Computers, Environment and Urban Systems 108 (2024) 102071

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus

Inferring storefront vacancy using mobile sensing images and computer vision approaches

Yan Li [a], Ying Long [b,*]

[a] School of Architecture, Tsinghua University, Beijing, China
[b] School of Architecture and Hang Lung Center for Real Estate, Key Laboratory of Ecological Planning & Green Building, Ministry of Education, Tsinghua University, Beijing, China

ARTICLE INFO

Keywords:
Commercial vacancy
Mobile sensing images
Deep learning
Text recognition
Retail landscape

ABSTRACT

Storefront vacancy has been a widespread and worldwide phenomenon, raising concerns about the changing characteristic of the retail landscape, loss of community vitality, and hollowing out of cities. Although the causes leading to this phenomenon have been extensively debated, little granular data are available to evaluate the issue in a timely manner. Therefore, this study aims to develop a data-driven approach to capture the commercial structure of vacant storefronts on a store-by-store basis as well as to analyze their evolution patterns. First, street-level images were collected using mobile sensing in a low-cost, large-scale and efficient manner; then, a storefront vacancy estimation model was developed using computer vision techniques to infer the storefront location, operation status, business category, and vacancy rates. Three volunteers spent five days collecting street-level images from an urban area of 964 km^2 in the case city of Xining, China. As a result, 93,069 stores were identified in the city in March 2022, of which 25,488 were vacant. Moreover, the storefront vacancy rate increased significantly after the epidemic, from 21.8% in 2018 to 30.0% in 2022. Stores in shopping, catering, and life services had the maximum vacancies. The factors that had the greatest impact on storefront vacancy were, in order of importance, far from commercial zonings, low population density, and far from the urban center. However, these factors influenced the vacancy in diverse and complex ways, and in the future, urban planning strategies to address vacancy issues should be well considered and differentiated.

1. Introduction

Today, storefronts play a crucial role in shaping the commercial landscape of cities by providing economic vitality and creating essential shopping corridors for service-oriented users. Their highly visible presence adds to the vibrancy of urban areas (Cavan, 2016). They are vital in street design, providing not only access to goods and services but also civic and social spaces where people encounter each other, thus creating vibrant sidewalks. However, store closures have been a widespread worldwide phenomenon due to financial crisis (Chowdhury & Sarkar, 2017), e-commerce (Berman, 2019; Oskam, 2021; van Zweeden, 2009), imbalance between rising construction and shrinking consumption (Lee, Kim, & Kim, 2021), and global pandemic (Wang et al., 2020), leaving few corners of the world unscathed and inducing urban changes that government decision-makers and urban planners should be aware of (Yeates & Montgomery, 1999).

The causes of shop vacancies are complex and varied, ranging from changes in business structures and market competition to urban planning and management issues. The definition of vacancy as “the non-let situation of an available property” (Remøy & van der Voordt, 2007) is a valuable tool for comprehending the economic aspects of the subject matter under investigation in this study. Whereas a clear distinction among different forms of vacancy has been made based on their economic effects (Lee & Newman, 2021; van Zweeden, 2009). The category of low-impact or less detrimental vacancies can be broadly categorized into three groups: (1) initial vacancy, which arises when a building has just been completed; (2) mutation vacancy, which arises due to a change in property ownership or transfer; and (3) frictional vacancy, where the property has been vacant for less than one year and a new tenant has not yet been secured. In the harmful or high-impact class, long-term vacancy generally occurs for more than three consecutive years and, according to some studies, for 6 months or 1 year (Myers & Wyatt, 2004). The following three forms of vacancy have fewer clear distinctions: (1) structural vacancy, which results from the mismatch between housing

* Corresponding author at: School of Architecture, Tsinghua University, Haidian District, Beijing 100084, China.
E-mail address: ylong@tsinghua.edu.cn (Y. Long).

https://doi.org/10.1016/j.compenvurbsys.2023.102071
Received 30 October 2022; Received in revised form 13 December 2023; Accepted 26 December 2023

market supply and demand; (2) functional vacancy, resulting from incomplete internal functions or poor external environment and supporting facilities; and (3) technical vacancy, resulting from not meeting general market requirements (e.g., loan restrictions and purchase restrictions). The diverse nature and purpose of storefront vacancies create complexities in their definition, making it difficult to establish clear boundaries for determining when a store is considered vacant and for how long it remains so. As a result, the duration of storefront vacancies can appear arbitrary. The recognition of wasted capacity in the definition is crucial (Myers & Wyatt, 2004). Therefore, the study of shop vacancy is significant in promoting the development of commercial districts and the sustainability of cities.

The aim of the study is to understand the current situation, causes and solutions to the storefront vacancy issue, in order to inform the sustainable development of street businesses. Specific research questions include: (1) What is the definition and identification criteria of storefront vacancy? (2) How can a fine-grained database of storefront vacancies be obtained on a large scale and at low cost? (3) What are the characteristics of the spatial distribution and temporal changes of storefront vacancy? What are the influencing factors and solutions for storefront vacancies? How can targeted policies and measures be developed to address the issue? These questions are answered in detail in the following sections.

2. Literature review

There are differences in urban studies and development history of storefront vacancy between Western countries and China. Early research on vacant stores mainly focused on Europe and North America. In the early 1990s, Europe began to pay attention to the problem of vacant commercial properties, and the UK was one of the earliest countries to conduct research. By conducting surveys on different commercial districts in central London, it was found that the vacancy rate was on the rise, and reasons for London's vacant shops include high rents, the recession, e-commerce competition and urban planning (Home, 1983). Research in North America also began to increase gradually. A study conducted in 1997, which surveyed commercial districts in Manhattan, New York City, revealed a correlation between the decline of these districts and the presence of vacant commercial properties (Zukin, 1996). In 2000, a study was conducted on the phenomenon of vacant properties in global city centers, and it was believed that this phenomenon was related to globalization and changes in urban economic structure (Bromley, 2000). In recent years, research has gradually shifted its focus to Asia and Latin America. Asia, too, has its share of vacant commercial properties in commercial districts such as Hong Kong, Singapore, and Japan. A recent survey conducted in 2021 of commercial districts in Hong Kong, China revealed a high vacancy rate of up to 9.8% (Hui, Yiu, & Yau, 2007). A study of the commercial districts in the central business district of Singapore suggested that vacant commercial properties were caused by various factors such as economic downturn, high rents, Market disequilibrium resulting from a mismatch between supply and demand and other factors (Sim, Yu, & Malone-Lee, 2002). Commercial districts in Japan also face the problem of vacant properties. For example, a survey was conducted on commercial districts in Tokyo, Japan, and it was believed that vacant commercial properties were caused by factors such as population decline, urban planning, and market competition (Kanayama & Sadayuki, 2021). Chinese research on vacant commercial properties mainly focuses on large cities such as Beijing, Shanghai, and Guangzhou (Chin & Chow, 2012), and the research focus includes vacancy rates, reasons for vacancy, and countermeasures. Research results show that the vacancy rate of commercial properties in China has been increasing year by year, and the main reasons include real estate market regulation and commercial mode transformation. At the same time, countermeasures have also been proposed, such as optimizing commercial models, promoting urban renewal, and strengthening policy guidance. In addition, with the continuous development of China's economy and urbanization, the scope of research has gradually expanded, and the problem of vacant commercial properties in small and medium-sized cities as well as urban-rural areas has also begun to receive attention (Zhang, Zhu, & Ye, 2016).

The pros and cons of different research methods for studying storefront vacancies and their application scenarios also vary. Questionnaires and statistical analysis are suitable for gaining a comprehensive understanding of shop vacancy and the factors influencing it, but may suffer from sample selection bias and inaccurate information collection; Case studies can provide insight into the specific causes and solutions to shop vacancy, but may suffer from case specificity and non-replicability; GIS techniques and spatial econometric models can analyze the spatial distribution and influencing factors of shop vacancy, but require large amounts of data and computing power. The data used in existing studies mainly include store-by-store in-person or phone reviews (Talen & Park, 2022; Yeates & Montgomery, 1999); commercial real-estate databases, such as CoStar (Geurts & Black Jr, 2015) and Live XYZ, a New York based technology company that has mapped every ground floor use in the city; official records sourced from the city's finance department, which are provided by owners of commercial properties situated on the ground or second floor (Stringer, 2019); and community data obtained from non-profit organizations (e.g., MAPSCorps) and research institutes (e.g., Experian Goad surveys) (Tselios, Lambiri, & Dolega, 2018). However, interview and questionnaire data have a long update cycle as their collection requires high labor and time costs, while real-estate data are mostly only transaction data, which are collected by property managers and often fail to cover all vacant stores. Previous studies have faced challenges in obtaining comprehensive data, resulting in the analysis of commercial district and real estate market changes from a macroscopic perspective across an entire region. To better understand the trends and impacts of vacancies, a microscopic analysis is required.

3. Data and materials

3.1. Study area

The project partner city, Xining, the capital city of Qinghai Province, is our study area, and it is a highly representative economic center in Northwestern China (see Fig. 1), with the highest gross domestic product (GDP) growth rate of any provincial capital in Northwest China (over 150 billion yuan in 2021 according to the Qinghai Provincial People's Government). Xining is the sole central city on the Qinghai-Tibet Plateau with a population exceeding one million. In 2020, the city had a residential population of 2.46 million and an urbanization rate of 68.92%, according to the data of the Seventh National Population Census. It has the typical economic characteristics of a third-tier city in China, with a commercial area per capita (0.92 m^2/person) that already far exceeds the standard line (0.5 m^2/person) and even almost on a par with first-tier cities (1 m^2/person). And with many commercial complexes currently being planned and built, the overall commercial scale of Xining is continuing to rise, causing a tide of store closures under the commercial fever.

The central area of the city is composed of four districts (i.e., Chengdong, Chengxi, Chengbei, and Chengzhong) and two towns (i.e., Duoba and Lushaer), covering an area of 964 km^2 and 1446 km of urban roads. The locations of landmarks were also plot in Fig. 1, including Xining railway station and eight commercial zones (Dongguan Street in Chengdong district, Xiaoqiao and Minghui Cheng in Chengbei district, Dashizi and Shangri-La in Chengzhong district, and Limeng and Wanda Plaza in Chengxi district).

3.2. Method overview

The method framework mainly involves two steps, as shown in Fig. 2. The first step is to acquire mobile sensing data, which mainly

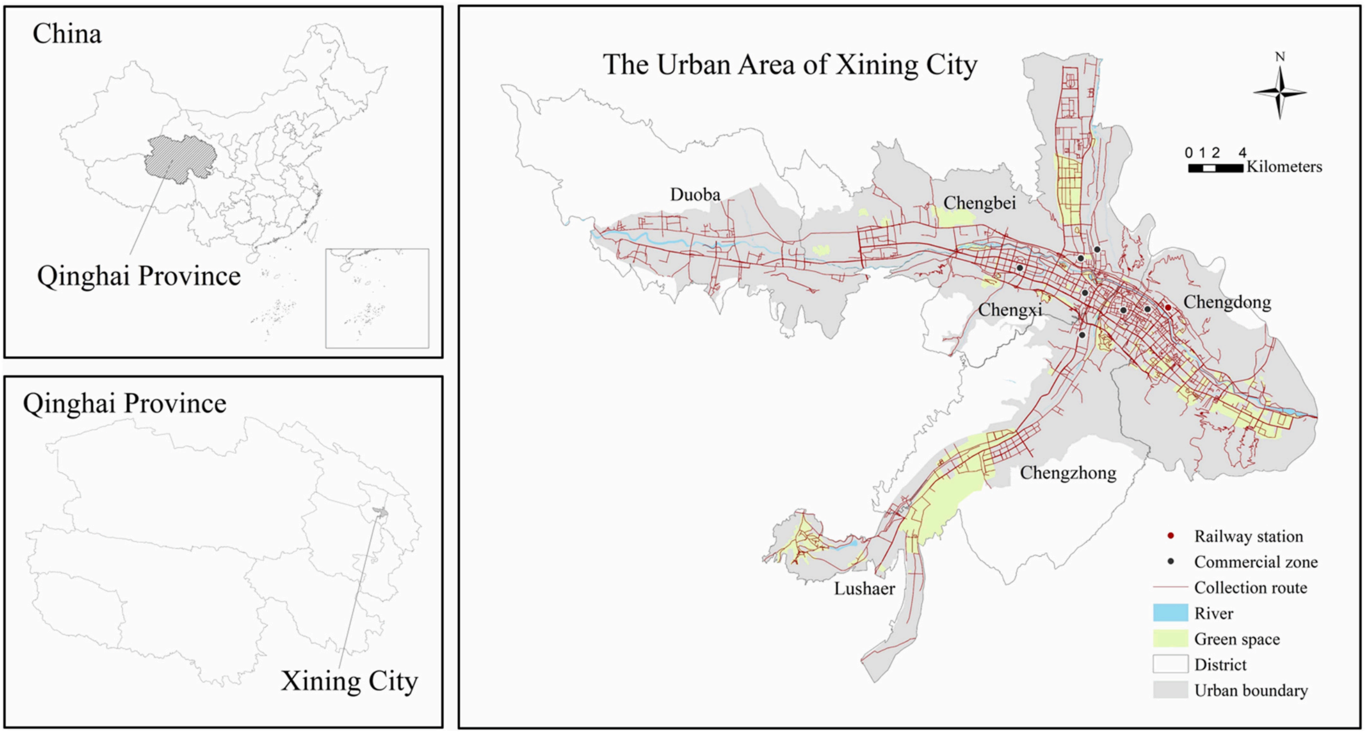

Fig. 1. Central area of Xining as the study area and storefronts along the urban roads as the study objects.

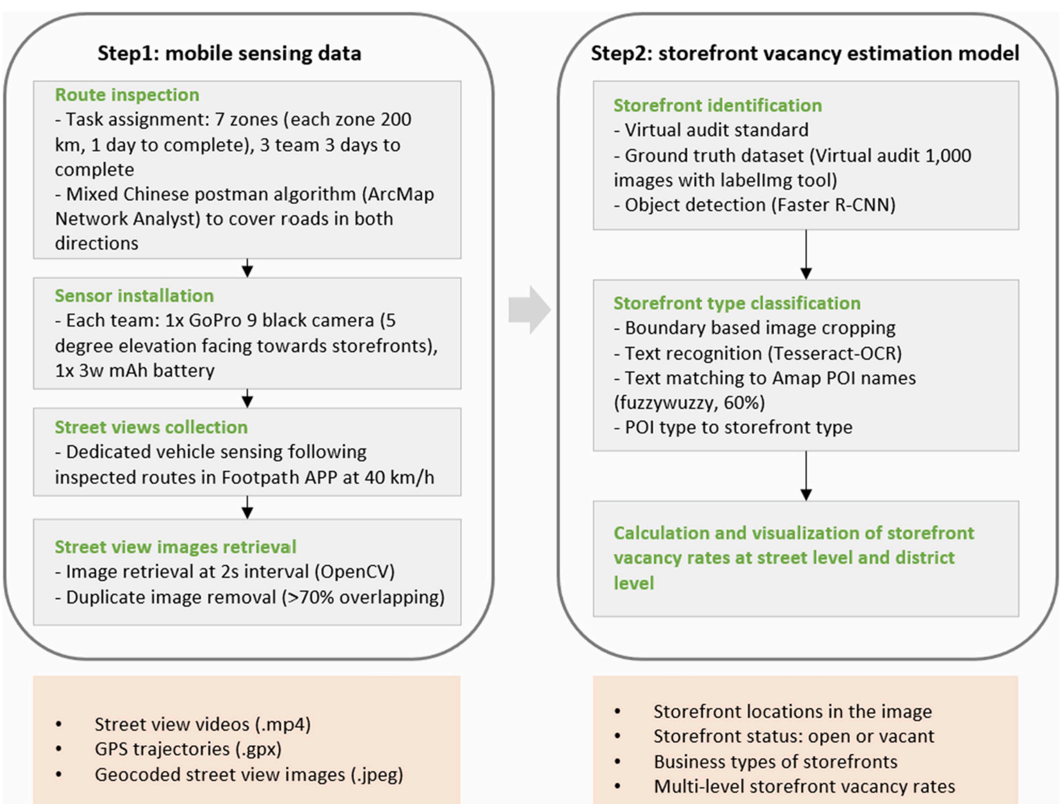

Fig. 2. Data collection and vacancy estimation model.

includes route planning, device installation, data storage, and processing; the second step is to construct a storefront vacancy estimation model, which mainly includes the virtual audit criteria of storefront vacancy, storefront status identification, business type classification, and vacancy rate calculation. The details are described in the following sections.

3.3. Mobile sensing data

The mobile sensing images are street view images obtained through the mobile sensing campaign (Li, Meng, Zhao, Li, & Long, 2023; Zhang, Zhao, Li, Long, & Liang, 2023), which was conducted from March 23 to March 27 in 2022 (see Fig. 3). Retail stores such as convenience stores, supermarkets, and department stores typically operate for extended hours every day, including weekends and public holidays, from 9 am to 10 pm. Although the financial institutions were closed on weekends, we manually labeled as open if there was no obvious damage to the exterior of the building or the shop front. So, we collected data continuously from the 23rd to the 27th. Weather conditions were considered in the experimental design and the collection was carried out in March with local temperatures ranging from -3° to 10°. The weather forecast was confirmed in advance to avoid rain and snow and to ensure that the objects in the images were clear, so the weather during data collection was sunny or cloudy. The data collection was conducted between 9 am and 4 pm to avoid congestion during the morning and evening rush hours which reduces the efficiency of the acquisition. First, route planning involved assigning tasks according to the availability of agents and designing navigation routes to cover the whole street network. This study aimed to evaluate the commercials along the main streets, as well as commercial activities inside the neighborhoods, which covered the main roads and interior roads with a total length of 1446 km based on the road data obtained from Amap, a local navigation company in China. The collection task was divided into seven tasks based on a daily workload of approximately 200 km with 7 working hours and a speed of 40 km/h. The speed was designed to ensure a reasonable image sampling interval, as GPS returns coordinate positions every 1 s, so the sampling interval is about 11.11 m (40 km/h times 1 s); also, high speeds can lead to blurred images. In the actual data acquisition process, the speed of acquisition is affected by road conditions, such as congestion. According to our statistics on the final GPS routes, the actual driving speed varies from 29.12 km/h to 52.73 km/h, with an average speed of 41.26 km/h. Three volunteers with dedicated vehicles were recruited to collect data in parallel, ending up with 3 days for all data collection. For each region, the navigation routes were calculated using the ArcMap Network Analysis tool to ensure that all planned routes in the region were traveled with the least time cost. The final navigation route in .kml format was imported into the navigation application "Footpath" in a mobile phone. To cover the entire space in a given area, the route planning algorithm (also known as the Chinese postman problem) is usually applied to compute optimized route paths, or a mixed Chinese postman problem (MCP) is used for mixed paths (Minieka, 1979). Mixed paths are considered in this study because there are three types of streets: one-way streets that can only be passed in one direction, narrow streets that can accommodate both directions but only need to be passed once, and wide streets whose agents must serve both sides of the street separately. Assuming that agents move at varying speeds between two road intersections, and that travel times between the intersections should exhibit exponential variation within a range of 0 to 10 units, the problem at hand can be reformulated as a linear problem.

$$MinimumZ = \sum \sum t_{ij} x_{ij} \tag{1}$$

Subject to

$$l_{ij} \le x_{ij} \le u_{ij}$$

where t_{ij} is a stochastic variable for the traverse time from node i to node j, x_{ij} is the travel speed from node i to node j, l_{ij} is the lower bound of the speed from node i to node j, and u_{ij} is the upper bound of the speed from node i to node j.

During the collection, each volunteer was provided a GoPro 9 camera with a 3 W mAh charging battery to support a whole-day data collection. The GoPro was fixed to the right window shield with a supporter, maintaining a 5° elevation angle and facing toward the commercials along the roads. The videos were recorded in video mode with a standard view setting of 1920×1080. Meanwhile, the Footpath application was initiated to start navigation while recording the GPS routes along the collection routes.

The collected dataset contained 7 GPS-recorded routes (.gpx), each corresponding to a region, and 142 video files (.avi) recording storefront videos, due to a 12-min limit for each GoPro video. In the data processing, to ensure a spatial interval of 25 m, the image frames were extracted from the video every 2 s, and timestamps were recorded using OpenCV, a Python library for manipulating video and image data. Similarly, GPS coordinates were extracted every 2 s with timestamp and geographic coordinates using the Python gpxpy library. On this basis, the geographic coordinates and images were matched by timestamps. For duplicate images resulting from traffic light stops, an image similarity check was performed to remove images with >70% similarity, where the similarity value is the average of the histogram similarity of the three RGB channels. Finally, 64,264 georeferenced images were obtained at approximately 25-m intervals.

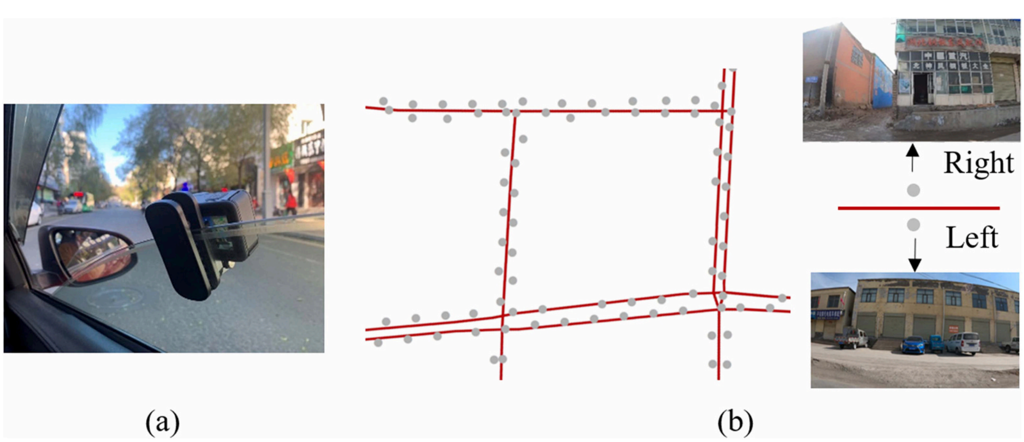

Fig. 3. Vehicle-based mobile sensing data: (a) Installation of cameras in taxis facing the storefronts, (b) from the routes to the street points, covering streetscapes on both sides of the road.

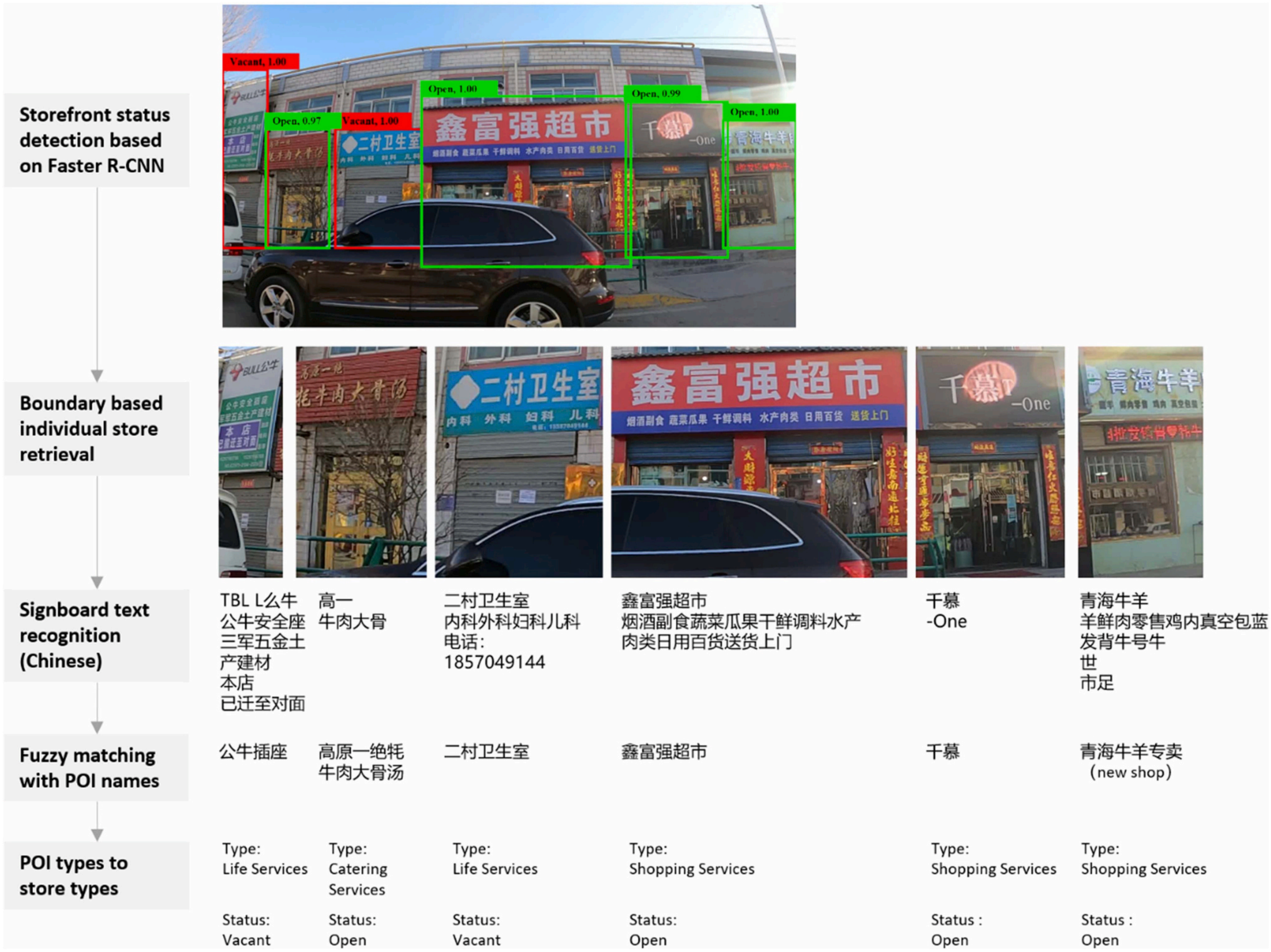

Fig. 4. Identification of store status and types.

3.4. *Storefront vacancy estimation model*

The flowchart for the storefront vacancy estimation Model is depicted in Fig. 4, providing a visual representation of the model's structure and sequential processes.

3.4.1. *Definition of storefront vacancy*

This study analyses storefronts facing main streets that have an impact on street design and street vibrancy, including pedestrian-oriented and auto-oriented corridors but excluding storefronts in the shopping malls that do not face the streets. The street design involved three levels of subdivision for the street edge: Plinths that are defined based on their morphology (the ground floors of buildings), segments defined based on territorial boundaries (representing areas of territorial ownership), and microsegments were defined based on spatial criteria, specifically areas of the space that were separated by pillars or partitions (Simpson et al., 2022). Considering that storefront closures are generally the act of the owner, the second scale (i.e., the territorial part within the plinth expressed through individual and group ownership) is defined as the unit of analysis. The duration of a commercial vacancy is a significant characteristic to consider. Typically, a vacancy that lasts for less than a year is referred to as a frictional vacancy. It is essential to have information about such vacancies to enable proper market functioning, ensuring that individuals and businesses seeking spaces do not have to wait needlessly and their requirements are met promptly. However, the literature does not provide a definitive answer on the duration of long-term (structural) vacancies. Given that the information obtained by this study is a snapshot of the city's storefronts and that the duration can only be determined if a subsequent scan is performed at another time, the study can be regarded as exploratory in nature. The virtual audit criteria for storefront vacancy are described in detail in Table 1.

3.4.2. *Identification of storefront status*

Based on the specified storefront vacancy identification criteria, the deep learning method was used to automatically identify stores from images and determine their open or vacant status. In addition, based on the geographic coordinates of the images, the store locations were determined. Faster R-CNN is the third-generation R-CNN family of convolutional neural network (CNN) models. It achieves the highest accuracy on several object detection tasks and was, therefore, chosen as the storefront recognition model in this study (Ren et al., 2015). We trained the faster R-CNN storefront detection model on our labeled dataset, which included 1000 randomly selected images from the collection dataset and 811 labeled storefronts. Each of these images significantly varied in store size, business type, and the number of stores, making this labeled dataset useful for training. The storefront detection model was implemented in Keras with the architecture of ResNet-50 pre-trained on the ImageNet model. Initially, a stochastic gradient descent (SGD) solver was employed for 50 iterations with a base learning rate of 0.001, followed by an additional 50 iterations with a reduced base learning rate of 0.0001.

Table 1
Virtual audit criteria for storefront vacancy.

Classes	Definition	Image samples
Vacant store	Closed not in business, or with the words demolition, sale, transfer, etc. The exterior usually has a roll-up door locked or signboard broken.	
	Signboards spanning multiple storefronts count as one label;	
	Stores without signboards count as one label per storefront.	
Open store	The state of being in business, such as the roll-up door is open, or the door is not open, but there are goods or people inside and signs of business.	
	Multi-story stores with inconsistent signboards are marked more than one.	
	The existence of partial vacancy is marked two; the overall store is open and part of the store is vacant.	

3.4.3. *Identification of business type*

For the identified vacant stores, their business types were further analyzed according to the point of interest (POI) classification criteria. This was done to analyze the categories of vacant stores. The image cropping method was applied to the rectangular shapes of stores identified in the previous section, to retrieve individual store images to obtain the text information on each store and avoid interference from other stores. To identify the store information from the individual store images, we compared different types of optical character recognition (OCR) software— Google Docs OCR, Tesseract, ABBYY FineReader, and Transym—and selected Tesseract because of its highest accuracy rate (Smith, 2007). Tesseract is a suite of algorithms that have been highly optimized, and the module "Python-tesseract" utilizes tesseract-OCR to convert images into text within a Python environment. To convert the given image to simplified Chinese text, we utilized the method "image_to_string(image, lang = 'chi_sim')" as defined. The recognition content contained the basic information of the store, including the signboard name, business type, and contact information. To classify the store information, the classification standard of POIs was referred. The 2,356,972 POIs (2018–2020) within the study area were crawled from Amap, containing POI names, addresses, and type information. The FuzzyWuzzy package in Python was used for fuzzy matching, setting the matching rate to above 60%. The defined function "token_sort_ratio()" was used to avoid problems caused by an inconsistent order.

3.4.4. *Calculation of the storefront vacancy rate*

The storefront vacancy rate (SVR) for a street is defined as the sum of all vacant and open stores on the street divided by the total number of stores.

$$SVR_i = \frac{\sum VS_j}{\sum OS_j + \sum VS_j} \tag{2}$$

where VS_j and OS_j are the numbers of vacant stores and open stores at street point j, and SVR_i indicates the vacancy rate of street i, ranging from 0% to 100%.

The vacancy rate for a district is defined as the average of the vacancy rates for all streets and the calculation can be expressed as follows:

$$SVR_k = \frac{\sum SVR_i}{I} \tag{3}$$

where I is the overall count of streets in district k and SVR_k is the vacancy rate of district k, ranging from 0% to 100%.

3.4.5. *Identification of commercial structure*

Based on the observed spatial heterogeneity of storefront vacancy rates (SVRs) across the city, the determinants of storefront vacancy were further investigated (see Table 2) and an Ordinary Least Squares (OLS) regression model was constructed to explore the impact of various urban geographic and socioeconomic factors on storefront vacancy. The literature identifies common causes of storefront vacancy as business underperformance, realignments of trading areas, closures due to bankruptcy, and the seizing of emerging opportunities (Cavan, 2016). Apart from external economic factors and social issues, geographical characteristics such as the accessibility (Geurts & Black Jr, 2015; Lee

Table 2
Explanatory variables for OLS regression.

Category	Explanatory variables	Definition	Data sources	References
Geographical features	Distance to urban center (km)	Distance to the Xining railway station	Amap POI data	(Geurts & Black Jr, 2015; Lee et al., 2021)
	Distance to nearest bus stop (km)	Distance to nearest bus stop	Amap bus stop data	(Geurts & Black Jr, 2015; Tselios et al., 2018)
	Distance to nearest bus route (km)	Distance to nearest bus route	Amap bus route data	(Lee et al., 2021; Stringer, 2019)
	Distance to rivers (km)	Distance to rivers	Amap river data	(Geurts & Black Jr, 2015; Tselios et al., 2018)
	Distance to parks and recreation facilities (km)	Distance to parks and recreation facilities	Amap AOI data	(Geurts & Black Jr, 2015; Lee et al., 2021; Stringer, 2019)
	Distance to nearest commercial zoning (km)	Distance to eight commercial zonings	Xining urban planning document	(Geurts & Black Jr, 2015; Lee et al., 2021; Stringer, 2019; Talen & Park, 2022; Tselios et al., 2018)
	Road density (km/km^2)	Total road length divided by buffer area	Amap road network data	(Geurts & Black Jr, 2015; Stringer, 2019)
	Mean NDVI (ratio: 0–1)	Mean NDVI of the buffer area	USGS Landsat	(Stringer, 2019)
Demographical features	Population density (people/km^2)	Population density per square kilometer	WorldPop	(Geurts & Black Jr, 2015; Lee et al., 2021; Stringer, 2019; Talen & Park, 2022; Tselios et al., 2018)
	Median housing age (year)	Median housing age	Anjuke real-estate data	(Stringer, 2019)
	Neighborhood housing price (CNY)	Average housing price	Anjuke real-estate data	(Stringer, 2019)

et al., 2021; Stringer, 2019; Tselios et al., 2018) and location of commercial districts (Geurts & Black Jr, 2015; Lee et al., 2021; Stringer, 2019; Talen & Park, 2022; Tselios et al., 2018), as well as demographic features such as housing prices (Stringer, 2019), escalating rents, and a transient population (Bushnell, 1931; Deng & Ma, 2015; Saraiva, Sá Marques, & Pinho, 2019; Tokosh & Chen, 2022), significantly influence the survival and closure rates of businesses. For each identified variable, data sources were designated to ensure accuracy and consistency. Primary data were sourced from Amap's multiple datasets, an extensive geographic information system that provides real-time data on transportation and urban layout. Additionally, we utilized information on population density from the WorldPop database, as well as data on housing age and prices from the Anjuke real estate database. These sources provided us with high-resolution and up-to-date geographic and socioeconomic data, ensuring the reliability of our analysis results. Finally, informed by the theoretical frameworks outlined in the literature and subsequent data availability screening, we selected 11 explanatory factors for quantitative analysis, as indicated in Table 2, and examined their relationships with SVR.

Specifically, for the operationalization of each variable, the Xining railway station was designated as the urban center reference point. Bus stop data were extracted from the 2018–2020 Amap POI data. The locations of the eight commercial districts were obtained from Xining urban planning documents, including train station and Dongguan Street in Chengdong district, Xiaoqiao and Minghui Cheng in Chengbei district, Dashizi and Shangri-La in Chengzhong district, and Limeng and Wanda Plaza in Chengxi district. Road density calculations were based on the total road length within a 1-km buffer zone of each road segment divided by the area of the zone, using Amap network data. The mean Normalized Difference Vegetation Index (NDVI) was determined by averaging NDVI values from Landsat 8 satellite imagery within a 1-km buffer zone surrounding each road segment. Population density figures were derived from the 2020 WorldPop world population data, calculated by dividing the total population within intersecting grids of a 1-km road buffer by the area of these grids.

The association between street-level SVRs and related factors was analyzed using an OLS regression model, based on the following equation:

$$SVR = \sum G_k g_k + \sum D_t d_t + b \tag{4}$$

where SVR is the street-level SVR and the dependent variable of the regression model; G_k and D_t are the geographic and demographic variables, respectively; g_k and d_t are their coefficients; and b is the model constant.

4. Results

4.1. Accuracy of the storefront vacancy estimation model

Next, we evaluated the accuracy of two parts of the model: storefront recognition and signboard text recognition. Fig. 5 illustrates the learning results of storefront identification. The accuracy of storefront identification was described by two indicators: accuracy and kappa coefficient. Accuracy measures the proportion of correctly predicted instances out of all instances. Kappa coefficient measures the agreement between the actual and predicted classifications, taking into account the possibility of chance agreement and is calculated as follows.

$$Kappa = \frac{Po - Pe}{1 - Pe} \tag{5}$$

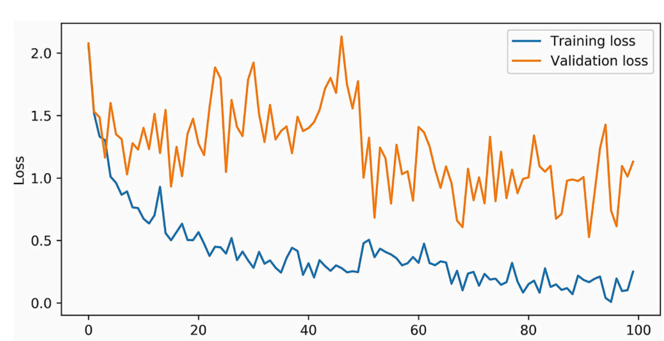

Fig. 5. Learning accuracy and loss curves, validation, and training loss in storefront identification.

where *Po* (Observed Agreement) is the proportion of instances that were correctly classified, which is equal to Accuracy. *Pe* (Expected Agreement) is the proportion of instances that would be expected to be correctly classified by chance.

The validation accuracy decreased sharply in the first few epochs; as the learning proceeded, the curve fluctuated and the accuracy became stable at approximately 83.0%. This is reflected in the fact that the validation and training loss reduced in the beginning, fluctuated over the period, and eventually stabilized. The overall accuracy rate was 85.0%, slightly below the validation accuracy. The recall and accuracy for the vacant stores were 92.6% and 80.3%, respectively, as shown in Table 3. The value of 0.71 for the kappa coefficient demonstrated a high level of concordance between the forecast and actual results (0.6–0.8).

In terms of signboard text recognition, line recognition accuracy (LRA) was used as the evaluation metric. The successful recognition of a line of text is defined as the correct recognition of each character in the text. LRA is defined as the proportion of lines that are correctly recognized and can be calculated as follows:

$$LRA = \frac{|l|}{|L|} \tag{6}$$

where l and L are the count of accurately recognized lines and the overall number of lines identified, respectively.

The text recognition accuracy was validated on the labeled storefront dataset, achieving an LRA of 76.42%. In addition, to determine the best choice of our model to achieve high performance, our results were compared with those obtained by the other well-known OCR software: Google Docs OCR, ABBYY FineReader, and Transym. As shown in Table 4, Tesseract outperformed the other software and proved our choice. In particular, Tesseract is open-source and trainable, whereas other products are mature software and do not support improvements.

The business status and store type when using the model for prediction were double checked and manually corrected for false positives. Errors in store business status were mainly in false positives due to trees along the street, bus stop signs and newsstand occlusions. The errors in store types were mainly in text recognition, where special fonts were harder to recognize and obscured resulting in incomplete names, which needed to be corrected manually.

4.2. Storefront vacancy in 2022

A total of 93,069 storefronts were identified, comprising 25,488 vacant storefronts and 67,581 operational storefronts. The distribution of these storefronts is illustrated in Fig. 6. We found an average SVR of 30.0% for stores in the urban central area of Xining in 2022. However, SVR is spatially heterogeneous across the districts and towns, among which the streets in the town of Lushaer had the highest average SVR (45.6%), followed by Chengbei (37.6%), Duoba (32.2%), Chengzhong (30.9%), Chengdong (27.1%), and Chengxi (21.6%). As shown in Fig. 8 (a), the percentage of streets with the SVR in the lower quantile (below 20%) was 51.2% of all streets, implying a relatively high occupancy rate. These streets were mainly located in the old downtown area. Meanwhile, streets with SVRs above 60% accounted for 19.4% of all streets. These streets had very low occupancy rates and were mainly located in newly developed areas on the urban peripheries. Overall, we observed an upward gradient of SVRs from the center to the periphery. It was found that 7.1% of the streets were completely unoccupied, with SVRs equal to 1. These were small alleys close to residential areas. Only 22 of the 70 streets with total vacancies had >10 stores, while the other 48 streets had <10 stores or only 1 store.

Table 3
Confusion matrix for storefront identification with faster R-CNN.*

Storefronts		Ground truth			
		Open stores	Vacant stores	Total	Precision
Prediction	Open stores	315	32	347	90.8%
	Vacant stores	90	374	464	80.6%
	Total	405	406	811	
	Recall	77.8%	92.1%		Overall: 85.0%

* Kappa coefficient: 0.71, indicating substantial agreement.

Table 4
Comparison of Tesseract with other OCR software. All experiments were performed on our labeled storefront dataset.

Method	LRA
Tesseract	76.42%
ABBYY FineReader	73.68%
Google Docs OCR	52.86%
Transym	41.36%

Fig. 7 depicts the business types of storefronts. Of all the stores in 2022, stores operating in shopping, catering, and life services had the highest number of vacancies (i.e., 7495, 6148, and 3795, respectively). The results of a further analysis of store type (i.e., the POI's secondary categories) showed that in the shopping category, franchise stores had the highest proportion of vacancies (19.4%), followed by home building material markets (17.7%) and convenience stores (10.1%); in the catering category, the highest proportion of vacant stores was in Chinese food restaurants (54.8%), followed by fast-food restaurants (15.0%) and tea houses (1.8%); in the lifestyle category, beauty salons had the highest proportion of vacancies (22.2%), followed by lottery stores (5.1%) and bath and massage centers (4.7%). There were 2001 identified stores that failed to classify the store type, of which 1689 were vacant and 412 were open. This is due to some shops not having signage so shop name information could not be obtained, or new shops opening in 2022 not being able to match 2018 POI data.

4.3. Commercial structure of storefront vacancy

Table 5 presents the outcomes of the OLS regression model. The *p*-value indicates that seven explanatory variables were significant (i.e., Sig < 0.05). The factors having the greatest impact on storefront vacancy were, in order of importance, distance from commercial zonings, population density, and distance from urban centers, as indicated by the standardized coefficients (Beta). This model explained 41.2% of the variance of SVR, as indicated by the value of R^2. Meanwhile, no significant multicollinearity was observed among all variables, as the collinearity statistics suggested tolerance >0.1 and VIF < 3.

5. Discussion

5.1. Comparison of storefront vacancy data

We compared the proposed approach with two other types of storefront vacancy data available in the Chinese context, namely commercial street view images and store transaction data from real-estate websites, as shown in Table 7. Online trading of storefronts is not particularly common in China, and although considerable transaction information is available, such as monthly rent, construction area, and operating status, which can be obtained directly from websites without relying on algorithms, the number of store transactions is very small, with only 997 stores in the whole of Xining, of which only 265 are actually vacant. The main advantage of the commercial street view images is global coverage, partly covering many years, from which the storefront status and types can be evaluated using our developed image processing algorithm. However, the main drawback is that commercial companies no longer update their data. Xining is four years out of date in

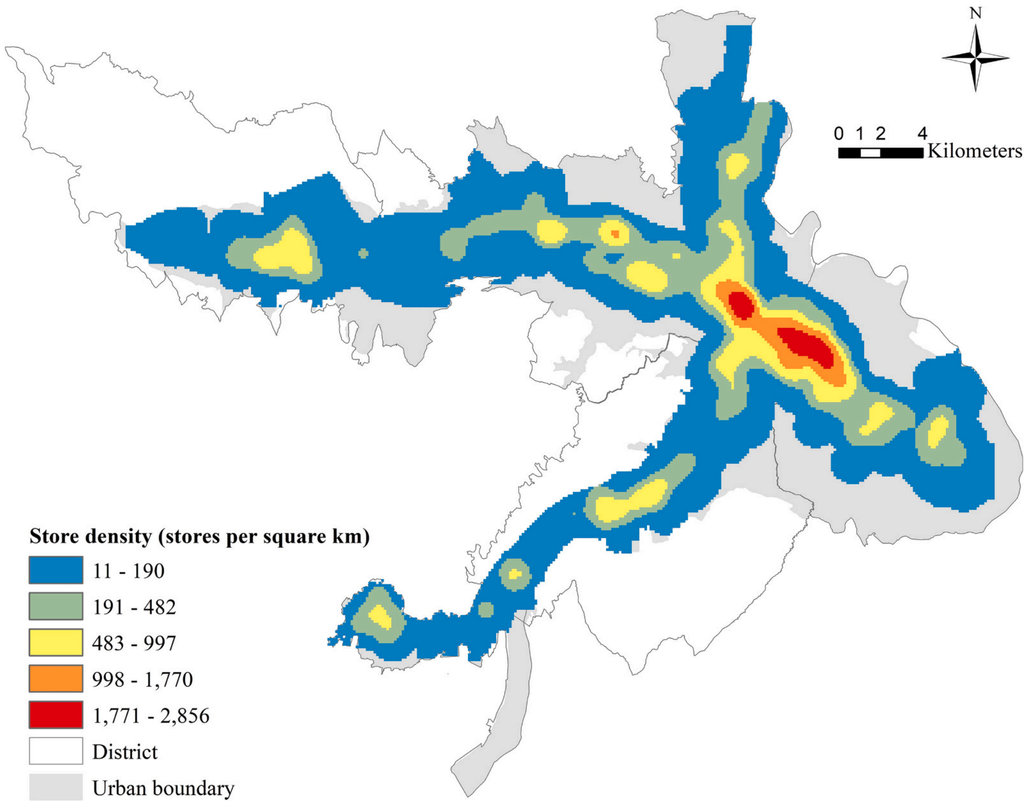

Fig. 6. Identified store density (kernel density) in 2022.

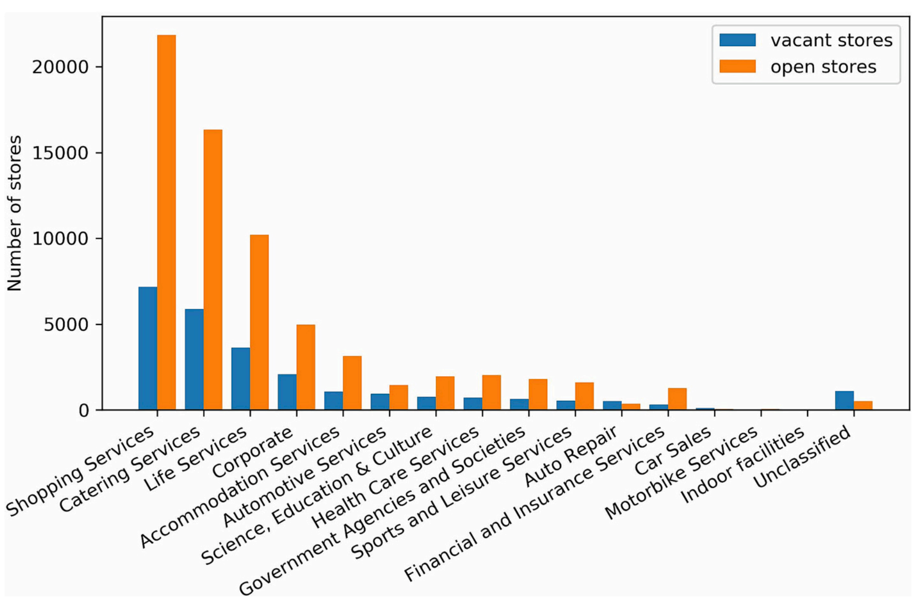

Fig. 7. Store types of open and vacant stores in 2022.

terms of time, and the spatial coverage cannot be extended with the expansion range of the city; for example, the suburban towns of Duoba and Lushaer are not fully covered. In contrast, our proposed algorithm is flexible, can cover a wide area at a low cost, and is planned to be multitemporal in the future, covering 93,069 stores in Xining with little labor cost and car rental cost.

5.2. *Changing dynamics of storefront vacancy*

To further observe the temporal changes in SVRs, we obtained street view images from Baidu Map, with a time machine function, to observe the changes at some points over the years. The commercial street view images in the study area of Xining were obtained in this study. First, the survey points on the roads were created on the map every 25 m, which matches the spacing of the mobile sensing images. Every point was associated with latitude and longitude coordinates, as well as additional

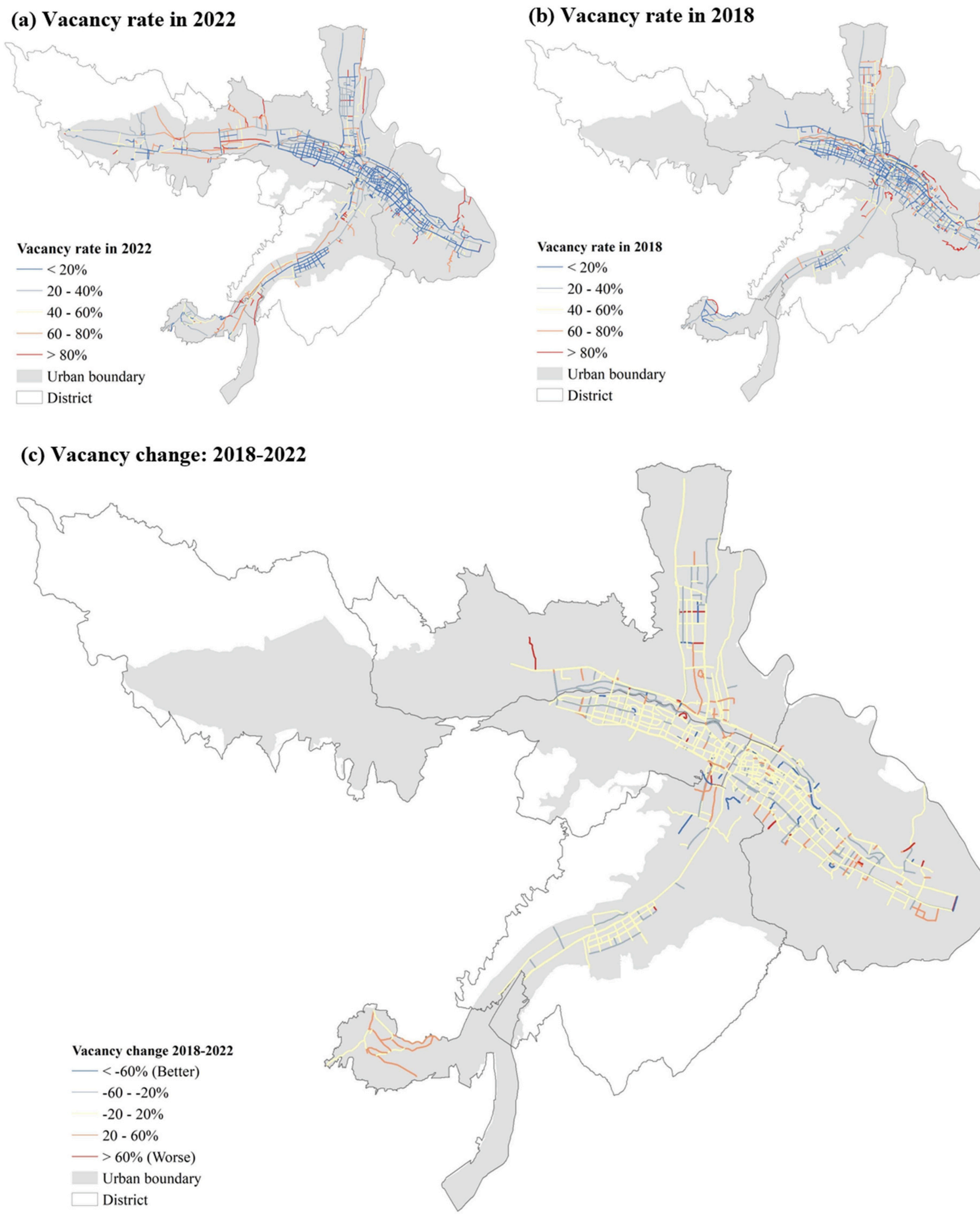

Fig. 8. Spatial distribution of street-level storefront vacancy and changes over the years.

Table 5
OLS regression results for street-level storefront vacancy.

Independent variables	Beta	*p*-value	Tolerance	VIF
Constant	0.273			
Distance to urban center (km)	0.312	0.001	0.582	1.731
Distance to nearest bus stop (km)	0.234	0.001	0.899	1.110
Distance to nearest commercial zoning (km)	0.475	0.000	0.487	2.049
Road density (km/km^2)	−0.176	0.014	0.906	1.108
Population density (people/km^2)	−0.309	0.002	0.539	1.826
Neighborhood housing price	−0.301	0.000	0.901	1.112
Mean NDVI (ratio: 0–1)	0.184	0.018	0.807	1.238

Dependent variable: street-level SVR.
Adjusted R^2: 0.412.

Table 6
Storefront statistics.

Year	Points/images	Open stores	Vacant stores	Total	Streets	SVR of streets	Data sources
2014	34,582	34,815	8460	43,275	678	20.7%	Commercial street view images
2016	32,038	28,659	7366	36,025	608	20.2%	Commercial street view images
2018	53,836	48,982	13,733	62,715	852	21.8%	Commercial street view images
2022	76,676	67,581	25,488	93,069	981	30.0%	Mobile sensing images

Table 7
Comparison of storefront vacancy data.

	Real-estate data (Anjuke.com)	Commercial street view images (Baidu.com)	Mobile sensing images (our approach)
Time coverage	2022	2014, 2016, 2018 No further updates	2022 Determined by data collection plans
Spatial coverage (no. of vacant stores/total no. of stores)	2022: 265/997	2014: 8460/43,275 2016: 7366/36,025 2018: 13,733/62,715 Urban central area of global cities	2022: 25,488/93,069 Customized or full coverage, determined by planned routes
Store status	Vacant or still in business	Vacant or open	Vacant or open
Store types	12 categories: Catering, beauty salon, shopping, leisure and entertainment, supermarkets, household services, electrical appliances and communications, auto repair, medical equipment, furniture and construction, education and training, hotels, others	POI level-1 type, 23 categories	POI level-1 type, 23 categories
Monthly rent	100–300,000 CNY	Not available	Not available
Construction area	16–5700 m^2	Not available	Not available
Cost	Data download	Data download, image processing algorithm development or manually virtual audit	Cost of cameras, labor cost for data collection, vehicle hire cost, and cost for image processing algorithm development

parameters, such as a size of 1280 × 720 pixels, a pitch angle of 0°, a heading angle of 270° facing the storefront, and a request key, on the basis of which a URL was concatenated and obtained to download the street view images over the years with the support of Python. Finally, 34,582, 32,038, and 53,836 street view images were collected for 2014, 2016, and 2018, respectively, representing equal numbers of investigation points. Based on the acquired commercial street view images, our proposed SVR estimation model was adopted for storefront identification and SVR calculation.

The estimation results from previous years and our identification results are presented in Table 6 and Fig. 8 (b). The table shows a significant increase in SVR in 2022, which remained at approximately 21.0% until 2018 but increased to 30.0% in 2022. Fig. 8 (c) displays streets with increased and decreased vacancy rates, represented in red and blue respectively. Thus, we can reasonably assume that the COVID-19 outbreak at the end of 2019 had a knock-on effect on the real

(a)

(b)

Fig. 9. Minghuicheng community stores in 2018 (a) and 2022 (b).

economy, as demonstrated in the literature (Ahsan, 2020; Mouratidis & Yiannakou, 2022).

5.3. Case study: Minghuicheng community business

Community business was greatly affected by the epidemic. Minhuicheng, one of the eight major commercial zones in Xining, not only has shopping, restaurants, bookstores, coffee, banks and other businesses, introducing major chain brands, but also provides a number of specialized services such as childcare and car maintenance to meet the personalized, diversified and specialized needs of residents. Before the epidemic, as a major community business in Xining, it served not only the residents of its own community, but also radiated to the surrounding communities, making the commercial area always popular. Prior to the epidemic, the street vacancy rate for the main road, Chaoyang West Road was 18.2% in 2018. However, the vacancy rate for 2022 during the epidemic was 59.4%, with an increase of 41.2% (see Fig. 9).

Although the vacancy rate rose sharply during the epidemic, in the post-epidemic era, the daily life of residents may change, such as more attention to personal health, more concern for quality of life, more rational and cautious consumption, and a more focused scope of daily life, all of which require a better environment and quality of community living. In addition to the improvement of public facilities in the city, the environment and quality of community life also depend to a large extent on the continuous optimization of community businesses.

6. Conclusions

To address the data paucity of individual-level storefront vacancy and the underestimation of storefront vacancy research, this study develops a framework for assessing the SVR based on mobile sensing data and a computer vision approach. At the data level, the mobile sensing images are obtained at low cost, on a large scale, and with high efficiency, to support the construction of a fine-scale storefront vacancy database, which outperforms other types of data in terms of spatial coverage and update time. At the methodological level, a storefront vacancy estimation model is developed based on deep learning and text recognition techniques to assess the commercial structure of storefronts at multiple scales, inferring the total number of storefronts, their operating conditions, business categories, and multiscale vacancy rates. A total of 93,069 stores are identified in the case city, of which 25,488 are vacant, demonstrating the efficiency of the model compared to only 265 vacant stores on China's largest real-estate website. The model can be applied to not only the mobile sensing images collected in this study but also other commercial street view images, such as Google and Baidu. In terms of the experimental findings, the overall SVR in the case city of Xining is relatively high, reaching 30.0%, compared to other cities, such as New York, which was at 11.3% in 2020. The increasing vacancy rate, which rose from 21.8% in 2018 to 30.0% in 2020, highlights the need for the government and urban management authorities to take appropriate measures. Stores in shopping, catering, and life services categories have the highest number of vacancies. Distance to commercial zonings has the greatest impact on SVRs, followed by population density and distance to urban centers. These findings can provide a reference for urban planners to optimize the nature of land use and commercial structure configuration.

In the future, another mobile sensing campaign is planned for autumn as an approach to obtain a multitemporal database of storefront vacancy, which can be used to assess the vacancy duration, often as an indicator of the degree of resource waste. Meanwhile, a wide range of factors influencing vacancy in diverse and complex ways can be further analyzed, and urban planning strategies to address vacancy issues should be well considered and differentiated.

CRediT authorship contribution statement

Yan Li: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Ying Long:** Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by National Natural Science Foundation of China [Grant Number 52178044], with a focus on intelligent identification, mechanism analysis and planning countermeasures of various types of urban vacancy. Additionally, it was supported by the urban renewal project in Xining City, which includes research on the identification of urban spatial issues and the development of urban renewal management measures, funded by the China Urban Construction Design & Research Institute. The code, pre-trained models, training samples, and recognition results have been made publicly available for download on Mendeley Data: https://doi.org/10.17632/9v37g2y9fc.1.

References

Ahsan, M. M. (2020). Strategic decisions on urban built environment to pandemics in Turkey: Lessons from COVID-19. *Journal of Urban Management, 9*(3), 281–285.
Berman, B. (2019). Flatlined: Combatting the death of retail stores. *Business Horizons, 62*(1), 75–82.
Bromley, R. (2000). Street vending and public policy: A global review. *International Journal of Sociology and Social Policy, 20*(1/2), 1–28.
Bushnell, J. D. (1931). Problems in analyzing vacancy statistics. *Journal of the American Statistical Association, 26*(173A), 47–52.
Cavan, D. R. (2016). Analyzing retail store closures. *The Appraisal Journal, 84*(4), 353–360.
Chin, H., & Chow, A. (2012). *The case for China retail: Issues and opportunities. Parsippany, NJ*.
Chowdhury, J., & Sarkar, S. (2017). The financial impact of retail store closure announcements. *International Journal of Physical Distribution and Logistics Management, 47*(6), 536–556.
Deng, C., & Ma, J. (2015). Viewing urban decay from the sky: A multi-scale analysis of residential vacancy in a shrinking US city. *Landscape and Urban Planning, 141*, 88–99.
Geurts, T. G., & Black, J. F., Jr. (2015). Analyzing commercial real estate market and property data: Techniques for the classroom using CoStar information services. *Journal of Real Estate Practice and Education, 18*(1), 55–76.
Home, R. K. (1983). Inner city vacant land: UK policies. *Cities, 1*(1), 59–70.
Hui, E., Yiu, C. Y., & Yau, Y. (2007). Retail properties in Hong Kong: A rental analysis. *Journal of Property Investment & Finance, 25*(2), 136–146.
Kanayama, Y., & Sadayuki, T. (2021). What types of houses remain vacant? Evidence from a municipality in Tokyo, Japan. *Journal of the Japanese and International Economies, 62*, Article 101167.
Lee, J., Kim, H., & Kim, H. (2021). Commercial vacancy prediction using LSTM neural networks. *Sustainability, 13*(10), 5400.
Lee, R. J., & Newman, G. (2021). The relationship between vacant properties and neighborhood gentrification. *Land Use Policy, 101*, Article 105185.
Li, Y., Meng, X., Zhao, H., Li, W., & Long, Y. (2023). Identifying abandoned buildings in shrinking cities with mobile sensing images. *Urban Informatics, 2*(1), 3.
Minieka, E. (1979). The Chinese postman problem for mixed networks. *Management Science, 25*(7), 643–648.
Mouratidis, K., & Yiannakou, A. (2022). COVID-19 and urban planning: Built environment, health, and well-being in Greek cities before and during the pandemic. *Cities, 121*, Article 103491.
Myers, D., & Wyatt, P. (2004). Rethinking urban capacity: Identifying and appraising vacant buildings. *Building Research and Information, 32*(4), 285–292.
Oskam, M. C. (2021). *Drivers of retail vacancy pre-Covid19: The effect of building and location factors*.
Remøy, H. T., & van der Voordt, T. J. (2007). A new life: Conversion of vacant office buildings into housing. *Facilities, 25*(3), 88–103.
Ren, S., et al. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. In *, 1. Proceedings of the 28th international conference on neural information processing systems* (pp. 91–99). Montreal, Canada: MIT Press.
Saraiva, M., Sá Marques, T., & Pinho, P. (2019). Vacant shops in a crisis period–A morphological analysis in portuguese medium-sized cities. *Planning Practice and Research, 34*(3), 255–287.
Sim, L.-L., Yu, S.-M., & Malone-Lee, L.-C. (2002). Re-examining the retail hierarchy in Singapore: Are the town centres and neighbourhood centres sustainable? *The Town Planning Review*, 63–81.
Simpson, J., et al. (2022). Street edge subdivision: Structuring ground floor interfaces to stimulate pedestrian visual engagement. *Environment and Planning B: Urban Analytics and City Science, 49*(6), 1775–1791.
Smith, R. (2007). An overview of the Tesseract OCR engine. In *Ninth international conference on document analysis and recognition (ICDAR 2007)*. IEEE.
Stringer, S. M. (2019). *Retail vacancy in New York city: Trends and causes, 2007-2017*.
Talen, E., & Park, J. (2022). Understanding urban retail vacancy. *Urban Affairs Review, 58*(5), 1411–1437.
Tokosh, J., & Chen, X. (2022). Did the Macy's in my mall close? Revisiting the closures of Macy's, Sears, and JCPenney stores. *GeoJournal, 87*(4), 2551–2575.
Tselios, V., Lambiri, D., & Dolega, L. (2018). Performance within a recession: The converging trajectories of retail centres in the UK. *Regional Science Policy & Practice, 10*(4), 347–365.
Wang, J., et al. (2020). *Bankruptcy and the COVID-19 crisis*. Available at SSRN 3690398.
Yeates, M., & Montgomery, D. (1999). The changing commercial structure of non-metropolitan urban centres and vacancy rates. *Canadian Geographer/Le Géographe canadien, 43*(4), 382–399.
Zhang, D., Zhu, P., & Ye, Y. (2016). The effects of E-commerce on the demand for commercial real estate. *Cities, 51*, 106–120.
Zhang, Y., Zhao, H., Li, Y., Long, Y., & Liang, W. (2023). Predicting highly dynamic traffic noise using rotating mobile monitoring and machine learning method. *Environmental research, 229*, 115896.
Zukin, S. (1996). Space and symbols in an age of decline. In *Re-presenting the city: Ethnicity, capital and culture in the twenty-first century metropolis* (pp. 43–59).
van Zweeden, J. (2009). *Retail vacancy in Dutch city centers: How can differences in retail vacancy between cities be explained*. Erasmus University.