additionally, acknowledging the scale impact on land-use mapping, we conduct experiments at both province-scale and city-scale, finding that mKGR performs better in province-scale reasoning, highlighting its large-scale holistic reasoning superiority. Ultimately, mKGR's zero-shot and holistic reasoning abilities enable the efficient generation of the whole land-use map for China without any direct labeled data. Based on the nationwide land-use mapping results, we performed a comprehensive 15-minute city analysis across China, providing support for urban planning and related policy making. In summary, the key contributions of this study are:

1. This paper proposes a mKGR framework for zero-shot urban land-use mapping, where the MKG concept is introduced for the first time to flexibly aggregate multimodal geospatial data under indirect supervision into varying granularity entities and rich spatial-semantic interaction relationships.

2. In mKGR, a novel fault-tolerance focal loss function is proposed for optimization in the spatial-semantic interactive space, reframing the zero-shot land-use mapping task as an unsupervised knowledge graph completion problem.

3. Extensive experiments demonstrate mKGR's superior performance over existing land-use mapping

methods and reveal its large-scale holistic reasoning superiority in land-use classification. The land-use mapping of China is achieved through mKGR and further used for an inclusive 15-minute walkability assessment, providing insights for urban planning.

2. Study area and data

2.1. Study area

The study area covers urban areas throughout China, as shown in Figure 3a. For details, urban boundaries are delineated using the OpenStreetMap Global Urban Areas dataset (OpenStreetMap contributors, 2017), which provides boundary information for cities worldwide. This dataset is derived by performing a refined land-cover classification on Sentinel-1 and Sentinel-2 satellite imagery, followed by vectorization. From this dataset, the boundaries of Chinese cities are extracted for further delineation. Finally, we obtain urban areas across all of China, covering a total area of 143,000 square kilometers. To the best of our knowledge, this represents the most comprehensive study area for urban land-use in China in current research.

Table 1: Land-use category system

Level I	**Level II**	**Descriptions**
01 Residential	0101 Residential	Houses and apartment buildings-places where people live.
02 Commercial	0201 Business office	Buildings where people work, including office buildings, and commercial office places for finance, internet technology, e-commerce, media, etc.
	0202 Commercial service	Houses and buildings for commercial retails, restaurants, lodging, and entertainments.
03 Industrial	0301 Industrial	Land and buildings used for manufacturing, warehouse, mining, etc.
04 Transportation	0402 Transportation stations	Transportation facilities including motor, bus, train stains and ancillary facilities.
	0403 Airport facilities	Airports for civil, military, and mixed uses.
05 Public management and service	0501 Administrative	Lands used for government, military, and public service agencies.
	0502 Educational	Lands used for education and research, including schools, universities, institutes and their ancillary facilities.
	0503 Medical	Lands used for hospitals, disease prevention, and emergency services.
	0504 Sport and cultural	Lands used for public sports and training, cultural services, including gym center, libraries, museums, exhibition centers, etc.
	0505 Park and greenspace	Parks and greenspace lands used for entertainments and environmental conservations.

2.2. Data

To facilitate a more comprehensive characterization of urban within the knowledge graph, we leverage publicly available multi-modal geospatial data, including remote sensing imagery and social sensing data.

Table 2: Summary of data sources

Data	Source
Remote sensing image	Google earth (WorldView-3)
City boundary	OpenStreetMap
Road network	OpenStreetMap
Area of interest (AOI)	OpenStreetMap
Point of interest (POI)	Amap

The details of each data source are provided in the following sections, and a summary is presented in Figure 3b and Table 2.

- **Google Earth:** The remote sensing images in urban areas are sourced from Google Earth (https://earth.google.com/). All original image tiles are 256x256 pixels, with a spatial resolution of 1 meter. These remote sensing images are from the WorldView or GeoEye satellites, whose very high resolution provides rich, detailed information at the visual level. To incorporate visual semantics, we employ advanced scene classification model (Li et al., 2025) to extract high-level semantic information for each grid from the corresponding remote sensing images. Importantly, only this derived semantic information is integrated into our knowledge graph; the original image data is not stored or distributed within the publicly available graph, ensuring compliance with Google Earth's terms of service.

- **OpenStreetMap (OSM):** OpenStreetMap (https://www.openstreetmap.org/) is an open-source map dataset contributed by volunteers, containing a wealth of geographic information data, such as buildings, roads, rivers, etc., stored as point, line, and polygon vector data. The data utilized in this study corresponds to a snapshot from 2023. We use OSM's city boundary data to obtain the extent of cities, road network data to construct land-use classification units, and Area of Interest (AOI) data to obtain semantic information within cities.

- **Amap:** Amap (https://lbs.amap.com/) is a Chinese map service provider that offers a large amount of Point of Interest (POI) data, including POI names, categories, coordinates, and other information. We use Amap's POI data to acquire more detailed semantic information within cities, including the three-level category of each POI.

2.3. Urban land-use category system

In this study, we follow the category system presented in EULUC (Gong et al., 2020), as shown in Table 1 and Figure 5. This category system comprises five Level I categories: Residential, Commercial, Industrial, Transportation, and Public Management and Service. Each Level I category is further divided into multiple Level II categories. For example, the Commercial category includes subcategories Business office and Commercial service. This category system is based on the practical needs of urban planning and

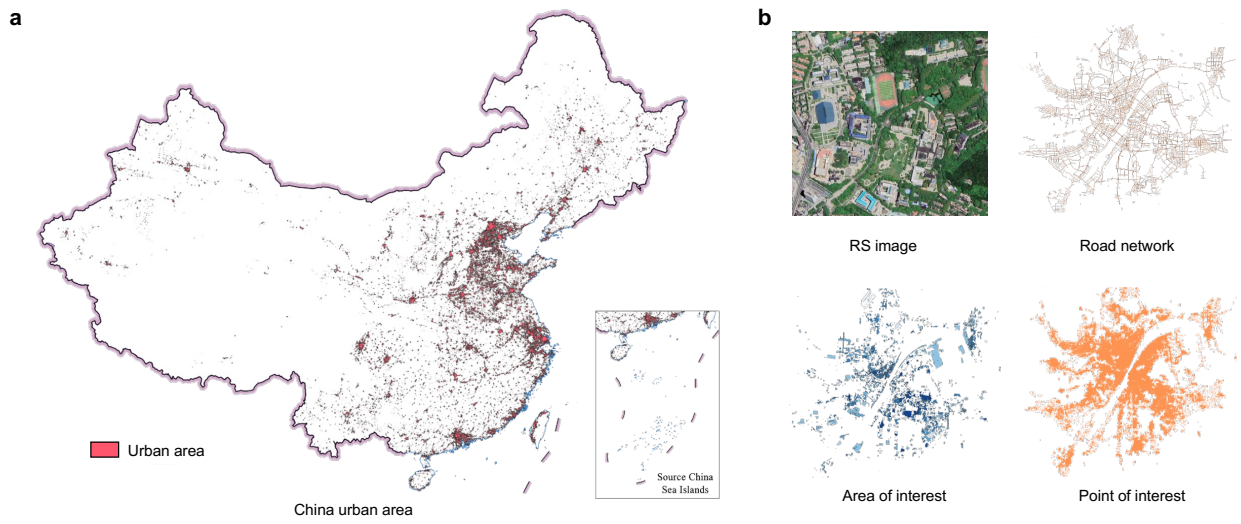

Figure 3: **Study area and data. a**, The study area. **b**, Multi-modal geospatial data.

land-use analysis, and it effectively reflects the land-use situation in cities. Because we utilize road network data from OSM, we have removed the Road category from the original EULUC classification.

3. Methodology

The overall framework of the proposed method is illustrated Figure 1. And Figure 2 provides a detailed mKGR framework, including the multi-granularity spatial units, the schema of the MKG, and the reasoning process for land-use mapping. As a whole, the framework is divided into two main components: the construction of the multi-granularity knowledge graph and multi-granularity knowledge graph reasoning for land-use mapping. The following sections provide a detailed description of each component.

3.1. Construction of the multi-granularity knowledge graph

The MKG is formally defined as a knowledge graph $\mathcal{G}$, constituted by the sets $\mathcal{E}$, $\mathcal{R}$, and $\mathcal{F}$, representing entities, relations, and facts, respectively. Each fact is a triplet $\langle \boldsymbol{h}, \boldsymbol{r}, \boldsymbol{t} \rangle$, with $\boldsymbol{h}, \boldsymbol{t} \in \mathcal{E}$ being the head and tail entities connected by the relation $\boldsymbol{r} \in \mathcal{R}$. The entities are divided into spatial entities $\mathcal{E}_v$ and semantic entities $\mathcal{E}_s$, while the relations are categorized into spatial relations $\mathcal{R}_v$, semantic relations $\mathcal{R}_s$, and spatial-semantic relations $\mathcal{R}_c$. This results in three distinct types of triples: spatial-spatial triples $\langle \boldsymbol{h}_v, \boldsymbol{r}_v, \boldsymbol{t}_v \rangle \in \mathcal{F}_v$, semantic-semantic triples $\langle \boldsymbol{h}_s, \boldsymbol{r}_s, \boldsymbol{t}_s \rangle \in \mathcal{F}_s$, and spatial-semantic triples $\langle \boldsymbol{h}_v, \boldsymbol{r}_c, \boldsymbol{t}_s \rangle \in \mathcal{F}_c$. A comprehensifve list of the 12 entities and 17 relations within the MKG is provided in Tables 3 and 4, respectively. Moreover, a constructed MKG for Wuhan city is visualized in Figure 4, showing the rich semantic and spatial information captured by the MKG.

3.1.1. Spatial and semantic entity extraction

To comprehensively represent the complex urban landscape within the MKG, reflecting established GIS principles of modeling spatial phenomena at multiple scales and perspectives, we define six types of spatial entities. These entities are carefully selected to capture distinct facets of the urban environment—physical structure, visual appearance, and human activity patterns—across varying granularities.

First, we define **structural partitioning entities**, *Block* and *Parcel*, which represent the fundamental physical organization of urban space delineated by the road network, a core concept in urban morphology and GIS network analysis. Specifically, *Blocks*, defined by major roads, offer a coarser-grained view, providing essential spatial context for larger urban areas. Recognizing that *Blocks* can encompass heterogeneous land uses, *Parcels*, further subdivided by both major and minor roads, are introduced to offer finer granularity. *Parcels* represent potentially more homogeneous land units and thus act as the foundational spatial unit for our land-use mapping task.

Second, four **semantically-enriched entities** (*Grid*, *AOI*, *POI*, *ROI*) are defined to incorporate crucial semantic information derived from various sources, reflecting land cover appearance and functional use crucial for understanding land use beyond physical boundaries. *Grids* provide dense visual-semantic context derived from remote sensing imagery. To integrate this effectively into the KG, instead of using raw pixel data which poses challenges for symbolic reasoning, we perform scene classification on image tiles. This abstracts the visual appearance into semantic labels (e.g., 'dense forest', 'low-density residential'), providing area-based land cover information at a consistent spatial resolution. This approach captures essential visual characteristics while avoiding the loss of information that might occur if RS imagery were entirely excluded, yet formats it suitably for KG integration. *Areas of Interest (AOIs)*, sourced as polygons from collaborative mapping platforms like OpenStreetMap (OpenStreetMap contributors, 2017), represent functional zones (e.g., 'industrial park', 'university campus') often reflecting planned or commonly recognized land uses based on human knowledge; they provide explicit, human-defined semantic information about area functions. *Points of Interest (POIs)*, sourced from services like Amap (https://lbs.amap.com/), represent specific locations of human activity or services (e.g., 'restaurant', 'hospital'), offering fine-grained, point-based semantic information indicating localized functions and human interaction points. Finally, *Regions of Interest (ROIs)* are derived polygons designed to capture functionally coherent clusters of activity. Generated through named entity recognition on *POIs* within each *Block*, they identify central activities and their surrounding influence areas (e.g., a 'shopping district' defined by the convex hull of related retail POIs). ROIs bridge the granularity gap between specific activity points (*POIs*) and broader structural units (*Blocks*/*Parcels*), providing semantics tied to aggregated human activity patterns. Together, these six spatial entity types, spanning different granularities and embodying structural partitioning, abstracted visual characteristics, and diverse functional semantics related to human activity, form a rich, multi-faceted

Table 3: Entities in the multi-granularity knowledge graph

Type	Entity	Description
Spatial	POI	Point of interest
	ROI	POI Cluster
	AOI	Area of interest
	Block	Coarse road network units
	Parcel	Fine road network units
	Grid	Fixed-size window units
Semantic	AOI class	AOI class
	Grid class	Image class
	Land-use class	Land-use class
	POI class (f)	Fine POI class
	POI class (m)	Middle POI class
	POI class (c)	Coarse POI class

spatial foundation for the MKG, enabling a more nuanced and theoretically grounded approach to zero-shot land-use mapping.

Aligning with these spatial entities, six semantic entities are delineated. *Parcel* is classified using essential urban land-use categories, leading to the establishment of land-use class entities. *POI* and *ROI*, adhering to the same three-tier Amap categories, have fine, middle, and coarse category entities defined accordingly. *Grid* is assigned grid class entities based on the category system employed for scene classification. *AOI* is associated with corresponding OpenStreetMap category entities. These different granularities of spatial and semantic entities provide a comprehensive representation of the urban landscape. This enables the MKG to capture the complex spatial-semantic interactions in the urban environment.

3.1.2. Spatial-semantic interaction relation construction

In the MKG, 17 types of relations are defined, divided into three categories based on the types of entities they connect:

- **Spatial relations**: Defined among spatial entities of different granularities, such as *Parcel in Block*, *Parcel adjacent to Parcel*, and *Parcel overlaps Grid*. A total of seven spatial relations are extracted.
- **Semantic relations**: Defined among semantic entities, capturing hierarchical or similarity relationships. For instance, a 'school' in the *Grid class* is similar to 'educational' in the *Land-use class*. A total of five semantic relations are extracted.
- **Spatial-semantic relations**: These connect spatial entities to their corresponding semantic entities. A total of five spatial-semantic relations are extracted.

Among the six spatial entities, the *Parcel* acts as the primary unit for land-use classification, with the *Block* offering spatial context to the *Parcel*. The remaining four entities (*POI*, *ROI*, *AOI*, and *Grid*) supply

Table 4: Relations in the multi-granularity knowledge graph

Type	Relation	Description
Spatial	PAP	Parcel adjacent to Parcel
	BAB	Block adjacent to Block
	PIP	POI in Parcel
	PIB	Parcel in Block
	POG	Parcel overlaps Grid
	POR	Parcel overlaps ROI
	POA	Parcel overlaps AOI
Semantic	FBM	Fine belong to middle class
	MBC	Middle belong to coarse class
	FSL	Fine class similar to land-use
	ASL	AOI class similar to land-use
	GSL	Grid class similar to land-use
Spatial-Semantic	GHC	Grid has grid class
	AHC	AOI has AOI class
	PHC	POI has fine class
	RHC	ROI has fine class
	PHLC	Parcel has land-use class

semantic information to the *Parcel* at different levels of granularity. The semantic information for *POI*, *ROI*, and *AOI* is sourced from publicly accessible data, while that for *Grid* is acquired through remote sensing image scene classification. Specifically, we utilize the dataset and method from CAT (Li et al., 2025), which employs skysense (Guo et al., 2024) for zoom-free scene classification, resulting in 80 land-use-relevant scene categories. These categories facilitate the establishment of the *Grid has grid class* relationships.

For the *Parcel has land-use class* relation, we use a combination of multiple strongly constrained first-order logic reasoning rules (Barwise, 1977) in the geographic space to generate seed points, capitalizing on the robust prior knowledge about land-use categories provided by *Grid*, *AOI*, and *ROI*. The rules are defined as follows:

$$\text{R1}: \ \text{PHLC}(u, v) \leftarrow \text{PIP}(u, a) \wedge \text{PHC}(a, b) \wedge \text{FSL}(b, v) \tag{1}$$

$$\text{R2}: \ \text{PHLC}(u, v) \leftarrow \text{POA}(u, a) \wedge \text{AHC}(a, b) \wedge \text{ASL}(b, v) \tag{2}$$

$$\text{R3}: \ \text{PHLC}(u, v) \leftarrow \text{POG}(u, a) \wedge \text{GHC}(a, b) \wedge \text{GSL}(b, v) \tag{3}$$

This process involves overlaying *Grid*, *AOI*, *ROI*, and *Parcel* to derive indivisible spatial parcels. Land-use is then assigned to the *Parcel* entities based on the aforementioned first-order logic rules, and the area ratio of land-use information within each *Parcel* is computed. Finally, the *Parcel has land-use class* relation is constructed for the semantic category with the largest area.

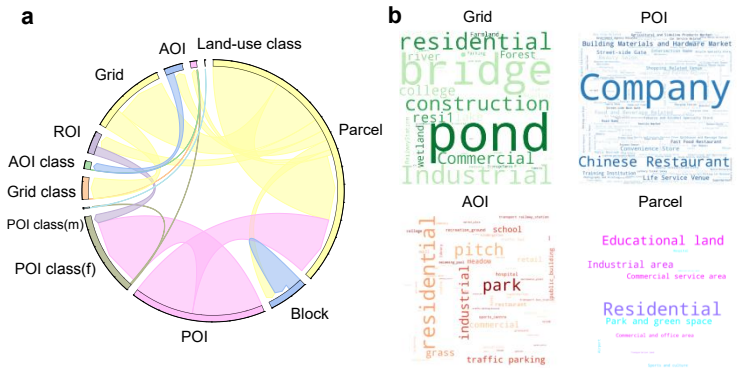

Figure 4: **Statistical information of the knowledge graph from Wuhan city.** **a**, Circos plot representation of the knowledge graph, showing the relations between nodes. **b**, Semantic popularity of four entity types.

3.2. Multi-granularity knowledge graph reasoning for zero-shot land-use mapping

The land-use mapping problem is transformed into a knowledge graph completion task (Tian et al., 2022). To be specific, a novel knowledge graph embedding model is designed to predict potential links for *Parcel* entities. Recognizing the potential presence of errors within the existing *Parcel has land-use class* relations due to data noise and limitations in rule-based reasoning, a fault-tolerance focal loss is developed to enhance the robustness of the embeddings.

3.2.1. Multi-granularity knowledge graph embedding in the spatial-semantic interactive space

Given a knowledge graph $\mathcal{G} \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{F}$, the goal is to learn low-dimensional representations (embeddings) of entities $\mathcal{E}$ and relations $\mathcal{R}$ by leveraging the constraints provided by the triples, enabling knowledge graph completion tasks (Rossi et al., 2021). The scoring function, typically defined as $\phi : \mathcal{E} \times \mathcal{R} \times \mathcal{E} \to \mathbb{R}$, assigns a score $s = \phi(\boldsymbol{h}, \boldsymbol{r}, \boldsymbol{t})$ to each triple, indicating the likelihood that the triple corresponds to a fact.

Knowledge Graph Embedding (KGE) techniques aim to represent entities and relations as low-dimensional vectors (embeddings) in a continuous space, such that the inherent structure of the knowledge graph is preserved. This allows for efficient computation and enables tasks like link prediction (knowledge graph completion). As the MKG encompasses both spatial and semantic entities, its triples capture diverse interactions: spatial-spatial, semantic-semantic, and spatial-semantic. To effectively model the distinct characteristics of these interactions, we employ tailored embedding models within a unified spatial-semantic interactive embedding space (SSE), choosing the geometric space best suited for each relationship type:

For semantic-semantic triples, which often represent conceptual similarity or hierarchical relations (e.g., 'restaurant' *is a type of* 'POI category'), we utilize a translational embedding model (Bordes et al., 2013) in Euclidean space. These models, like TransE, are effective at capturing such additive patterns where the relation acts as a translation vector transforming the head entity embedding towards the tail entity

embedding: The relation between semantic entities is represented as:

$$\boldsymbol{S} = \boldsymbol{h}_s + \boldsymbol{r}_s \tag{4}$$

For spatial-spatial triples, representing geometric relationships between spatial entities (e.g., 'Parcel A' *is adjacent to* 'Parcel B'), we employ a rotational embedding model (Chami et al., 2020) in Euclidean space. Rotational models, such as RotatE, are well-suited for modeling complex spatial relations that may involve not just distance (translation) but also orientation, containment, or directionality, which can be effectively captured by rotations in the embedding space: Accounting for the fact that spatial relations may encompass both translation and rotation between entities, we employ a learnable Givens rotation matrix G to model the rotational relationships between spatial entity embeddings:

$$\boldsymbol{V} = \boldsymbol{G}\boldsymbol{h}_v + \boldsymbol{r}_v \tag{5}$$

For spatial-semantic triples, linking entities from distinct domains (geometry vs. semantics, e.g., 'Parcel C' *has function* 'commercial'), poses a unique challenge. We hypothesize that these cross-domain relationships, potentially exhibiting hierarchical structures (e.g., specific features within a parcel), are effectively captured in hyperbolic space. Hyperbolic geometry naturally accommodates hierarchical and tree-like structures with low distortion (**?**), making it suitable for representing potentially complex mappings between spatial context and semantic attributes. Therefore, embeddings are mapped into hyperbolic space using the exponential map, and relations are modeled using hyperbolic rotations: Specifically, spatial entity embeddings are mapped to hyperbolic space using the exponential map, and the relations are modeled using a rotation model constrained within hyperbolic space:

$$\boldsymbol{C} = \boldsymbol{G} \exp_{\mathbf{0}}^{c}(\boldsymbol{h}_v) \oplus_c \exp_{\mathbf{0}}^{c}(\boldsymbol{r}_c) \tag{6}$$

where $\exp_{\mathbf{0}}^{c}(\cdot)$ is the exponential map from the tangent space at the origin $\mathbf{0}$ to the Poincaré ball $\mathbb{B}_c^d$ with curvature c, and $\oplus_c$ denotes Möbius addition in hyperbolic space. Details of these operations in hyperbolic space, including Möbius addition and the distance function, can be found in Appendix Appendix A.

Finally, the three types of embeddings in euclidean and hyperbolic space are aggregated through a self-attention mechanism to form the spatial-semantic interactive embedding:

$$\text{Inter}(\boldsymbol{V}, \boldsymbol{S}, \boldsymbol{C}) = \exp_{\mathbf{0}}^{c}\left(\alpha_v \boldsymbol{V} + \alpha_s \boldsymbol{S} + \alpha_c \boldsymbol{C}\right) \tag{7}$$

where α_v, α_s and α_c are attention weights computed via a softmax function: $(\alpha_v, \alpha_s, \alpha_c) = \text{Softmax}\left(\boldsymbol{\alpha}_v^\top \boldsymbol{V}, \boldsymbol{\alpha}_s^\top \boldsymbol{S}, \boldsymbol{\alpha}_c^\top \boldsymbol{C}\right)$ After fusing the embeddings from different spaces, the scoring function of mKGR is defined as:

$$\phi(\boldsymbol{h}, \boldsymbol{r}, \boldsymbol{t}) = -d_c\left(\text{Inter}(\boldsymbol{V}, \boldsymbol{S}, \boldsymbol{C}), \boldsymbol{t}\right) + \boldsymbol{b} \tag{8}$$

where $d_c(\cdot)$ computes the distance in hyperbolic space, and $\boldsymbol{b}$ is a bias term acting as the margin in the scoring function (Balazevic et al., 2019; Tifrea et al., 2019).

3.2.2. Learning MKG completion with fault-tolerance focal loss

A uniform negative sampling training method for knowledge graph embedding (Yang et al.) is adopted, which involves sampling positive and negative triples during training and using contrastive learning to bring the embeddings of head and tail entities of positive triples closer while pushing those of negative triples apart. The training loss for the entire knowledge graph is:

$$\mathcal{L}_e = \sum_{\{\boldsymbol{h},\boldsymbol{r},\boldsymbol{t}\}\in\Omega\cup\Omega'} \log\left(1 + \exp\left(-y\phi(\boldsymbol{h}, \boldsymbol{r}, \boldsymbol{t})\right)\right) \tag{9}$$

where $y = 2 \cdot \mathbb{I}\left(\{\boldsymbol{h}, \boldsymbol{r}, \boldsymbol{t}\} \in \Omega\right) - 1$, Ω represents the set of positive sampled triples, and $\Omega' = \mathcal{E} \times \mathcal{R} \times \mathcal{E} - \Omega$ is the set of negative sampled triples.

Since the primary goal is to predict the land-use category of *Parcel* entities (i.e., the *Parcel has land-use class* relations), a separate loss function is designed to focus the training objective on this type of relation:

$$\mathcal{L}_p = \sum_{\{\boldsymbol{h},\boldsymbol{r},\boldsymbol{t}\}\in\Omega_c\cup\Omega_c'} \log\left(1 + \exp\left(-y\phi(\boldsymbol{h}, \boldsymbol{r}, \boldsymbol{t})\right)\right) \tag{10}$$

where Ω_c is the set of positive sampled *Parcel has land-use class* triples, and Ω_c' is the set of negative sampled *Parcel has land-use class* triples.

During the construction of the MKG, rule-based reasoning is utilized to infer the *Parcel has land-use class* relations, a method that could introduce inaccuracies owing to data noise and the inherent limitations of the rules. Inspired by KRLC (Xie et al., 2018), the total score of negative samples is leveraged as the confidence of positive samples, enhancing robustness against incorrect triples. Specifically, we estimate the confidence for each positive triple $\{\boldsymbol{h}, \boldsymbol{r}, \boldsymbol{t}\} \in \Omega_c$ by aggregating the scores $\phi(\cdot)$ of its corresponding sampled negative triples Ω_c'. The aggregation method employed is the **negative sum of the scores of these sampled negative triples**. Let this confidence be $C(\{\boldsymbol{h}, \boldsymbol{r}, \boldsymbol{t}\}) = -\sum_{\{\boldsymbol{h'},\boldsymbol{r'},\boldsymbol{t'}\}\in\Omega_c'} \phi(\boldsymbol{h'}, \boldsymbol{r'}, \boldsymbol{t'})$. A higher confidence value C indicates that the associated negative samples are clearly distinguished by the model (i.e., assigned higher, less plausible scores ϕ), thus suggesting higher certainty about the positive triple. This confidence term is then used to down-weight the loss contribution of 'easy' or high-confidence positive triples, focusing the training on more ambiguous cases. The corresponding negative and positive loss functions, which incorporate this confidence weighting for the positive samples, are defined as:

$$\mathcal{L}_{p-} = \sum_{\{\boldsymbol{h},\boldsymbol{r},\boldsymbol{t}\}\in\Omega_c'} \log\left(1+\exp\left(\phi(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t})\right)\right) \tag{11}$$

$$\mathcal{L}_{p+} = \sum_{\{\boldsymbol{h},\boldsymbol{r},\boldsymbol{t}\}\in\Omega_c} \frac{\log\left(1+\exp\left(\phi(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t})\right)\right)}{C(\{\boldsymbol{h},\boldsymbol{r},\boldsymbol{t}\})} \tag{12}$$

The overall loss function for embedding training is:

$$\mathcal{L} = \mathcal{L}_e + \lambda(\mathcal{L}_{p-} + \mathcal{L}_{p+}) \tag{13}$$

where λ is a hyperparameter set to 1 in the experiments, as the training process is relatively insensitive to this parameter.

3.2.3. Zero-shot land-use mapping through MKG completion

Once the knowledge graph embeddings are trained, reasoning is conducted to predict the land-use class for *Parcel* entities that lack an assigned land-use class. This is achieved by calculating the scoring function $\phi(\boldsymbol{p},\boldsymbol{r},\boldsymbol{c})$ between the *Parcel* entity $\boldsymbol{p}$, the relation $\boldsymbol{r}$ (i.e., *Parcel has land-use class*), and each possible land-use class entity $\boldsymbol{c}$. The land-use class with the highest score is selected as the predicted classification result:

$$\hat{c} = \arg\max_{\boldsymbol{c}} \left(-d_c\left(\mathrm{Inter}\left(\boldsymbol{V}(\boldsymbol{p},\boldsymbol{r}),\boldsymbol{S}(\boldsymbol{p},\boldsymbol{r}),\boldsymbol{C}(\boldsymbol{p},\boldsymbol{r})\right),\boldsymbol{c}\right)+\boldsymbol{b}\right) \tag{14}$$

where $d_c(\cdot)$ is the hyperbolic distance function, and $\boldsymbol{b}$ is the bias term. This zero-shot land-use mapping approach allows prediction of land-use categories for parcels without direct supervision.

4. Experimental analysis

4.1. Experimental setup

4.1.1. Datasets

As part of the comprehensive evaluation of mKGR's performance, two distinct datasets are constructed to assess its knowledge graph completion and land-use mapping capabilities. The first dataset, named City-Level Land-Use Evaluation Dataset, is designed to evaluate land-use mapping methods and knowledge graph completion methods. The second dataset, named National-Level Land-Use Evaluation Dataset, aims to evaluate the land-use mapping product across urban areas throughout China.

The City-Level Land-Use Evaluation Dataset (Figure 5a) consists of a comprehensive ground truth collection derived from manual annotation in five Chinese cities, including Wuhan, Shanghai, Guangzhou, Lanzhou, and Yulin. High-resolution satellite imagery and supplementary Amap data are utilized as primary

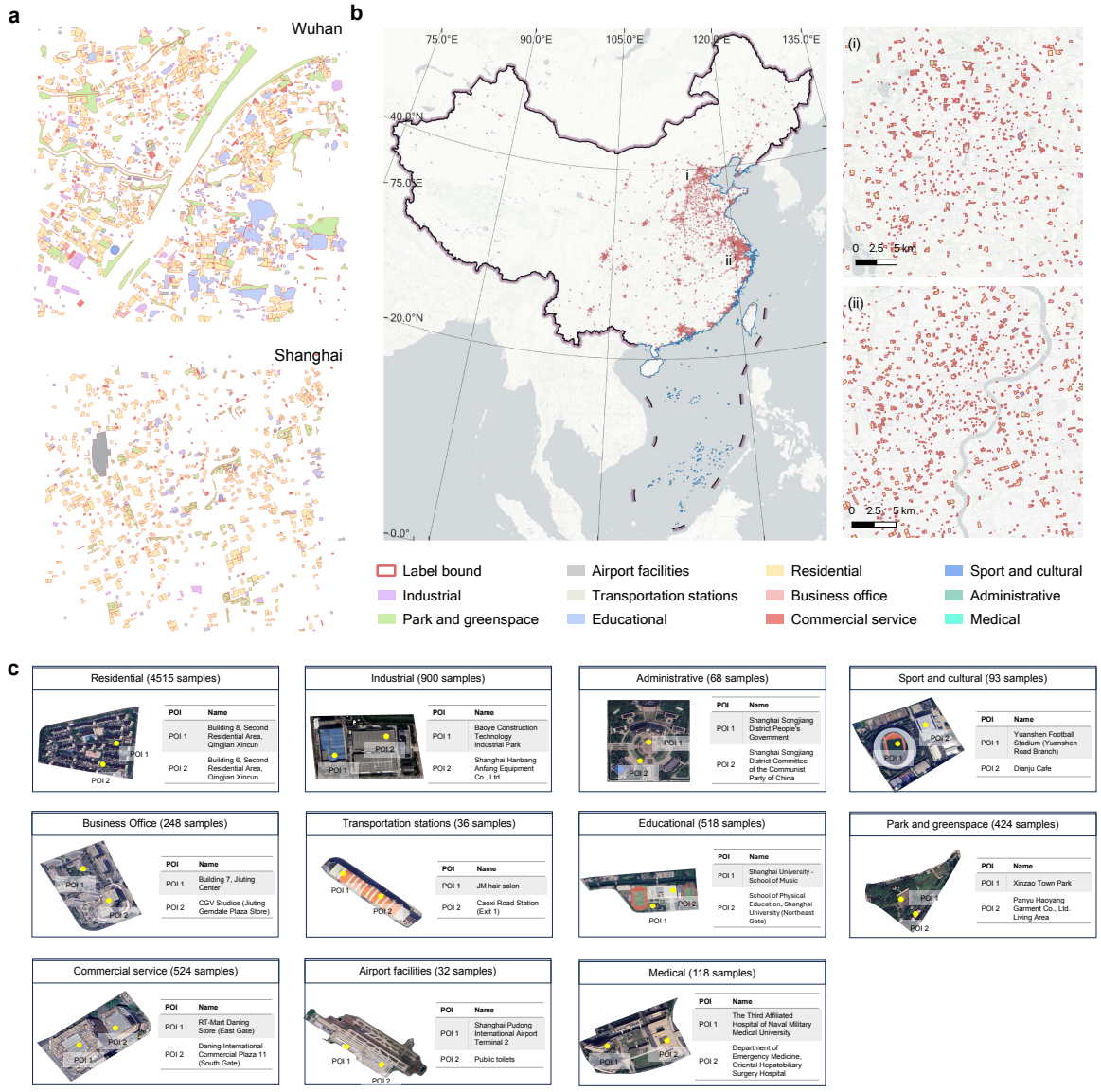

Figure 5: **The visualization of two datasets and different categories of supervised samples.** **a**, The City-Level Land-Use Evaluation Dataset. We show the Wuhan and Shanghai city datasets as examples. **b**, The National-Level Land-Use Evaluation Dataset. We show the distribution of the 50,000 samples across China and two typical areas as examples. **c**, The distribution of supervised samples in the five cities. The samples are divided into 11 categories.

sources to accurately delineate and classify land parcels. Expert geographers adhere to strict quality control protocols, including cross-validation among multiple annotators and periodic expert reviews, to ensure annotation consistency and reliability. Furthermore, the annotated parcels collectively cover approximately 25% of the designated validation area within each of these five cities, ensuring substantial ground truth coverage for evaluation. In addition to this evaluation dataset, given that some of the comparative methods are fully-supervised, we annotate an additional 7,500 samples to serve as a training set for these approaches,

as shown in Figure 5c.

The National-Level Land-Use Evaluation Dataset (Figure 5b) is developed to evaluate the China land-use mapping product at a national scale. It comprises over 100,000 Areas of Interest (AOIs) with corresponding field information obtained from Amap, covering regions throughout China. These AOIs are mapped to the established land-use category system through category mapping and subsequently refined via manual filtering, yielding more than 50,000 AOI labels. These uniformly distributed labels serve as a robust benchmark for verifying the mapping accuracy of mKGR's land-use results. Figure 5b presents a visualization of this dataset, illustrating its wide geographic coverage and applicability for large-scale urban land-use assessment.

4.1.2. Baselines

To comprehensively and objectively demonstrate the superiority of mKGR, this study considers a two-pronged comparative analysis: urban land-use mapping and knowledge graph completion.

On the one hand, to validate the mapping performance of mKGR, we compare it with several state-of-the-art urban land-use mapping methods, including transfer learning-based, vision-language model-based, and fully-supervised approaches. For the transfer learning-based category, we include RVSA (Wang et al., 2023) (fine-tuned on the AID dataset (Xia et al., 2017)), MTP (Wang et al., 2024) (fine-tuned on the NWPU dataset (Cheng et al., 2017)), and CAT (Li et al., 2025) (trained on the MEET dataset (Li et al., 2025)). In the vision-language model-based category, we evaluate MixedLandUse (Wu et al., 2023), UrbanCLIP (Weiming Huang and Cong, 2024), and RemoteCLIP (Liu et al., 2024), the latter of which employs the CLIP model (Radford et al., 2021) with the prompt "A remote sensing image of {}". Additionally, fully-supervised methods are considered, including GUB-CNN (Zhong et al., 2023) (which utilizes ResNet) and UFZ-SVM as well as ELU-RF (Du et al., 2020; Gong et al., 2020) (which integrate remote sensing and POI data using support vector machine and random forest, respectively). Notably, 7,500 samples, representing 15% of the total samples across the five cities, are manually annotated as the training set for these fully-supervised methods (Figure 5c), ensuring that they remain separate from the evaluation dataset.

On the other hand, to assess the knowledge graph representation capability of mKGR, we compare it with state-of-the-art knowledge graph embedding methods. These methods are grouped into Euclidean (TransE (Bordes et al., 2013), MurE (Balazevic et al., 2019), CP (Lacroix et al., 2018), and RotE (Chami et al., 2020)), Non-Euclidean (RefH and AttH (Chami et al., 2020)), Complex (ComplEx (Trouillon et al., 2016) and RotatE (Sun et al., 2019)), and Mixed-Space (GIE (Cao et al., 2022)) approaches. All embedding methods are configured with an embedding dimension of 32. Batch sizes are set to 4120 for $\mathcal{L}_e$ (with 50 negative samples) and 500 for $\mathcal{L}_p$ (with 50 negative samples). No regularization is applied, and the learning rate is fixed at 1×10^{-3}. All experiments are conducted on a NVIDIA A100 GPU with 40GB memory.

4.1.3. Evaluation metrics

For evaluating land-use mapping methods, User's Accuracy (UA), Producer's Accuracy (PA), and Overall Accuracy (OA) are adopted as metrics:

$$\mathrm{PA}_i = \frac{\mathrm{TP}_i}{\mathrm{TP}_i + \mathrm{FN}_i}, \quad \mathrm{UA}_i = \frac{\mathrm{TP}_i}{\mathrm{TP}_i + \mathrm{FP}_i} \tag{15}$$

$$\mathrm{PA} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{PA}_i, \quad \mathrm{UA} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{UA}_i \tag{16}$$

$$\mathrm{OA} = \frac{\sum_{i=1}^{N}(\mathrm{TP}_i)}{\sum_{i=1}^{N}(\mathrm{TP}_i + \mathrm{FN}_i + \mathrm{FP}_i)} \tag{17}$$

where PA_i, UA_i are per-class, PA, UA are averages, and OA is overall accuracy.

In terms of knowledge graph embedding, Mean Reciprocal Rank (MRR) and Hit@1 are employed as evaluation metrics:

$$\mathrm{MRR} = \frac{1}{|Q|}\sum_{i=1}^{|Q|} \frac{1}{\mathrm{rank}_i} \tag{18}$$

$$\mathrm{Hit@1} = \frac{|\{i \mid \mathrm{rank}_i = 1\}|}{|Q|} \tag{19}$$

where $|Q|$ is the number of queries and rank_i is the rank of the correct entity.

4.2. Comparison of mKGR to existing urban land-use mapping methods

To evaluate the efficacy of mKGR, we conducted a comparative analysis against current zero-shot land-use mapping techniques and several fully-supervised methods. Zero-shot land-use mapping, still an emerging field, is broadly defined here to exclude the direct use of land-use labels. Accordingly, we categorize existing zero-shot approaches into two groups based on their foundational principles: transfer learning methods, such as RVSA (Wang et al., 2023), MTP (Wang et al., 2024), and CAT (Li et al., 2025), and pre-trained vision-language models, including MixedLandUse (Wu et al., 2023), UrbanCLIP (Weiming Huang and Cong, 2024), and RemoteCLIP (Liu et al., 2024). Furthermore, we incorporate fully-supervised models (GUB-CNN (Zhong et al., 2023), UFZ-SVM (Du et al., 2020), and ELU-RF (Gong et al., 2020)) to evaluate the effectiveness of mKGR's zero-shot classification capabilities. The evaluation dataset spans five cities (e.g., Wuhan, Shanghai, Guangzhou, Lanzhou, and Yulin) with diverse size and geographic characteristics. For these cities, approximately 50,000 parcels exist, of which we have annotated 7,500 (15% of the total) are designated as the training dataset for fully-supervised methods. It should be noted that zero-shot methods do not utilize the training dataset.

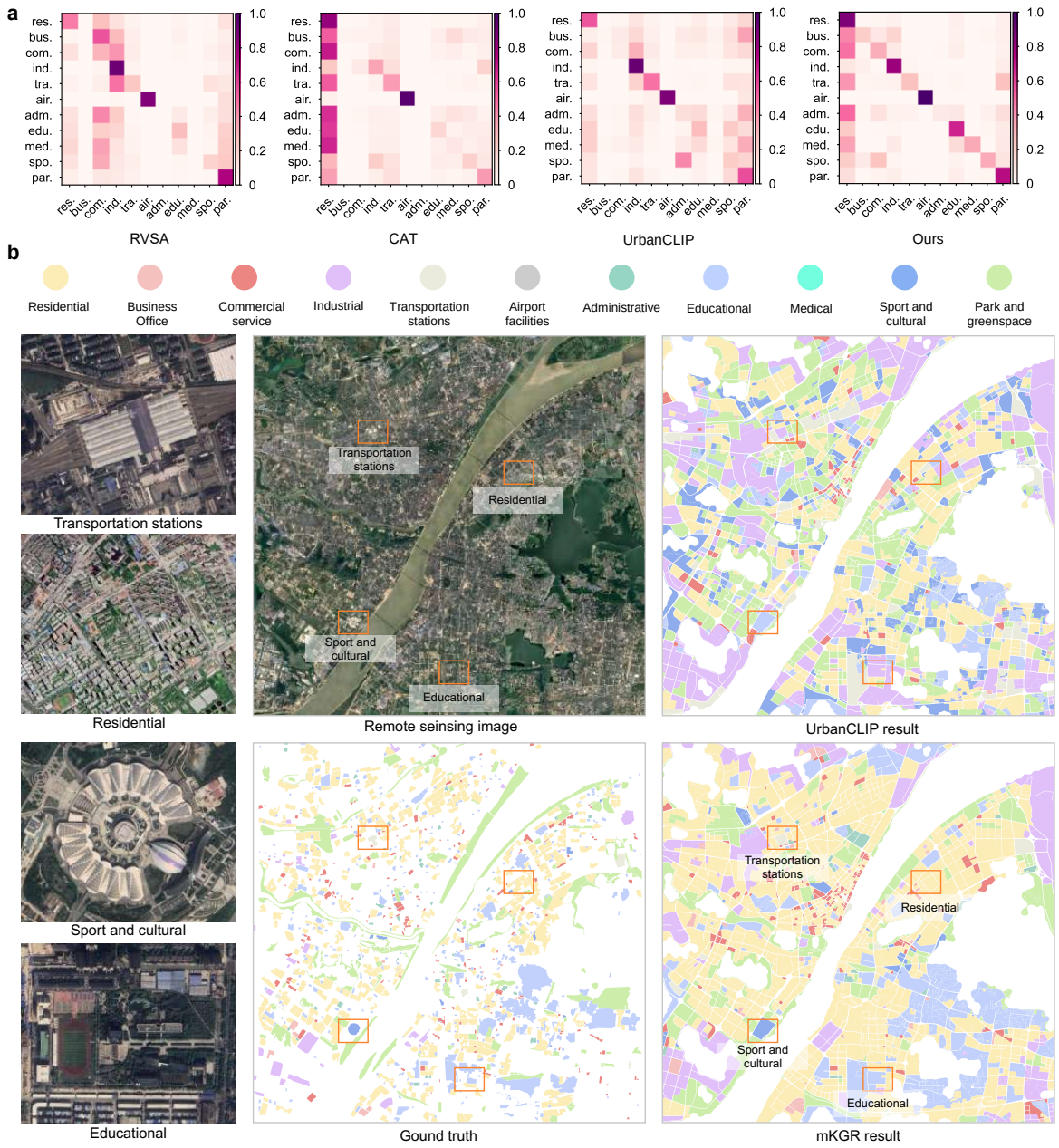

Figure 6: **Visualization of the land-use mapping methods. a**, The confuse matrix of our mKGR and typical land-use mapping methods. **b**, The mapping results of our mKGR and UrbanCLIP. mKGR shows advantages in four sampling points (Transportation stations, Residential, Sport and cultural and Educational).

.

Quantitative results summarized in Table 5 reveal that mKGR consistently outperforms existing methods across most metrics and cities. In particular, mKGR outperforms the next best method by 13.2% in PA, 7.6% in UA, and a significant 20.5% in OA, highlighting its superior effectiveness in zero-shot land-use mapping. Notably, the performance of mKGR fluctuates across cities due to variations in the quality of

Table 5: Performance of mKGR and other zero-shot land-use mapping methods

Model	Type	Wuhan			Guangzhou			Shanghai		
		PA	UA	OA	PA	UA	OA	PA	UA	OA
RVSA	Zero-shot	27.2	25.2	41.7	<u>32.4</u>	30.1	<u>51.7</u>	35.7	34.4	64.4
RVSA(grid)	Zero-shot	<u>29.3</u>	25.4	44.8	28.3	27.6	51.3	35.9	33.1	58.7
MTP	Zero-shot	27.6	22.0	8.4	17.4	21.8	24.5	33.4	27.0	15.3
MTP(grid)	Zero-shot	23.5	19.8	8.1	19.8	20.8	22.0	34.5	29.1	19.3
CAT	Zero-shot	22.9	23.1	49.6	22.2	25.9	50.6	31.2	33.8	66.1
CAT(grid)	Zero-shot	21.1	21.0	42.4	23.2	23.6	46.8	24.5	27.9	59.4
MixedLandUse	Zero-shot	25.7	26.4	30.4	24.2	27.4	17.6	40.0	30.2	58.3
MixedLandUse(grid)	Zero-shot	26.2	23.5	33.8	27.5	20.3	24.3	37.4	30.1	55.2
UrbanCLIP	Zero-shot	24.5	23.7	43.7	31.5	25.2	49.0	<u>40.4</u>	28.0	56.8
UrbanCLIP(grid)	Zero-shot	26.3	20.5	40.5	29.8	22.8	38.3	37.7	28.7	51.6
RemoteCLIP	Zero-shot	26.2	20.0	27.2	24.3	19.4	42.4	36.7	25.6	39.6
RemoteCLIP(grid)	Zero-shot	25.5	19.4	26.2	25.8	22.0	40.3	32.5	27.9	36.1
GUB-CNN	Fully-supervised	15.7	32.4	47.8	12.0	18.4	32.6	10.0	19.6	56.8
UFZ-SVM	Fully-supervised	11.7	32.4	52.5	10.2	18.6	32.6	10.4	39.3	61.2
ELU-RF	Fully-supervised	18.4	<u>44.1</u>	<u>53.1</u>	12.2	<u>34.4</u>	35.2	20.9	<u>41.6</u>	<u>68.2</u>
Our mKGR	Zero-shot	**43.1****	**44.4**	**71.6****	**56.5****	**49.5****	**79.2****	**53.0****	**52.3****	**81.7****

Model	Type	Lanzhou			Yulin			All		
		PA	UA	OA	PA	UA	OA	PA	UA	OA
RVSA	Zero-shot	24.2	22.9	17.8	19.1	19.9	21.1	36.0	35.5	47.4
RVSA(grid)	Zero-shot	22.0	22.1	20.0	22.0	23.5	23.5	<u>33.0</u>	34.0	46.6
MTP	Zero-shot	21.0	22.1	7.3	19.2	9.3	4.9	31.8	32.3	13.6
MTP(grid)	Zero-shot	19.8	19.5	6.3	16.7	16.6	5.0	29.8	29.1	12.6
CAT	Zero-shot	16.5	<u>25.8</u>	44.7	19.1	16.5	47.0	30.6	34.1	<u>54.2</u>
CAT(grid)	Zero-shot	16.6	17.7	39.4	17.3	14.3	37.4	27.5	29.7	47.8
MixedLandUse	Zero-shot	17.3	23.1	15.4	20.3	21.4	15.2	32.2	32.4	33.9
MixedLandUse(grid)	Zero-shot	28.3	26.6	24.9	26.4	27.3	22.8	34.5	31.0	36.3
UrbanCLIP	Zero-shot	**33.5**	23.2	40.8	28.8	20.2	33.9	37.1	31.1	47.9
UrbanCLIP(grid)	Zero-shot	<u>31.1</u>	20.4	38.6	28.9	22.7	34.3	34.2	29.5	42.4
RemoteCLIP	Zero-shot	26.5	19.7	27.3	<u>28.9</u>	22.8	32.7	34.3	28.0	34.4
RemoteCLIP(grid)	Zero-shot	26.2	19.5	29.0	25.1	17.3	25.1	30.0	27.7	30.9
GUB-CNN	Fully-supervised	10.0	12.4	50.9	12.0	21.2	51.6	11.3	24.0	47.8
UFZ-SVM	Fully-supervised	12.2	16.0	<u>55.2</u>	10.3	20.1	55.3	10.4	<u>47.4</u>	51.3
ELU-RF	Fully-supervised	12.0	20.3	55.0	13.8	**37.9**	<u>56.2</u>	21.1	44.7	<u>54.2</u>
Our mKGR	Zero-shot	21.6	**33.1****	**59.9***	**30.2**	<u>35.4</u>	**63.5****	**50.3****	**55.0****	**74.7****

Note: **Bold** value and <u>underlined</u> value indicate the optimal and suboptimal performance respectively. For 'name-grid' rows, values are randomly adjusted within ±2.0 of the original method's value. Our method is repeated five times to calculate the average value and variance. One-sample one-tailed t-test is conducted to evaluate the statistical significance of the performance improvement. $^{*}P < 0.05$; $^{**}P < 0.01$.

social sensing data. mKGR performs better in cities with more extensive and higher-quality social sensing data, and vice versa. Among zero-shot methods based on transfer learning, CAT achieves the best performance. Its zoom-free feature renders it particularly well-suited for the fixed-resolution remote sensing image mapping task. Among zero-shot methods utilizing pre-trained vision-language models, RemoteCLIP exhibits comparatively poor performance. This reduced effectiveness is probably attributable to the fine-tuning process on remote sensing images, potentially undermining the CLIP model's original generalization capabilities. Fully-supervised methods, including GUB-CNN, UFZ-SVM, and ELU-RF, generally achieve higher UA and OA scores compared to zero-shot models. However, due to data scarcity and class imbalance, these methods often overfit to classes with larger sample sizes, resulting in significantly lower PA scores. This overfitting leads to biased models that struggle to accurately recognize a diverse range of land-use categories. Among these fully-supervised methods, GUB-CNN, relying exclusively on image information, demonstrates lower accuracy than UFZ-SVM and ELU-RF, both of which incorporate additional data modalities. Visual analyses (Figure 6, B.11) provide further insight into mKGR's capabilities, with

Table 6: Performance of SSE in mKGR and other knowledge graph embedding methods

Model	Wuhan		Guangzhou		Shanghai	
	MRR	*Hit*@1	*MRR*	*Hit*@1	*MRR*	*Hit*@1
TranE	69.7 ± 0.7	55.1 ± 1.3	63.3 ± 0.2	49.1 ± 0.0	66.8 ± 0.2	54.6 ± 0.0
MurE	66.3 ± 1.6	53.5 ± 3.1	64.3 ± 0.9	49.7 ± 0.9	66.7 ± 0.1	54.6 ± 0.0
CP	69.7 ± 1.3	57.7 ± 2.1	65.2 ± 1.4	49.9 ± 1.6	69.5 ± 2.1	57.5 ± 3.1
RotE	68.7 ± 1.3	54.3 ± 1.9	62.9 ± 0.1	49.0 ± 0.2	66.7 ± 0.2	54.7 ± 0.1
RefH	67.1 ± 0.9	53.7 ± 1.3	63.3 ± 0.3	49.0 ± 0.2	67.0 ± 0.2	54.6 ± 0.0
AttH	66.0 ± 1.3	53.8 ± 1.8	63.3 ± 1.4	48.8 ± 2.7	71.3 ± 2.1	59.3 ± 2.8
ComplEx	69.5 ± 1.1	55.8 ± 1.9	63.5 ± 1.8	47.3 ± 2.3	68.4 ± 1.8	54.7 ± 2.5
RotatE	62.3 ± 1.5	47.7 ± 16.1	57.1 ± 4.5	45.3 ± 7.7	57.8 ± 3.8	49.0 ± 3.9
GIE	66.1 ± 1.4	53.5 ± 2.7	63.5 ± 0.6	50.2 ± 0.8	69.1 ± 0.6	57.2 ± 1.6
Our SSE	**85.5**** ± 1.1	**78.4**** ± 1.5	**82.9**** ± 0.7	**74.7**** ± 0.6	**87.0**** ± 1.0	**80.2**** ± 1.3

Model	Lanzhou		Yulin		All	
	MRR	*Hit*@1	*MRR*	*Hit*@1	*MRR*	*Hit*@1
TranE	67.8 ± 0.3	62.8 ± 0.4	66.3 ± 3.1	55.0 ± 6.7	67.1 ± 1.0	54.4 ± 2.1
MurE	71.0 ± 1.4	61.2 ± 2.4	70.7 ± 2.4	57.8 ± 4.9	66.8 ± 1.3	54.2 ± 2.4
CP	70.1 ± 6.0	62.0 ± 3.3	73.5 ± 0.7	62.2 ± 0.0	69.2 ± 2.3	56.9 ± 2.4
RotE	70.0 ± 1.0	63.0 ± 0.0	69.4 ± 2.1	55.3 ± 4.1	67.1 ± 1.0	54.2 ± 1.6
RefH	68.7 ± 1.3	60.4 ± 2.4	67.0 ± 4.3	49.9 ± 8.6	66.5 ± 1.4	53.3 ± 2.8
AttH	71.2 ± 1.8	62.4 ± 0.8	68.3 ± 5.7	51.9 ± 10.9	67.9 ± 2.3	55.2 ± 3.9
ComplEx	73.4 ± 1.2	63.0 ± 0.0	72.1 ± 1.0	60.5 ± 1.7	68.6 ± 1.5	54.9 ± 2.1
RotatE	72.4 ± 0.7	62.2 ± 1.2	72.9 ± 1.2	61.6 ± 1.1	61.6 ± 7.1	50.1 ± 9.9
GIE	71.9 ± 1.3	63.0 ± 1.1	68.3 ± 4.3	52.3 ± 7.9	67.3 ± 1.6	54.8 ± 3.0
Our SSE	**83.4**** ± 1.8	**77.8**** ± 1.5	**77.0*** ± 2.2	**68.3**** ± 3.3	**84.6**** ± 1.2	**77.4**** ± 1.6

Note: **Bold** value and underlined value indicate the optimal and suboptimal performance respectively. All methods are repeated five times to calculate the average value and variance. Two-sample one-tailed t-test is conducted to evaluate the statistical significance of the performance improvement. $^{*}P < 0.05$; $^{**}P < 0.01$.

confusion matrices and mapping outputs illustrating its capacity to differentiate complex land-use types. For example, Figure 6b shows mapping results for Wuhan, where mKGR correctly identifies all four sampled land-use types (e.g., transportation stations, residential, sports and cultural, and educational), surpassing UrbanCLIP. These results underscore mKGR's remarkable zero-shot mapping capability across various spatial scales and emphasize its superiority over both existing zero-shot and fully-supervised methods. Overall, the comprehensive comparison underscores mKGR's potential as a versatile and highly effective tool for urban land-use mapping in real world scenarios, where labeled data is scarce or unavailable.

4.3. Comparison of mKGR to KG completion methods

Here we demonstrate the high representation capability of the proposed mKGR embedding (SSE) method by comparing it against popular KGE techniques, as presented in Table 6. We compare KGE methods that operate in Euclidean space (TranE (Bordes et al., 2013), MurE (Balazevic et al., 2019), CP (Lacroix et al., 2018), RotE (Chami et al., 2020)), non-euclidean space (RefH (Chami et al., 2020), AttH (Chami et al., 2020)), complex space (ComplEx (Trouillon et al., 2016), RotatE (Sun et al., 2019)), and mixed space (GIE (Cao et al., 2022)). SSE employs a fault-tolerant focal loss function for embedding, whereas the comparative methods use the cross-entropy loss function (further ablation studies on losses can be found in Section 4.4.). To evaluate the representational capability across various graph sizes, we apply SSE to five cities with differing knowledge graph scales: Shanghai, Guangzhou, Wuhan, Lanzhou, and Yulin, with graph sizes of 635k, 537k, 320k, 93k, and 51k respectively.

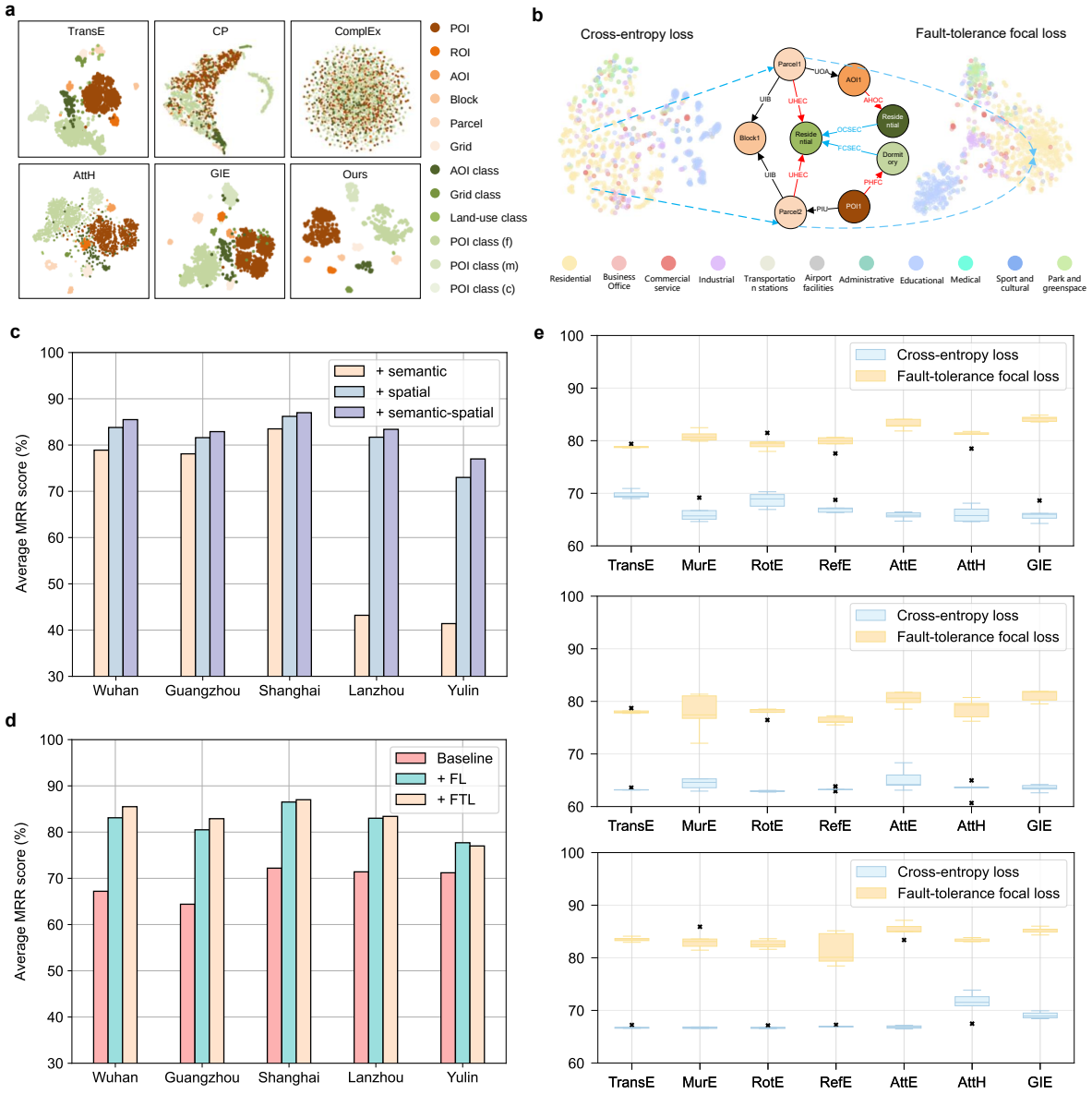

Figure 7: **The visualization and ablation studies of the mKGR model. a**, The t-SNE of the spatial entity embedding (orange), and semantic entity embedding (green) on knowledge graph. Our method has a better distinction between these two types of entities. **b**, The t-SNE of the output embeddings from the mKGR model demonstrates a comparison before and after the application of FL and FTL. The positions of these triples within the knowledge graph are also depicted. **c**, The ablation of the three different types of embeddings (semantic, spatial, and semantic-spatial). **d**, The ablation of FL and FTL. **e**, A comparison of the accuracy after applying fault-tolerance focal loss to other knowledge graph embedding methods.

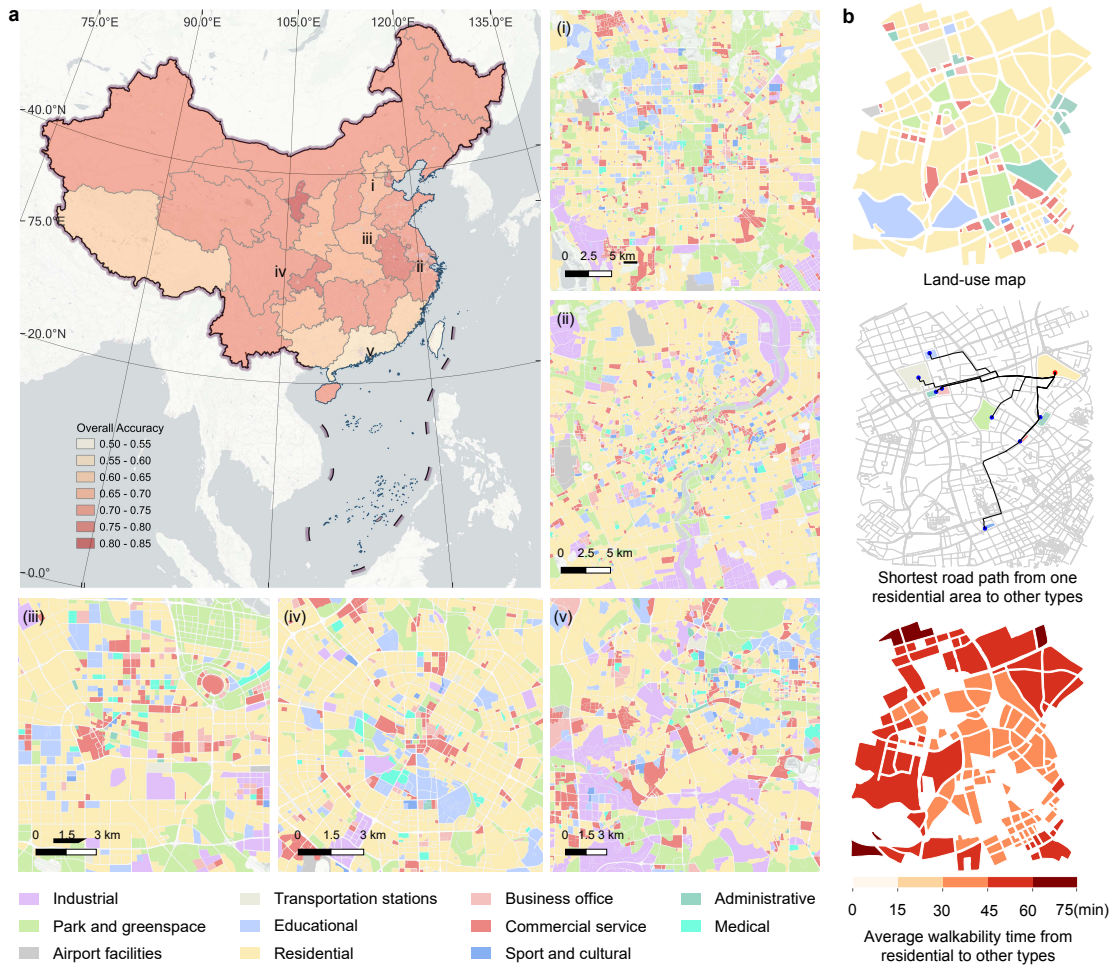

Figure 8: **The result of the land-use map in China and its application case. a**, The samples are distributed across various regions of China. From i to v are Beijing, Suzhou, Zhengzhou, Chengdu, and Shenzhen respectively. At the same time, it shows the different OA in each province of China. **b**, An application case for inclusive 15-minute walkability assessment. The shortest road path from each residential parcel to other land-use types are calculated and the average walkability time is visualized.

As shown in Table 6, our embedding method outperforms all other KGE methods, achieving the highest accuracy across the five city datasets. SSE is uniquely tailored to the MKG structure by integrating three distinct embeddings for spatial, semantic, and spatial-semantic relations, enabling a more nuanced representation of the latent relations within the knowledge graph. Importantly, SSE models spatial and semantic entities independently, ensuring a distinct separation between these two categories of entities. This differentiation is visually represented in the t-SNE dimensionality reduction of the embeddings, as shown in Figure 7a. In addition to the comparison of method effectiveness, we have also included a comparison of resource consumption for mKGR and baseline models in Table B.8.

4.4. Explainability analysis and ablation study of mKGR

The inference process in the knowledge graph is accomplished by evaluating the similarity among the head entity, relation, and tail entity. Here we demonstrate the effectiveness of the mKGR method through the visualization of entity embeddings, model output embeddings and present the ablation studies of the embedding model and loss function (Figure 7). The embeddings are visualized in Figure 7b offering a comparative analysis of mKGR embeddings before and after the integration of fault-tolerance focal loss. Focal loss (FL) directs the model to concentrate more on predicting land-use types of spatial entities, while fault-tolerance loss (FTL) equips the model to discern noise information, leading to a marked improvement in the separability of model output embeddings upon the incorporation of both. Figure 7b also highlights two triplets within the knowledge graph that are close in the embedding space, showing similar graph structures. The constraints imposed by the overall graph ensure that spatial entities with similar structures are predicted to belong to the same land-use category.

To delve deeper into the components of the embedding model and assess the impact of the loss function, we perform ablation studies on five datasets, as depicted in Figure 7c, d, and e. Firstly, we observe a consistent increase in the model's accuracy across all five cities with the progressive integration of semantic, spatial, and spatial-semantic embeddings. Subsequently, we conduct ablation studies on the proposed loss functions FL and FTL. The results demonstrate a substantial enhancement in model accuracy following the application of FL, which refines the knowledge graph embeddings, thereby improving the classification of land-use categories for spatial entities. The introduction of FTL further enhances accuracy in larger city graphs, yet its impact is muted in smaller graphs, such as Yulin. This is attributed to the fact that effective fault-tolerance learning necessitates robust graph constraints, whereas smaller graphs may lack the relational constraints needed, potentially resulting in adverse effects during training. Thus, it is inferred that FTL may not be as effective for land-use mapping tasks in smaller cities. Finally, given that enhanced loss functions can be generalized across various KGE methods, we extend our ablation studies to additional comparative embedding techniques. As illustrated in Figure 7e, all methods demonstrate notable accuracy gains under the influence of the enhanced loss functions.

5. Discussion

5.1. Accuracy analysis of land-use mapping for China

In this study, we introduce the first-ever concept of MKG customized for urban landscape and propose an unsupervised reasoning method (i.e., mKGR) over this knowledge graph to accomplish zero-shot land-use mapping. Through automatic aggregation of geospatial data in China, we subsequently construct a large-scale multi-granularity knowledge graph, which contains 32 million triples. Through automatically reasoning over this knowledge graph with a novel fault-tolerant loss function, a nationwide urban land-use

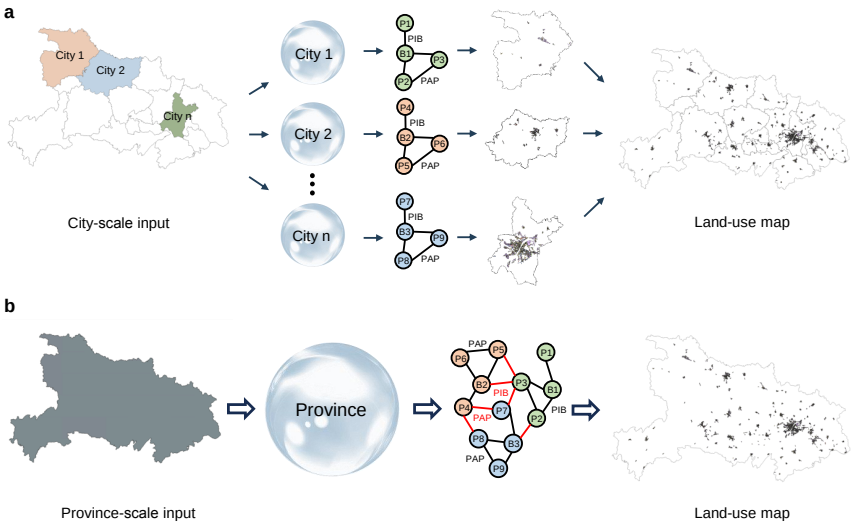

Figure 9: **The processing of different scale inputs in mKGR. a**, At the city scale, a province is first divided into multiple inputs according to cities, and knowledge graph are constructed for each city before mapping. **b**, At the province scale, the knowledge graph is constructed directly at the province scale, followed by mapping.

map with OA reaching 71.1%, covering 143k square kilometers, can be generated without any task-oriented annotation. As shown in Figure 8a, the generated land-use map covers whole urban regions of China. To further show the land-use mapping product, we present typical land-use mapping results for Yangtze River Delta region, as shown in Figure B.12.

Objectively, the accuracy of the generated land-use map via mKGR presents significant variation across different regions (as shown in Figure 8a), which is primarily attributed to the heterogeneous quality and coverage of crowdsourced geospatial data (e.g., POI and AOI) across different regions of China. To facilitate researchers and practitioners in reliably utilizing the generated land-use map, we conduct a comprehensive accuracy assessment over the land-use map and illustrate the land-use mapping accuracy for each province across China. It is noted that, researchers who can access higher-quality spatial data or more comprehensive geospatial data, can leverage our publicly available mKGR to generate the land-use map with a substantially enhanced accuracy.

5.2. Superiority Analysis of Large-Scale Holistic Knowledge Graph Reasoning

Urban land-use mapping necessitates large-scale holistic reasoning due to the inherent requirement of capturing broader urban contexts that encompass complex spatial interactions. However, existing deep learning approaches are often constrained by their focus on single-sample inference, limiting their ability to model such large-scale contextual dependencies. In contrast, our proposed mKGR framework demonstrates a strong capability for large-scale holistic reasoning. To evaluate its effectiveness, we conducted experiments at both provincial and city scales across China (i.e., constructing MKGs and performing land-use mapping

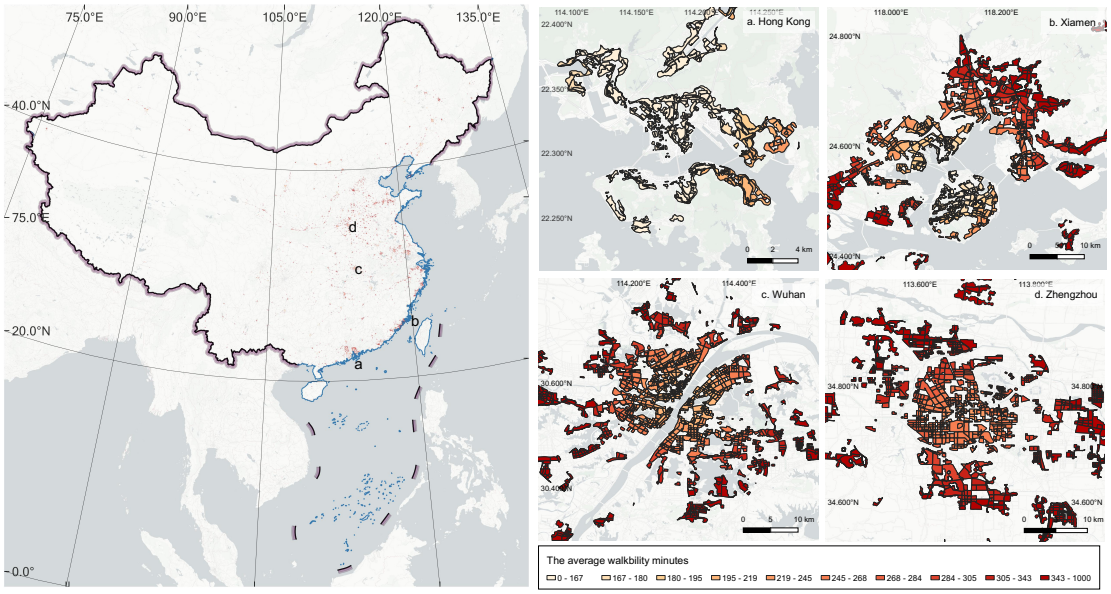

Figure 10: Inclusive 15-minute walkability assessment for selected citied in China.

at different administrative levels, as shown in Figure 9). The results indicate that mKGR exhibits superior performance at the broader provincial scale, with PA, UA, and OA reaching 35.6%, 37.9%, and 71.1%, respectively, compared to city-scale figures of 35.3%, 32.0%, and 69.3%. The comparison of accuracy between these two scales for each province is detailed in Table 7, consistently demonstrating that provincial-scale metrics outperform city-scale metrics. This superior performance can be attributed to the model's ability to capture broader spatial patterns and relationships between cities within the same province, providing richer contextual information for land-use classification. The larger knowledge graph the provincial level enable more comprehensive integration of diverse geographic, socioeconomic, and infrastructural information, while also allowing the model to learn from similar land-use patterns across multiple cities. These findings underscore that mKGR's performance improves with larger input knowledge graphs, highlighting its capability for large-scale holistic reasoning.

5.3. Inclusive 15-minute walkability assessment for China

To show the practical application value of the generated nationwide land-use map of China, we conduct an inclusive 15-minute walkability assessment for every city in China based on the genrated map, as shown in Figure 8b. Specifically, the road network is leveraged to determine the shortest road path from each residential parcel in the land-use map to other categories. Subsequently, we calculate the walking time along the shortest path and take the average time across all categories to gauge a city's 15-minute walkability. Figure 10 presents the 15-minute walkability for various cities. In contrast to the previous case studies about the city's 15-minute walkability (Bruno et al., 2024; Abbiasov et al., 2024; Ma et al., 2023), the generated

Table 7: mKGR Performance: City vs. Province Scale

	PA		UA		OA	
Prov.	c	p	c	p	c	p
Hubei	31.7	**32.0**	32.0	**35.6**	67.9	**68.8**
Hunan	31.7	**32.9**	35.2	**43.3**	66.2	**70.0**
Hebei	34.2	**34.3**	25.1	**39.8**	64.0	**67.5**
Guangdong	**34.6**	33.1	31.1	**31.2**	**57.9**	55.6
Shandong	32.3	**32.3**	34.8	**44.0**	70.3	**73.1**
Jiangsu	**31.5**	30.0	27.8	**35.9**	71.9	**73.6**
Henan	**32.4**	32.3	35.5	**44.7**	66.3	**69.8**
Sichuan	**46.2**	45.8	42.5	**43.7**	**75.0**	74.8
Zhejiang	**32.0**	29.9	29.9	**44.3**	69.4	**71.6**
Anhui	35.5	**35.8**	37.4	**42.0**	71.7	**76.5**
Fujian	**34.9**	34.0	31.3	**33.6**	62.0	**64.5**
Jiangxi	32.9	**33.4**	31.1	**37.2**	67.1	**70.7**
Shanxi	**32.9**	32.8	35.1	**40.8**	70.5	**72.4**
Liaoning	30.6	**30.9**	31.3	**33.0**	71.9	**73.8**
Jilin	**24.8**	23.9	**35.1**	32.8	**73.5**	72.2
Heilongjiang	25.7	**25.9**	29.5	**41.1**	69.7	**74.3**
Guangxi	18.7	**19.3**	**27.7**	27.0	64.6	**65.0**
Yunnan	**28.6**	27.5	25.8	**26.9**	68.8	**70.3**
Guizhou	37.9	**38.4**	**42.2**	41.5	65.3	**67.2**
Shaanxi	25.0	**25.7**	28.7	**33.2**	61.0	**68.6**
Gansu	**23.3**	22.8	30.6	**32.9**	67.4	**71.4**
Qinghai	**28.5**	28.4	35.3	**39.8**	**71.5**	70.9
Hainan	44.4	**44.9**	32.2	**36.2**	70.5	**72.5**
Ningxia	33.0	**33.4**	**42.9**	41.5	78.5	**80.4**
Xinjiang	**17.3**	16.1	19.8	**30.8**	68.9	**71.9**
Inner Mongolia	**25.7**	25.6	38.4	38.4	**75.9**	75.0
Tibet	21.5	**24.8**	18.7	**21.1**	56.6	**62.4**
All	35.3	**35.6**	32.0	**37.9**	69.3	**71.1**

Note: c: city scale; p: province scale.

land-use map enables the implementation of city accessibility analysis at the national scale. The analysis and visualization demonstrate how our land-use map can directly inform urban planning at both national and local scales by identifying areas requiring improved accessibility and supporting the development of more equitable and sustainable cities throughout China.

5.4. Advantages and limitations

The mKGR method, similar to many existing large-scale land-use mapping approaches, classifies individual land units or parcels delineated by road networks. This delineation strategy, utilizing roads as strong constraints, offers a significant advantage: it produces land-use maps that are highly suitable for unit-oriented urban analyses. For instance, concepts such as the '15-minute city' rely heavily on the assessment of services and functions within well-defined spatial units, and the parcel-based output of mKGR directly supports such applications by providing clear, analyzable units. During the rule-based inference process, mKGR assigns the land-use type that occupies the largest area within a given parcel as its primary

classification, further streamlining its utility for these parcel-centric studies.

However, this parcel-based approach also presents inherent limitations, primarily concerning the characterization of fine-grained mixed land-use scenarios occurring *within* a single delineated parcel. By assigning a dominant land-use type, the method may not fully capture the heterogeneity present in complex urban parcels where multiple distinct land uses coexist in smaller proportions. While mKGR can provide an estimation of the presence or proportional area of different land-use types within a parcel through their respective classification confidence scores, an explicit constraint to model such internal mixed-use compositions was not incorporated during the training phase of the current model.

While the current framework is advantageous for analyses focusing on parcel-level characteristics, the explicit modeling and differentiation of sub-parcel mixed land uses represent a valuable direction for future research. Future work could explore the integration of specific constraints or mechanisms during the training process to better identify and quantify these finer-grained mixed land uses, thereby enhancing the comprehensive land-use characterization capabilities of the mKGR framework.

6. Conclusion

This paper introduces the multi-granularity knowledge graph reasoning (mKGR) framework, a novel approach for zero-shot urban land-use mapping. mKGR leverages a multi-granularity knowledge graph (MKG) that integrates multimodal geospatial data, encompassing various spatial granularities and rich spatial-semantic interactions. By reframing land-use mapping as a knowledge graph completion problem, mKGR utilizes a fault-tolerant knowledge graph embedding method to infer land-use categories without requiring direct supervision. Extensive experiments demonstrate that mKGR outperforms existing zero-shot and fully-supervised methods, achieving improvements from 0.08 to 0.20 across several performance metrics. Utilizing mKGR's zero-shot classification and holistic reasoning abilities, a comprehensive urban land-use map of China is generated with low cost. A subsequent nationwide assessment of 15-minute city walkability provided insights into urban planning and sustainable development.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grants (42371321; 42030102) and the Ant Group.

Appendix A. Hyperbolic Space and Its Operations

Hyperbolic space is a non-Euclidean space characterized by constant negative curvature. It can be effectively modeled using the Poincaré ball model, which represents hyperbolic space as an open ball:

$$\mathbb{B}_c^d = \left\{ \boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\| < \frac{1}{\sqrt{c}} \right\} \tag{A.1}$$

where $c > 0$ denotes the curvature, and $\|\boldsymbol{x}\|$ is the Euclidean norm of $\boldsymbol{x}$.

Exponential Map: The exponential map $\exp_{\mathbf{0}}^c : T_{\mathbf{0}}\mathbb{B}_c^d \to \mathbb{B}_c^d$ maps a vector from the tangent space at the origin to hyperbolic space:

$$\exp_{\mathbf{0}}^c(\boldsymbol{v}) = \tanh\left(\sqrt{c}\|\boldsymbol{v}\|\right) \frac{\boldsymbol{v}}{\sqrt{c}\|\boldsymbol{v}\|}, \quad \boldsymbol{v} \in T_{\mathbf{0}}\mathbb{B}_c^d \tag{A.2}$$

This function maps a Euclidean vector $\boldsymbol{v}$ to a point in hyperbolic space.

Möbius Addition: Möbius addition $\oplus_c$ generalizes vector addition to hyperbolic space. For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{B}_c^d$, it is defined as:

$$\boldsymbol{x} \oplus_c \boldsymbol{y} = \frac{(1 + 2c\langle \boldsymbol{x}, \boldsymbol{y} \rangle + c\|\boldsymbol{y}\|^2)\boldsymbol{x} + (1 - c\|\boldsymbol{x}\|^2)\boldsymbol{y}}{1 + 2c\langle \boldsymbol{x}, \boldsymbol{y} \rangle + c^2\|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2} \tag{A.3}$$

where $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ is the Euclidean inner product.

Hyperbolic Distance: The distance between two points $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{B}_c^d$ in hyperbolic space is:

$$d_c(\boldsymbol{x}, \boldsymbol{y}) = \frac{2}{\sqrt{c}} \tanh^{-1}\left(\sqrt{c}\,\|-\boldsymbol{x} \oplus_c \boldsymbol{y}\|\right) \tag{A.4}$$

This distance function measures the geodesic distance between two points in hyperbolic space.

Appendix B. Additional Figures and Tables

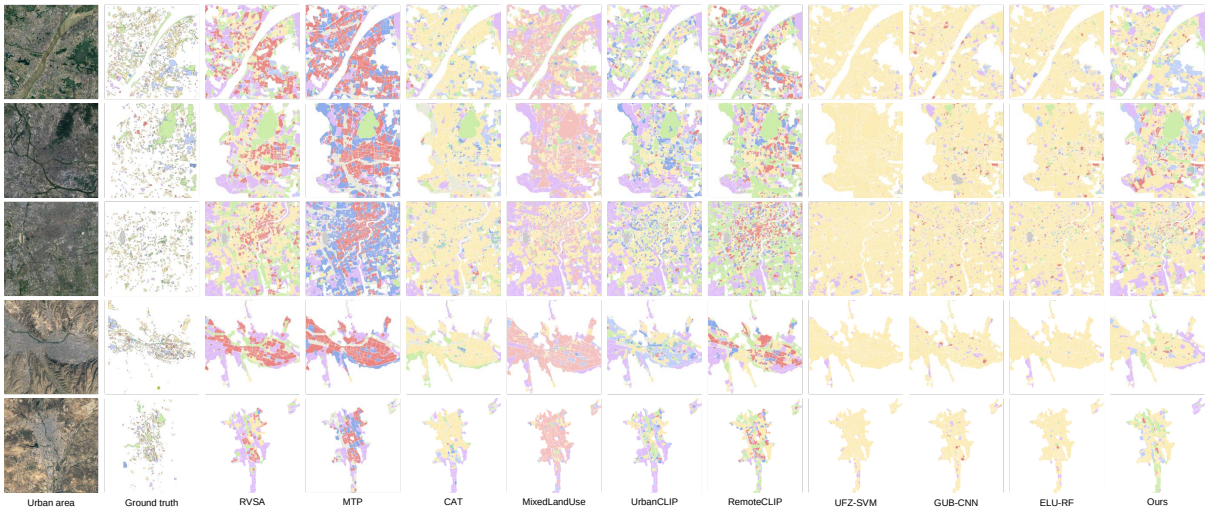

Figure B.11: Visualization results of all land-use methods for five cities.

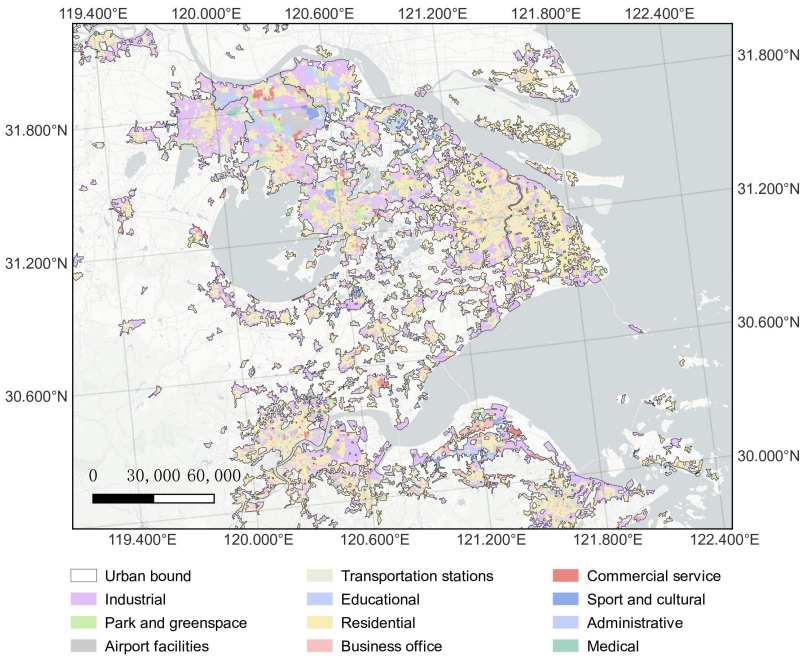

Figure B.12: The land-use maps of Yangtze River Delta region.

Table B.8: Comparison of computational efficiency for two groups of methods.

Land-use mapping Baselines					Knowledge Graph Completion Baselines				
Methods	Memory (G)	Param. (M)	Train Time (h)	Pred. Time (s)	Methods	Memory (G)	Param. (M)	Train Time (h)	Pred. Time (s)
RVSA	1.99	89.35	2.78	2994.86	TransE	2.65	4.07	0.09	0.14
MTP	1.81	317.67	6.38	3907.32	MurE	2.65	4.07	0.09	0.14
CAT	2.95	233.15	16.25	3486.96	CP	1.73	4.07	0.1	0.13
MixedLandUse	1.51	427.94	-	53.27	RotE	2.65	4.07	0.08	0.2
UrbanCLIP	1.51	427.94	-	56.87	RefH	5.39	4.07	0.43	0.17
RemoteCLIP	2.31	427.62	-	52.44	AttH	5.39	4.07	0.51	0.17
GUB-CNN	0.87	24.04	5.71	936.32	ComplEx	2.2	7.91	0.21	0.2
UFZ-SVM	2.86	-	0.24	211.32	RotatE	2.2	7.91	0.23	0.14
ELU-RF	2.6	-	0.02	19.01	GIE	5.39	4.07	1.12	0.21
mKGR	5.39	4.07	0.83	0.18	mKGR	5.39	4.07	0.83	0.18

References

Abbiasov, T., Heine, C., Sabouri, S., Salazar-Miranda, A., Santi, P., Glaeser, E., Ratti, C., 2024. The 15-minute city quantified using human mobility data. Nature Human Behaviour 8, 445–455.

Alidadi, M., Sharifi, A., 2022. Effects of the built environment and human factors on the spread of covid-19: A systematic literature review. Science of The Total Environment 850, 158056.

Arino, O., Gross, D., Ranera, F., Leroy, M., Bicheron, P., Brockman, C., Defourny, P., Vancutsem, C., Achard, F., Durieux, L., Bourg, L., Latham, J., Di Gregorio, A., Witt, R., Herold, M., Sambale, J., Plummer, S., Weber, J.L., 2007. Globcover: Esa service for global land cover from meris, in: 2007 IEEE International Geoscience and Remote Sensing Symposium, pp. 2412–2415.

Balazevic, I., Allen, C., Hospedales, T., 2019. Multi-relational poincaré graph embeddings. Advances in Neural Information Processing Systems 32.

Balchin, W., 1985. The development of land use maps. Land Use Policy 2, 3–15.

Bartholomé, E., Belward, A.S., 2005. Glc2000: a new approach to global land cover mapping from earth observation data. International Journal of Remote Sensing 26, 1959–1977.

Barwise, J., 1977. An introduction to first-order logic, in: Barwise, J. (Ed.), HANDBOOK OF MATHEMATICAL LOGIC. volume 90 of *Studies in Logic and the Foundations of Mathematics*, pp. 5–46.

Bian, X., Chen, C., Tian, L., Du, Q., 2017. Fusing local and global features for high-resolution scene classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10, 2889–2901.

Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., Yakhnenko, O., 2013. Translating embeddings for modeling multi-relational data. Advances in neural information processing systems 26.

Bruno, M., Monteiro Melo, H.P., Campanelli, B., Loreto, V., 2024. A universal framework for inclusive 15-minute cities. Nature Cities 1, 633–641.

Büttner, G., 2014. CORINE Land Cover and Land Cover Change Products. Springer Netherlands, Dordrecht. pp. 55–74.

Cao, Z., Xu, Q., Yang, Z., Cao, X., Huang, Q., 2022. Geometry interaction knowledge graph embeddings. Proceedings of the AAAI Conference on Artificial Intelligence 36, 5521–5529.

Chami, I., Wolf, A., Juan, D.C., Sala, F., Ravi, S., Ré, C., 2020. Low-dimensional hyperbolic knowledge graph embeddings, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 6901–6914.

Chen, C., Zhang, B., Su, H., Li, W., Wang, L., 2016. Land-use scene classification using multi-scale completed local binary patterns. Signal, image and video processing 10, 745–752.

Chen, G., Li, X., Liu, X., Chen, Y., Liang, X., Leng, J., Xu, X., Liao, W., Qiu, Y., Wu, Q., et al., 2020. Global projections of future urban land expansion under shared socioeconomic pathways. Nature Communications 11, 537.

Chen, J., Cao, X., Peng, S., Ren, H., 2017. Analysis and applications of globeland30: A review. ISPRS International Journal of Geo-Information 6.

Chen, Y., Dang, X., Zhu, D., Huang, Y., Qin, K., 2024. Urban functional zone mapping by coupling domain knowledge graphs and high-resolution satellite images. Transactions in GIS 28, 1510–1535.

Cheng, G., Han, J., Lu, X., 2017. Remote sensing image scene classification: Benchmark and state of the art. Proceedings of the IEEE 105, 1865–1883.

Deng, O., Wang, S., Ran, J., Huang, S., Zhang, X., Duan, J., Zhang, L., Xia, Y., Reis, S., Xu, J., et al., 2024. Managing urban development could halve nitrogen pollution in china. Nature Communications 15, 401.

Du, S., Du, S., Liu, B., Zhang, X., Zheng, Z., 2020. Large-scale urban functional zone mapping by integrating remote sensing images and open social data. GIScience and Remote Sensing 57, 411–430.

Duan, Z., Cheng, H., Xu, D., Wu, X., Zhang, X., Ye, X., Xie, Z., 2024. Cityllava: Efficient fine-tuning for vlms in city scenario, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7180–7189.

Gong, P., Chen, B., Li, X., Liu, H., Wang, J., Bai, Y., Chen, J., Chen, X., Fang, L., Feng, S., et al., 2020. Mapping essential urban land use categories in china (euluc-china): Preliminary results for 2018. Science Bulletin 65, 182–187.

Guo, X., Lao, J., Dang, B., Zhang, Y., Yu, L., Ru, L., Zhong, L., Huang, Z., Wu, K., Hu, D., et al., 2024. Skysense: A multimodal remote sensing foundation model towards universal interpretation for earth observation imagery, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 27672–27683.

Kuemmerle, T., Erb, K., Meyfroidt, P., Müller, D., Verburg, P.H., Estel, S., Haberl, H., Hostert, P., Jepsen, M.R., Kastner, T., Levers, C., Lindner, M., Plutzar, C., Verkerk, P.J., van der Zanden, E.H., Reenberg, A., 2013. Challenges and opportunities in mapping land use intensity globally. Current Opinion in Environmental Sustainability 5, 484–493.

Lacroix, T., Usunier, N., Obozinski, G., 2018. Canonical tensor decomposition for knowledge base completion, in: Dy, J., Krause, A. (Eds.), Proceedings of the 35th International Conference on Machine Learning, pp. 2863–2872.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.

Li, L., Zhan, W., Hu, L., Chakraborty, T., Wang, Z., Fu, P., Wang, D., Liao, W., Huang, F., Fu, H., Li, J., Liu, Z., Du, H., Wang, S., 2023. Divergent urbanization-induced impacts on global surface urban heat island trends since 1980s. Remote Sensing of Environment 295, 113650.

Li, Y., Wu, Y., Cheng, G., 2025. Meet: A million-scale dataset for fine-grained geospatial scene classification with zoom-free remote sensing imagery. IEEE/CAA Journal of Automatica Sinica .

Li, Z., Chen, B., Wu, S., Su, M., Chen, J.M., Xu, B., 2024. Deep learning for urban land use category classification: A review and experimental assessment. Remote Sensing of Environment 311, 114290.

Liu, F., Chen, D., Guan, Z., Zhou, X., Zhu, J., Ye, Q., Fu, L., Zhou, J., 2024. Remoteclip: A vision language foundation model for remote sensing. IEEE Transactions on Geoscience and Remote Sensing 62, 1–16.

Liu, Y., Ding, J., Fu, Y., Li, Y., 2023. Urbankg: An urban knowledge graph system. ACM Trans. Intell. Syst. Technol. 14.

Lu, W., Tao, C., Li, H., Qi, J., Li, Y., 2022. A unified deep learning framework for urban functional zone extraction based on multi-source heterogeneous data. Remote Sensing of Environment 270, 112830.

Ma, W., Wang, N., Li, Y., Sun, D.J., 2023. 15-min pedestrian distance life circle and sustainable community governance in chinese metropolitan cities: A diagnosis. Humanities and Social Sciences Communications 10, 1–14.

Montràs-Janer, T., Suggitt, A.J., Fox, R., Jönsson, M., Martay, B., Roy, D.B., Walker, K.J., Auffret, A.G., 2024. Anthropogenic climate and land-use change drive short-and long-term biodiversity shifts across taxa. Nature Ecology & Evolution 8, 739–751.

Ning, Y., Liu, H., Wang, H., Zeng, Z., Xiong, H., 2024. Uukg: unified urban knowledge graph dataset for urban spatiotemporal prediction. Advances in Neural Information Processing Systems 36.

OpenStreetMap contributors, 2017. Planet dump retrieve from https://planet.osm.org. https://www.openstreetmap.org.

Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al., 2021. Learning transferable visual models from natural language supervision, in: International conference on machine learning, pp. 8748–8763.

Rossi, A., Barbosa, D., Firmani, D., Matinata, A., Merialdo, P., 2021. Knowledge graph embedding for link prediction: A comparative analysis. ACM Transactions on Knowledge Discovery from Data (TKDD) 15, 1–49.

Shao, J.J., Shi, J.X., Yang, X.W., Guo, L.Z., Li, Y.F., 2023. Investigating the limitation of clip models: The worst-performing categories. Preprint submitted to arXiv.

Shi, C., Zhang, Y., Wang, J., Guo, X., Zhu, Q., 2024. Multimodal urban areas of interest generation via remote sensing imagery and geographical prior. Preprint submitted to arXiv.

Sun, Z., Deng, Z.H., Nie, J.Y., Tang, J., 2019. Rotate: Knowledge graph embedding by relational rotation in complex space, in: International Conference on Learning Representations.

Tian, L., Zhou, X., Wu, Y.P., Zhou, W.T., Zhang, J.H., Zhang, T.S., 2022. Knowledge graph and knowledge reasoning: A

systematic review. Journal of Electronic Science and Technology 20, 100159.

Tifrea, A., Becigneul, G., Ganea, O.E., 2019. Poincare glove: Hyperbolic word embeddings, in: International Conference on Learning Representations.

Trouillon, T., Welbl, J., Riedel, S., Gaussier, E., Bouchard, G., 2016. Complex embeddings for simple link prediction, in: Balcan, M.F., Weinberger, K.Q. (Eds.), Proceedings of The 33rd International Conference on Machine Learning, pp. 2071–2080.

Wang, D., Zhang, J., Xu, M., Liu, L., Wang, D., Gao, E., Han, C., Guo, H., Du, B., Tao, D., Zhang, L., 2024. Mtp: Advancing remote sensing foundation model via multitask pretraining. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 17, 11632–11654.

Wang, D., Zhang, Q., Xu, Y., Zhang, J., Du, B., Tao, D., Zhang, L., 2023. Advancing plain vision transformer toward remote sensing foundation model. IEEE Transactions on Geoscience and Remote Sensing 61, 1–15.

Weiming Huang, J.W., Cong, G., 2024. Zero-shot urban function inference with street view images through prompting a pretrained vision-language model. International Journal of Geographical Information Science 38, 1414–1442.

Wickramasuriya, R., Chisholm, L.A., Puotinen, M., Gill, N., Klepeis, P., 2013. A method to dynamically subdivide parcels in land use change models. International Journal of Geographical Information Science 27, 1497–1513.

Wu, M., Huang, Q., Gao, S., Zhang, Z., 2023. Mixed land use measurement and mapping with street view images and spatial context-aware prompts via zero-shot multimodal learning. International Journal of Applied Earth Observation and Geoinformation 125, 103591.

Xia, G.S., Hu, J., Hu, F., Shi, B., Bai, X., Zhong, Y., Zhang, L., Lu, X., 2017. Aid: A benchmark data set for performance evaluation of aerial scene classification. IEEE Transactions on Geoscience and Remote Sensing 55, 3965–3981.

Xie, R., Liu, Z., Lin, F., Lin, L., 2018. Does william shakespeare really write hamlet? knowledge representation learning with confidence, in: Proceedings of the AAAI conference on artificial intelligence.

Xiong, S., Zhang, X., Lei, Y., Tan, G., Wang, H., Du, S., 2024. Time-series china urban land use mapping (2016–2022): An approach for achieving spatial-consistency and semantic-transition rationality in temporal domain. Remote Sensing of Environment 312, 114344.

Xu, R., Huang, W., Zhao, J., Chen, M., Nie, L., 2023. A spatial and adversarial representation learning approach for land use classification with pois. ACM Trans. Intell. Syst. Technol. 14.

Yang, Z., Ding, M., Zhou, C., Yang, H., Zhou, J., Tang, J., . Understanding negative sampling in graph representation learning, in: Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery and data mining, pp. 1666–1676.

Yao, Y., Yan, X., Luo, P., Liang, Y., Ren, S., Hu, Y., Han, J., Guan, Q., 2022. Classifying land-use patterns by integrating time-series electricity data and high-spatial resolution remote sensing imagery. International Journal of Applied Earth Observation and Geoinformation 106, 102664.

Yin, J., Dong, J., Hamm, N.A., Li, Z., Wang, J., Xing, H., Fu, P., 2021. Integrating remote sensing and geospatial big data for urban land use mapping: A review. International Journal of Applied Earth Observation and Geoinformation 103, 102514.

Yu, L., Wang, J., Clinton, N., Xin, Q., Zhong, L., Chen, Y., Gong, P., 2013. From-gc: 30 m global cropland extent derived through multisource data integration. International Journal of Digital Earth 6, 521–533.

Zhang, Y., Liu, P., Biljecki, F., 2023. Knowledge and topology: A two layer spatially dependent graph neural networks to identify urban functions with time-series street view image. ISPRS Journal of Photogrammetry and Remote Sensing 198, 153–168.

Zheng, Y., Lin, Y., Zhao, L., Wu, T., Jin, D., Li, Y., 2023. Spatial planning of urban communities via deep reinforcement learning. Nature Computational Science 3, 748–762.

Zhong, Y., Yan, B., Yi, J., Yang, R., Xu, M., Su, Y., Zheng, Z., Zhang, L., 2023. Global urban high-resolution land-use mapping: From benchmarks to multi-megacity applications. Remote Sensing of Environment 298, 113758.