GUI: A Comprehensive Dataset of Global Urban Infrastructure Based on Geospatial Visual Foundation Models

GUI: A Comprehensive Dataset of Global Urban Infrastructure Based on Geospatial Visual Foundation Models

Zhenyu Han*
Department of Electronic Engineering, BNRist, Tsinghua University
Beijing, China

Xin Zhang*
Shenzhen International Graduate School, Tsinghua University
Shenzhen, China

Yanxin Xi
Department of Computer Science, University of Helsinki
Helsinki, Finland

Yan Luo
Media Lab, Massachusetts Institute of Technology
Department of Computing, The Hong Kong Polytechnic University
Hong Kong, China

Tong Xia
Department of Computer Science and Technology, University of Cambridge
Cambridge, UK

Yong Li†
Department of Electronic Engineering, BNRist, Tsinghua University
Beijing, China

ABSTRACT

The substantial social and financial costs of infrastructure identification impede in-depth analyses of sustainable urban design, especially in developing countries. In this paper, we present a novel framework with interactive web visualization based on geospatial visual foundation models. Leveraging this framework, we examine the urban infrastructure information in 1,178 cities worldwide, covering 93,088 km^2 areas. Cross-validation reveals that the overall accuracy of identified infrastructure achieves 67.0%. It sheds light on the sustainable development of cities and exposes the stark inequity in urban infrastructure provision for vulnerable populations. The identified urban infrastructure dataset of this study are available at https://github.com/tsinghua-fib-lab/GUI, and the interactive web application is at https://tinyurl.com/yz7xbfy3.

CCS CONCEPTS

• **Information systems → Data mining**; *Web searching and information discovery*; • **Computing methodologies → Computer vision**.

KEYWORDS

Urban Infrastructure, Computer Vision, Artificial Intelligence, Foundation Model, Satellite Imagery

ACM Reference Format:

Zhenyu Han, Xin Zhang, Yanxin Xi, Yan Luo, Tong Xia, and Yong Li. 2024. GUI: A Comprehensive Dataset of Global Urban Infrastructure Based on Geospatial Visual Foundation Models. In *The 32nd ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL '24), October 29-November 1, 2024, Atlanta, GA, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3678717.3691242

1 INTRODUCTION

The high population density poses significant challenges to urban infrastructure, which serves as the fundamental backbone of urban functionalities. This challenge is even more severe in developing countries, as a substantial portion of the global population, totaling 828 million people, still resides in slums with inadequate infrastructure support. To achieve the United Nations' Sustainable Development Goals (SDGs) of creating "inclusive, safe, resilient, and sustainable" cities by 2030, a critical step is to identify and assess the inequities in urban infrastructure.

Traditional methods for gathering infrastructure information heavily rely on census data and personal surveys. These approaches are often economically unfeasible in many developing countries [4]. Satellite imagery provides a promising proxy to census data [4, 9, 12, 14, 21]. However, due to the high quality variance of publicly available satellite images and the feature mismatch between training areas and inference areas, accurately identifying urban infrastructure globally is still a challenging task. Besides, the currently available satellite dataset usually concentrates on a few cities [3, 11], which are insufficient in capturing the global diversity of urban infrastructure.

With the help of valuable online tools, such as Esri World Imagery [5], Google Earth Engine [7], millions of high-quality satellite images can be easily collected and processed. Besides, geospatial visual foundation models leverage massive amounts of unlabeled satellite images to pretrain the dense computer vision backbone, which can effectively capture the varying remote sensing features globally and improve the generalization performance in out-of-distribution scenarios. Therefore, investigating the utilization of advanced network and model technologies to formulate a methodology for monitoring the SDGs constitutes a highly valuable research question.

In this paper, we present an automated framework for monitoring sustainable development progress through global infrastructure

*Both authors contributed equally to this research.
†Corresponding author (liyong07@tsinghua.edu.cn).

SIGSPATIAL '24, October 29-November 1, 2024, Atlanta, GA, USA

ACM ISBN 979-8-4007-1107-7/24/10
https://doi.org/10.1145/3678717.3691242

identification. Leveraging this framework, we examine the urban infrastructure in 1178 cities worldwide, covering 93,088 km^2 areas. We cross-validate the identified infrastructure with POI records from SafeGraph, which demonstrates a high overall accuracy of 67.0%. Based on these informative data, we showcase extensive sustainable topics can be investigated. For SDG 9.1 on developing quality, reliable, sustainable, and resilient infrastructure, we find that cities in high-income countries, which only occupy 35.7% of all the studied cities, enjoy a disproportionately high percentage of identified infrastructure at 54.2%. For SDG 10.1 on reducing income inequities and SDG 11.6 on reducing the adverse per capita environmental impact of cities, we prove that the identified urban infrastructure is highly informative in scaling law. Moreover, leveraging the identified urban infrastructure, a deeper insight of general development law can be examined.

In summary, our contributions in this paper are outlined below:

- We are the first to introduce a sustainable development framework designed for identifying global urban infrastructure, enabling automated crawling and identification of infrastructure in any city.
- We conducted several analyses on identified urban infrastructure, showcasing that the proposed framework greatly promotes understanding of several SDG targets.
- We present a comprehensive dataset with an interactive web application of urban infrastructure in 1178 representative cities globally, where 64.3% of them belong to vulnerable populations in non-high-income countries according to the World Bank classification [18].

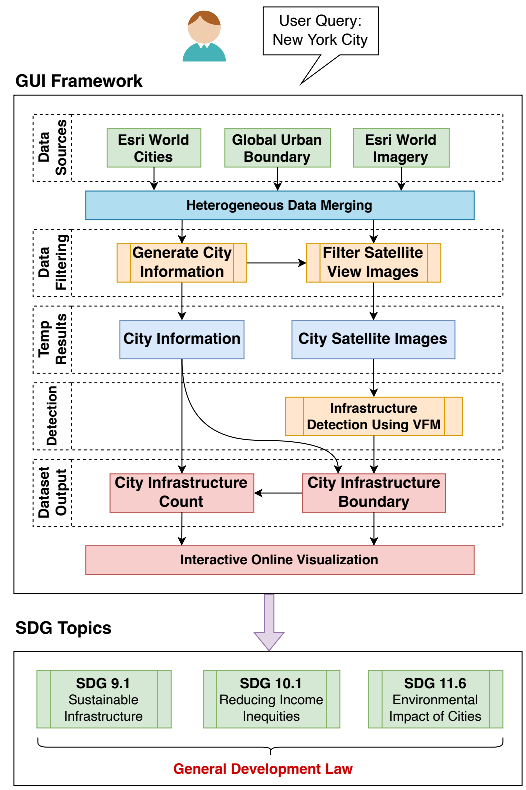

Figure 1: Framework overview.

2 METHODOLOGY

The schematic framework of this study is demonstrated in Figure 1. Given any user query, our framework collects the city statistics from Esri World Cities [6] and global urban boundary (GUB) [13]. Furthermore, we collect 2 million high-resolution images from Esri World Imagery covering the studied areas. We use the state-of-the-art geospatial VFM [17] finetuned on several object detection datasets to identify the urban infrastructure. Besides, we cross-validate the identified infrastructure and produce both the geometries and entity counts for various research communities. Finally, we build an interactive web application that enables instant data analyses, which supports downstream sustainable development applications.

2.1 Data Sources

In this study, we refer to the Esri World Cities dataset, which includes city center coordinates, the respective country name, and populations. It summarizes the basic information of global cities, acting as a basis for the following procedures.

We also refer to a remote sensing based global urban boundary dataset [13], which is derived from global artificial impervious area data. It leverages the fact that city functions typically build on concrete grounds to determine the urban areas. We match the Esri World Cities and the latest version (2018) of the global urban boundary dataset by calculating the inclusion relationship between the city center coordinates and the city boundaries.

As for the income level, we refer to the World Bank income level classification [18] as an agent for vulnerable populations. Specifically, the World Bank has categorized the countries into 4 groups: high-income country (HIC), upper-middle-income country (UMIC), lower-middle-income country (LMIC), and low-income country (LIC). We classify the cities according to the corresponding country income classification in our dataset. Furthermore, we also quantify such income level by linking the country with the GDP data in 2020 [19, 20] from the World Bank.

Finally, we collect the latest 1.2m resolution satellite tiles from Esri World Imagery that cover all the urban areas according to the code implementation in [8], which generates images in 256×256 resolution. Esri guarantees a typical 3-year update frequency of the provided satellite imagery.

2.2 Detection of Urban Infrastructure

Due to the limited training data of satellite imagery, accurately identifying urban infrastructure is a challenging task[11]. The varying infrastructure features may hinder traditional deep learning models when inferring unseen regions.

To tackle this challenge, we adopt a geospatial visual foundation model, ViTAE-RVSA [17], for the infrastructure detection problem. First, we apply the above two fine-tuned models to infer the collected satellite images of global cities simultaneously, which generates the bounding boxes of each entity of each category. Second, for the common categories that exist in both two models, we

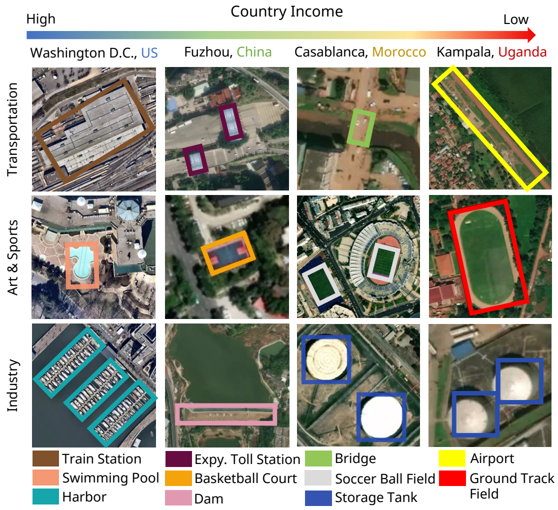

Figure 2: Illustration of identified infrastructure in countries of different income levels. Note that not all object categories are demonstrated in this figure due to the limited space.

propose a confidence-aware merging procedure to tackle the overlapping issue. Specifically, for bounding boxes without any overlap, we preserve both of them to ensure the coverage rate. For the overlapping boxes, we calculate the confidence levels and select results with higher values to prevent repeat detection.

In this study, we manually check the identified infrastructure to guarantee the accuracy of detection, where the results are demonstrated in Figure 2. Most of the infrastructure can be accurately detected, and the bounding box is able to rotate to the right angle that best fits the infrastructure.

3 RESULTS

3.1 Evaluation of Identified Urban Infrastructure

To properly evaluate the accuracy of the identified urban infrastructure, we refer to the SafeGraph Places Data Schema [16] as the ground truth of urban POIs. Compared with the OSM data that suffer from inequalities and biases [10], SafeGraph data has a much higher quality and has been adopted in high-quality research [9, 15].

For each of the identified urban infrastructure, we search for the corresponding SafeGraph POI around its central location with a 5 km buffer considering the relatively large sizes of the infrastructure. The accuracy of the identification is reported in Table 1, where the overall accuracy achieves 67.0%. The accuracy surpasses 70% in 10 categories, such as Bridge, Train Station, Dam and Stadium. Considering the potential ground truth missing in SafeGraph dataset, the real accuracy of the identified infrastructure should be even higher, especially for underappreciated POIs. Nonetheless, the high reliability of the identified urban infrastructure provides a new opportunity for downstream research, especially in low-income regions.

Table 1: Accuracy of identified urban infrastructure

Infrastructure	Accuracy
Bridge (BRG)	75.5%
Airport (APT)	77.1%
Expy. Service Area (ESA) / Expy. Toll Station (ETS)	17.0%
Overpass (OP)	68.9%
Train Station (TS)	78.0%
Roundabout (RA)	43.0%
Baseball Field (BF)	73.9%
Tennis Court (TC)	71.7%
Basketball Court (BC)	83.3%
Ground Track Field (GTF)	78.0%
Golf Field (GF)	70.9%
Stadium (STD)	84.2%
Soccer Ball Field (SBF)	71.7%
Swimming Pool (SP)	65.1%
Harbor (HBR)	77.8%
Storage Tank (ST)	32.7%
Chimney (CHM)	(No Data)
Dam (DM)	43.8%
Windmill (WM)	57.9%
All	67.0%

3.2 Distribution of Urban Infrastructure

To obtain a more in-depth understanding of infrastructure inequity around the world, we analyze the distribution of the identified infrastructure in Figure 3. We surprisingly find that 54.2% of the identified infrastructure is concentrated in HICs, which only occupy 35.7% of the studied cities. More importantly, this overwhelming advantage holds across every infrastructure category. 62.1% of *Transportation*, 53.2% of *Art & Sports*, and 45.0% of *Industry* are occupied by cities in HICs. The most severe inequity happens in *Tennis Court*, *Harbor*, and *Overpass*, where HICs occupy 81.9%, 70.6%, and 67.9% accordingly.

3.3 Scaling Law of Urban Infrastructure

In this section, we demonstrate how the proposed dataset can help to reveal inequitable development from an urban science perspective. Empirical studies have demonstrated the scaling-free property of urban properties [1], which can be well described by a power-law function $Y = Y_0 N^{\beta}$, where Y is our dependent variable of interest, N is the city population, and Y_0, β are constants in N. Leveraging this power-law function, we fit the city population with the number of identified urban infrastructure in our dataset in Figure 4. First, we observe the scaling component β in all income level groups is below 1, which forms a sublinear scaling law. This observation is consistent with the literature, which proves such a relationship from a statistical physics perspective [2]. Second, categorizing cities according to the corresponding country income levels, we find that cities in HICs have a much lower scaling component ($\beta_{HIC} = 0.591$) than the global average ($\beta = 0.801$), while UMICs demonstrate a higher $\beta_{UMIC} = 0.931$. Cities in HICs are associated with a nearly saturated infrastructure construction. For cities in UMICs, the high

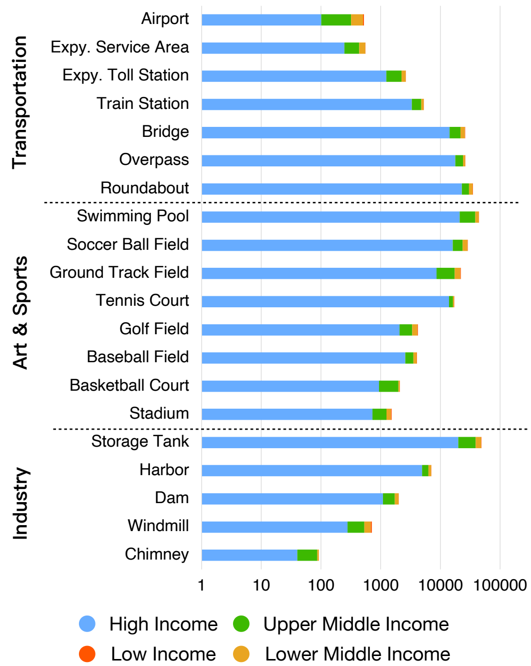

Figure 3: Number of identified infrastructure among countries by income level.

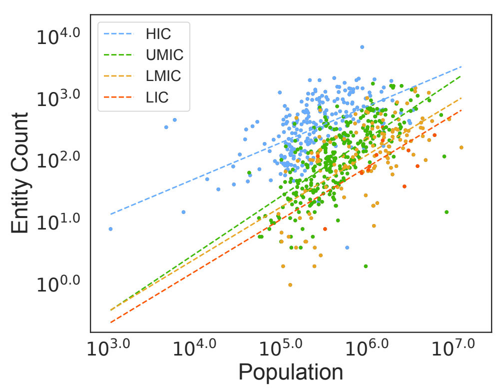

Figure 4: Scaling law of urban infrastructure with population. Each point represents a city.

scaling component reveals unbalanced development, where the largest cities enjoy a higher priority of infrastructure construction. Besides, the relatively small intercept reveals the insufficient development, where the overall infrastructure per capita still falls behind the HICs. Our observation is close to the theoretical model of $\bar{\beta} = 2/3$ [1] and other empirical observations of $\beta_{emp} = 0.8$ [2], validating the data quality of identified urban infrastructure in this study.

4 CONCLUSION

In this study, we present the first framework of global urban infrastructure identification leveraging the recent advances in geospatial visual foundation models. According to this framework, we successfully identify 20 object categories ranging from transportation, art & sports, and industry. The proposed framework supports extensive research topics on sustainable development, unveiling the profound urban infrastructure inequity. In future work, we aim to expand the published dataset to encompass more vulnerable cities and include additional infrastructure categories.

ACKNOWLEDGMENTS

This work was supported in part by National Natural Science Foundation of China under grant U23B2030, 62171260, and 62272260.

REFERENCES

[1] Luís MA Bettencourt. 2013. The origins of scaling in cities. *science* 340, 6139 (2013), 1438–1441.
[2] Luís MA Bettencourt, José Lobo, Dirk Helbing, Christian Kühnert, and Geoffrey B West. 2007. Growth, innovation, scaling, and the pace of life in cities. *PNAS* 104, 17 (2007), 7301–7306.
[3] Gong Cheng, Junwei Han, Peicheng Zhou, and Lei Guo. 2014. Multi-class geospatial object detection and geographic image classification based on collection of part detectors. *ISPRS* 98 (2014), 119–132.
[4] Guanghua Chi, Han Fang, Sourav Chatterjee, and Joshua E Blumenstock. 2022. Microestimates of wealth for all low-and middle-income countries. *PNAS* 119, 3 (2022), e2113658119.
[5] Esri. 2022. World Imagery. https://www.arcgis.com/home/item.html?id=10df2279f9684e4a9f6a7f08febac2a9%2F.
[6] Esri. 2023. World Cities. https://hub.arcgis.com/datasets/esri::world-cities/explore.
[7] Noel Gorelick, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *RSE* (2017). https://doi.org/10.1016/j.rse.2017.06.031
[8] Sungwon Han, Donghyun Ahn, Sungwon Park, Jeasurk Yang, Susang Lee, Jihee Kim, Hyunjoo Yang, Sangyoon Park, and Meeyoung Cha. 2020. Learning to score economic development from satellite imagery. In *SIGKDD*. 2970–2979.
[9] Zhenyu Han, Tong Xia, Yanxin Xi, and Yong Li. 2023. Healthy Cities, A comprehensive dataset for environmental determinants of health in England cities. *Scientific data* 10, 1 (2023), 165.
[10] Benjamin Herfort, Sven Lautenbach, João Porto de Albuquerque, Jennings Anderson, and Alexander Zipf. 2023. A spatio-temporal analysis investigating completeness and inequalities of global urban building data in OpenStreetMap. *Nature Communications* 14, 1 (2023), 3985.
[11] ISPRS. 2023. Potsdam. https://www.isprs.org/education/benchmarks/UrbanSemLab/2d-sem-label-potsdam.aspx.
[12] Tong Li, Shiduo Xin, Yanxin Xi, Sasu Tarkoma, Pan Hui, and Yong Li. 2022. Predicting Multi-level Socioeconomic Indicators from Structural Urban Imagery. In *CIKM*. 3282–3291.
[13] Xuecao Li, Peng Gong, Yuyu Zhou, Jie Wang, Yuqi Bai, Bin Chen, Tengyun Hu, Yixiong Xiao, Bing Xu, Jun Yang, et al. 2020. Mapping global urban boundaries from the global artificial impervious area (GAIA) data. *Environmental Research Letters* 15, 9 (2020), 094044.
[14] Yu Liu, Xin Zhang, Jingtao Ding, Yanxin Xi, and Yong Li. 2023. Knowledge-infused contrastive learning for urban imagery-based socioeconomic prediction. In *WWW 2023*. 4150–4160.
[15] Hamed Nilforoshan, Wenli Looi, Emma Pierson, Blanca Villanueva, Nic Fishman, Yiling Chen, John Sholar, Beth Redbird, David Grusky, and Jure Leskovec. 2023. Human mobility networks reveal increased segregation in large cities. *Nature* 624, 7992 (2023), 586–592.
[16] Safegraph. 2022. Safegraph Places Data Schema. https://docs.safegraph.com/docs/places.
[17] Di Wang, Qiming Zhang, Yufei Xu, Jing Zhang, Bo Du, Dacheng Tao, and Liangpei Zhang. 2022. Advancing plain vision transformer toward remote sensing foundation model. *TGRS* 61 (2022), 1–15.
[18] World Bank. 2022. The World by Income and Region. https://datatopics.worldbank.org/world-development-indicators/the-world-by-income-and-region.html.
[19] World Bank. 2023. GDP (Current US$). https://data.worldbank.org/indicator/NY.GDP.MKTP.CD.
[20] World Bank. 2023. GDP per capita (Current US$). https://data.worldbank.org/indicator/NY.GDP.PCAP.CD.
[21] Xin Zhang, Yu Liu, Yuming Lin, Qingmin Liao, and Yong Li. 2024. UV-SAM: Adapting Segment Anything Model for Urban Village Identification. In *AAAI*, Vol. 38. 22520–22528.