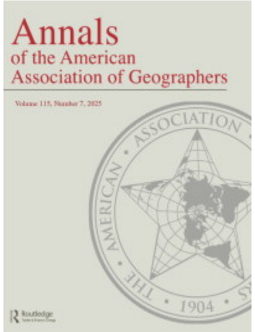

Annals of the American Association of Geographers

ISSN: 2469-4452 (Print) 2469-4460 (Online) Journal homepage: www.tandfonline.com/journals/raag21

From Street View Imagery to the Countryside: Large-Scale Perception of Rural China Using Deep Learning

Kunkun Zhu, Yu Gu, Yatao Zhang, Yandi Song, Zihao Guo, Xiaoqin Yan, Yao Yao, Qingfeng Guan & Xun Li

To cite this article: Kunkun Zhu, Yu Gu, Yatao Zhang, Yandi Song, Zihao Guo, Xiaoqin Yan, Yao Yao, Qingfeng Guan & Xun Li (2025) From Street View Imagery to the Countryside: Large-Scale Perception of Rural China Using Deep Learning, Annals of the American Association of Geographers, 115:7, 1720-1741, DOI: 10.1080/24694452.2025.2505682

To link to this article: https://doi.org/10.1080/24694452.2025.2505682

ANNALS OF THE AMERICAN ASSOCIATION OF GEOGRAPHERS
2025, VOL. 115, NO. 7, 1720–1741
https://doi.org/10.1080/24694452.2025.2505682

From Street View Imagery to the Countryside: Large-Scale Perception of Rural China Using Deep Learning

Kunkun Zhu,[a] Yu Gu,[b] Yatao Zhang,[c] Yandi Song,[a] Zihao Guo,[a] Xiaoqin Yan,[d,e] Yao Yao,[a,f,g,h] Qingfeng Guan,[a,f] and Xun Li[b]

[a]*UrbanComp Lab, School of Geography and Information Engineering, China University of Geosciences, China;* [b]*China Regional Coordinated Development and Rural Construction Institute, School of Geography and Planning, Sun Yat-sen University, China;* [c]*Future Resilient Systems, Singapore-ETH Centre, ETH Zurich, Singapore;* [d]*Institute of Remote Sensing and Geographic Information Systems, School of Earth and Space Sciences, Peking University, China;* [e]*Beijing Key Lab of Spatial Information Integration & Its Applications, Peking University, China;* [f]*National Engineering Research Center of Geographic Information System, China University of Geosciences, China;* [g]*LocationMind Institution, LocationMind, Inc, Japan;* [h]*Hitotsubashi Institute for Advanced Study, Hitotsubashi University, Japan*

Evaluating residents' subjective perceptions of the rural environment is crucial for formulating effective rural planning. Due to the difficulty in obtaining rural data and the limitations of traditional questionnaire survey methods, existing research mostly focuses on small-scale perception evaluations in specific areas, making it difficult to reveal rural perception characteristics at the regional scale. To address this issue, our study proposes a rural living environment perception evaluation model based on street view images and deep neural networks, achieving a quantitative evaluation of rural perception in 118 cities nationwide. We collected a large data set of rural street view images nationwide through crowdsourcing and established an index system comprising five subjective perception dimensions: wealthy, tidy, lively, habitable, and terroir. By training a multidimensional quantitative evaluation model, we comprehensively evaluated residents' subjective perceptions of China's rural environment. Furthermore, we explored the relationship between these subjective perceptions and objective socioeconomic indicators. The model achieves an average evaluation accuracy of 75 percent across five dimensions, with the wealthy dimension exceeding 80 percent. Rural environment perception is comprehensively influenced by various factors such as economic base and traditional feature protection, showing significant differences between different regions. The perception of rural environment in the eastern region is closely related to economic levels, whereas perception in the western region is more affected by infrastructure improvement and social development. Overall, this study provides scientific evidence for formulating more targeted and effective rural planning. *Key Words: human perception, rural environment, social sensing, street view.*

China's "industry-first, city-first" strategy rapidly transformed the urban landscape while neglecting rural living environments, leading to a pronounced disparity between urban and rural living conditions (Hu and Wang 2020). To address this issue, the Chinese government introduced the rural revitalization strategy in 2017—a national strategy aimed at promoting comprehensive rural development, encompassing rural economic growth, agricultural modernization, rural infrastructure development, and other areas (Y. Liu 2018; Y. Zhou, Li, and Xu 2020). Improving the rural living environment was highlighted as a critical objective within this strategy (Shi, Xu, and Ma 2023). As rural revitalization progresses, China's rural environment has been greatly improved (Fang et al. 2022). In this context, quantifying residents' subjective perceptions of the rural living environment is vital for measuring the effectiveness of environmental improvement and provides a scientific basis for devising more effective rural planning and development policies (Ho et al. 2020).

ARTICLE HISTORY
Initial submission, September 2024; revised submission, February 2025; final acceptance, March 2025

CORRESPONDING AUTHOR Yao Yao yaoy@cug.edu.cn

The rural living environment refers to the rural settlement environment involved in the daily life and production activities of rural residents. It includes both hard environments such as living conditions, infrastructure, and public service facilities, and soft environments such as economic development level, social service level, and life comfort level (R. Liu et al. 2024; H. Zhang, Liu, and Guo 2024). Compared with urban environments, rural environments often have more natural landscapes and lower building density, while also retaining more traditional characteristics in lifestyle and production methods (Y. Wang, Zhu, and Yu 2019; Y. Wang et al. 2024). Furthermore, the rural living environment in different regions also has considerable differences in economic level, infrastructure construction, social development, and more, making China's rural street view present significant diversity (Y.-R. Li et al. 2022). These disparities pose significant challenges to the subjective perception evaluation of rural environments.

Current assessments of subjective perceptions of the rural living environment primarily rely on traditional methods, such as questionnaire surveys, to capture specific groups' (e.g., villagers) perceptions of the rural environment (Y.-R. Li et al. 2022). For instance, P. Wang, Qin, and Li (2021) conducted field interviews, telephone surveys, and online surveys to explore the satisfaction levels of rural residents in the northwest region with their living environments. Xiao et al. (2022) undertook a questionnaire survey and in-depth interviews with forty-one villagers from two villages in Puyang City to analyze improvements in the rural environment. Traditional methods such as field interviews and questionnaire surveys, however, primarily depend on language and text, which might introduce recording and communication biases, potentially limiting their ability to accurately capture residents' perceptions. Furthermore, such studies are often limited by time, funding, and manpower, and can usually only cover specific areas and a small number of villages, making it difficult to apply to large-scale subjective perception evaluations.

In recent years, the emergence of street view images has provided a solid data foundation for large-scale capturing of residents' subjective perceptions. As a typical type of geo-tagged data, street view images can directly reflect environmental characteristics and have been widely used in urban environment perception research (J. Chen et al. 2023; Y. Zhang et al. 2023; Z. Wang, Ito, and Biljecki 2024; Yang et al. 2024). For example, the livability, safety, and aesthetics of urban areas can be revealed by analyzing green spaces, architectural styles, and the distribution of public facilities in street view images (Lu and Chen 2024; Qi et al. 2024; H. Zhou et al. 2024). Therefore, it is very feasible to obtain residents' subjective perceptions of the rural living environment through rural street view data. The collection and updating of street view data, however, are mainly concentrated in urban areas, whereas the coverage of rural street view images is relatively low, making the application of relevant research extremely limited in rural environments (Y. Yuan et al. 2023). To address this issue, our study employs a crowdsourcing approach to collect rural street view data, thereby supporting subsequent quantitative analysis.

Furthermore, combining street view data with deep learning technologies has become a mainstream method for evaluating residents' subjective perceptions (Ji et al. 2021; Y. Zhang et al. 2021; Ogawa et al. 2024; Z. Wang, Ito, and Biljecki 2024). Kruse et al. (2021) used a deep residual network (ResNet) model trained on tagged street view data to evaluate human perceptions of urban playability. S. Chen et al. (2024) used street view image data sets and deep learning methods to assess the sense of safety in public spaces. These studies primarily concentrate on modeling and evaluating subjective perceptions within urban settings, however, and there remains a significant deficiency of deep learning models tailored for rural environments.

To address the aforementioned issues, this study proposes a method for evaluating subjective perceptions of the rural living environment by combining rural street view images with deep learning models, filling the gap in existing research regarding large-scale quantification of rural environment perceptions. To our knowledge, this is the first study to use large-scale rural street view data and deep learning technologies to evaluate rural environment perceptions across multiple cities nationwide. Meanwhile, this study also makes several theoretical contributions. First, we expanded the index system of rural perception, replacing the single satisfaction indicator in traditional studies with rich dimensions. Second, the study reveals significant differences in rural environment perceptions between different regions, finding that these differences are influenced by various factors such as economic foundations and traditional characteristics. Third, the study finds that the relationship between rural subjective perceptions and objective indicators is complex, with obvious spatial heterogeneity. Perceptions in the eastern regions are mainly influenced by economic levels, whereas the western regions reflect the roles of infrastructure construction and social development more. These findings provide important scientific references for the improvement of rural environments.

Related Works

Subjective Perception Evaluation of Rural Environment

Residents' subjective perceptions of the rural environment serve as a crucial basis for formulating public health intervention measures. For instance, areas with poor sanitation conditions can be identified by evaluating residents' satisfaction with environmental sanitation, thus providing critical data support for the precise formulation of relevant policies (Fang et al. 2022). Traditional evaluations of rural perception primarily depend on interviews and questionnaires. Y. Wang, Zhu, and Yu (2019) conducted field surveys and collected 370 questionnaires to examine the satisfaction of residents in Xianju County, Zhejiang, regarding the habitability of rural areas. Moor et al. (2022) employed both written and online questionnaires to assess the satisfaction of all households in four pilot villages in the Netherlands with their living environments. Pang et al. (2024) gathered 466 questionnaires from twenty-five villages in Beijing to study respondents' satisfaction with the habitability of rural areas there. Although these methods offer relatively comprehensive research insights, their applicability to nationwide assessments of subjective rural environment perceptions is limited due to the significant labor and time costs of conducting questionnaires and interviews, as well as the restricted sample coverage.

With the rapid development of multisource geospatial data, a large amount of geotagged data has become more accessible (Y. Liu et al. 2015). Among these, street view data can accurately simulate the visual elements of cities (L. Liu et al. 2016). These specific physical scenes can evoke aesthetic and emotional responses in people, thereby forming subjective perceptions of a location in the human brain, such as beautiful, safe, or depressing (Dubey et al. 2016). In recent years, rural street view images have been shown to effectively reflect human perception of traditional rural areas (Sjaf et al. 2022; Xing and Leng 2024). Because the street view data collected by map service providers such as Google, Tencent, and Baidu do not cover extensive rural areas, however, researchers often need to collect rural street view data on their own. This data collection is often limited to individual villages or towns, making it challenging to meet the demands of large-scale research. Therefore, the difficulty in data acquisition has become the main obstacle to applying rural street view images in large-scale evaluations of the subjective perception of rural environments.

The rapid development of volunteered geographic information (VGI) provides an effective solution to this problem. Individuals can capture rural scenes through photos and upload them to open map services, providing a reliable data source for assessing rural perception (Goodchild 2007). The rural street view data obtained through crowdsourcing effectively addresses many of the limitations associated with traditional data collection methods, enabling quantitative evaluations of the subjective perception of the rural environment on a larger scale (Krupowicz, Czarnecka, and Grus 2020).

Evaluation System of Rural Environment Perception

In recent years, researchers have increasingly used street view data to study residents' perceptions of urban areas (Min et al. 2020; Yao et al. 2021; Qiu et al. 2023; Z. Wang, Ito, and Biljecki 2024). This type of research has developed a comprehensive evaluation system, particularly focusing on the dimension of perception. For example, Salesses et al. (2013) explored questions such as "Which place seems safer?" and "Which place seems wealthier?" by comparing street view images of New York, Boston, Linz, and Salzburg. Dubey et al. (2016) used the Place Pulse 2.0 data set to quantify global urban perceptions across six dimensions: safety, lively, boring, wealthy, depressing, and beautiful. These six dimensions have become significant reference points for numerous subsequent studies on urban perception (F. Zhang et al. 2018; Ji et al. 2021; Zhao, Lu, and Wang 2024). Moreover, Ogawa et al. (2024) further broadened the research by evaluating up to twenty-two subjective perception items in the Setagaya district of Tokyo.

Due to the significant differences between urban and rural landscapes, however, the perception dimensions captured by urban street views do not fully apply to rural areas (Yusheng and Ntarmah 2021). To address this gap, this study constructs an evaluation system based on the characteristics of rural environments, incorporating five dimensions: wealthy, lively, habitable, tidy, and terroir. The first three indicators (wealthy, habitable, and tidy) are derived from the explicit requirements put forward in the policy documents on rural construction by the Chinese government, namely building a rural environment with "wealthy life," "habitable ecology," and "tidy village appearance" (General Office of the State Council of the Communist Party of China 2021). "Wealthy" is the core dimension for measuring economic levels, reflecting the psychological feelings of residents toward the economic development status of rural areas (Lv, Xie, and Li 2019). "Habitable" focuses on whether the rural environment is comfortable, natural, and suitable for living, including considerations of the completeness of infrastructure, comfort of the living environment, and environmental affinity (Y. Wang, Zhu, and Yu 2019; X. Li et al. 2021). "Tidy" reflects the level of environmental sanitation and public space management in villages, and its quality directly affects the health and happiness of residents, serving as an important indicator for measuring the quality of the rural environment (Fang et al. 2022). Therefore, the wealthy, tidy, and habitable dimensions are undoubtedly important for assessing residents' subjective feelings toward the rural living environment and should be included in the evaluation system of rural environment perception.

Compared with the first three criteria, the lively and terroir dimensions are also crucial in evaluating rural environment perception. "Lively" captures the vibrancy and vitality of rural life, serving as a key dimension for evaluating social interactions within villages, population mobility, and residents' engagement in community activities (M. Yuan et al. 2024). A lively village is typically characterized by frequent daily activities, high utilization of public spaces, and proactive community involvement from residents (Zheng, He, and Tang 2023). The term *terroir* originates from French, initially used to describe the unique regional flavors of agricultural products such as wine. Its use has gradually expanded to other fields to denote local characteristics or regional features. Terroir focuses on the village's natural landscape, cultural traditions, and local characteristics, which are core features distinguishing rural from urban areas (M. Zhou, Chu, and Du 2019; F. Wang and Prominski 2020). Overall, the indicators selected in this study comprehensively reflect the actual conditions of the rural environment. They also provide important references for the scientific evaluation of rural environment perception. Notably, although *livable* is commonly used in urban perception studies, we adopt the term *habitable* to gauge the suitability of the environment for living. This choice aims to avoid potential confusion for readers due to the visual and semantic similarity between lively and livable.

Deep Learning Techniques in Existing Street View Research

With the rapid development of high-performance computing systems, current deep learning technology can extract higher dimensional information from images and has gradually become a mainstream method in the study of human subjective perception. For example, Dubey et al. (2016) collected volunteers' perceptions to construct a street view perception data set and developed a Siamese-structured convolutional neural network (CNN) to evaluate perceptions on a global scale. F. Zhang et al. (2018), based on the MIT Place Pulse data set, segmented street view images by landscape elements and proposed a deep convolutional neural network (DCNN) model to quantify the impact of physical scenes on perception dimensions. Rossetti et al. (2019) used Place Pulse 2.0 and applied discrete choice models and subdivision models to develop six binary models of subjective items. Yao et al. (2021) developed a scoring framework based on human–machine confrontation to evaluate the perception scores of street views.

Studies indicate that deep learning technology has significant effectiveness in urban perception research. In the rural domain, though, there is still a lack of deep learning models for the systematic evaluation of rural environment perception. Due to the significant differences in physical characteristics and residents' lifestyles between rural and urban environments, existing urban perception models are difficult to directly apply to the evaluation of rural settings. To address this issue, we develop a deep learning model for the subjective perception of rural environments based on previous studies and using the method of paired image comparison (Salesses, Schechtner, and Hidalgo 2013). This study aims to provide a more accurate, effective, and intuitive way to systematically evaluate the subjective perception of the rural environment.

Methodology

The technical process of this study is shown in Figure 1, primarily including three steps: (1) Collecting 23,287 rural street view images, constructing an evaluation index system after rigorous data processing, and evaluating the images by volunteers, with the scoring results stored in the rural street view database; (2) using 78,613 sets of comparison data in the database, training the evaluation models for various indicators through data augmentation and ResNet feature extraction, and scoring the comparison results with the TrueSkill model to generate the index score ranking chart; and (3) combining street view data and volunteer scoring data, calculating the rural landscape score through the TrueSkill model, followed by model accuracy analysis, national rural landscape score distribution pattern analysis, perceptual dimension correlation analysis, and the correlation analysis between subjective perception and objective indicators.

Construction of Rural Street View Data Set

To construct a comprehensive rural street view data set, this study organized a volunteer team to collect data across the country. In detail, we took prefecture-level cities in twenty-eight provinces of China as basic units and selected representative rural areas within each unit. The collection routes were planned according to the distribution of rural road networks to ensure that the diverse village landscape environment within each sampling unit was covered. Volunteers employed two methods for data collection: first, capturing static rural street view images using mobile phone cameras, and second, recording dynamic video data with the Trace Video 2.0 software. The integration of these two approaches enhanced the efficiency of data collection and ensured multiperspective coverage of rural environments.

In each village, volunteers took photos based on specified categories such as farmhouses, roads, and public facilities to ensure the data's comprehensiveness and representativeness. The principle was to reflect the true appearance of the countryside as accurately as possible, without considering the quality of the rural street view. This guideline ensured that the collected data genuinely represented the current state of the countryside and avoided the bias introduced by selective shooting, thereby providing a more objective and comprehensive basis for rural environment assessment.

To ensure data quality, the study conducted strict screening and processing of the collected rural street view images and videos. For video data, we extracted key frame images through frame rate control and conducted preliminary manual screening to remove invalid frames with excessive similarity or inverted images, reducing redundant data while ensuring image quality. Subsequently, a further fine-grained screening was conducted on the initially selected key frames and static captured images, eliminating invalid images that are blurry, incomplete, and lacking in information (e.g., those containing only the sky or tree). The high-quality street view images obtained after screening were stored in a database, forming the final rural street view data set.

Rural Street View Scoring Model Based on Human–Machine Collaboration

The rural street view evaluation method based on human–machine collaboration involves two main processes: comparison and scoring. We designed a deep learning network, termed PAIR-CNN, to facilitate large-scale comparisons of rural street view data, and employed the TrueSkill algorithm to compute the score of each rural street view. We also established a Web platform (http://computable-rural.urbancomp.net) to gather and record comparison data of rural street views while training the deep learning model in the background. Once the model stabilizes and achieves the required accuracy, volunteer scoring is discontinued. This human–machine collaboration approach significantly improves the decision-making process for comparisons, reduces

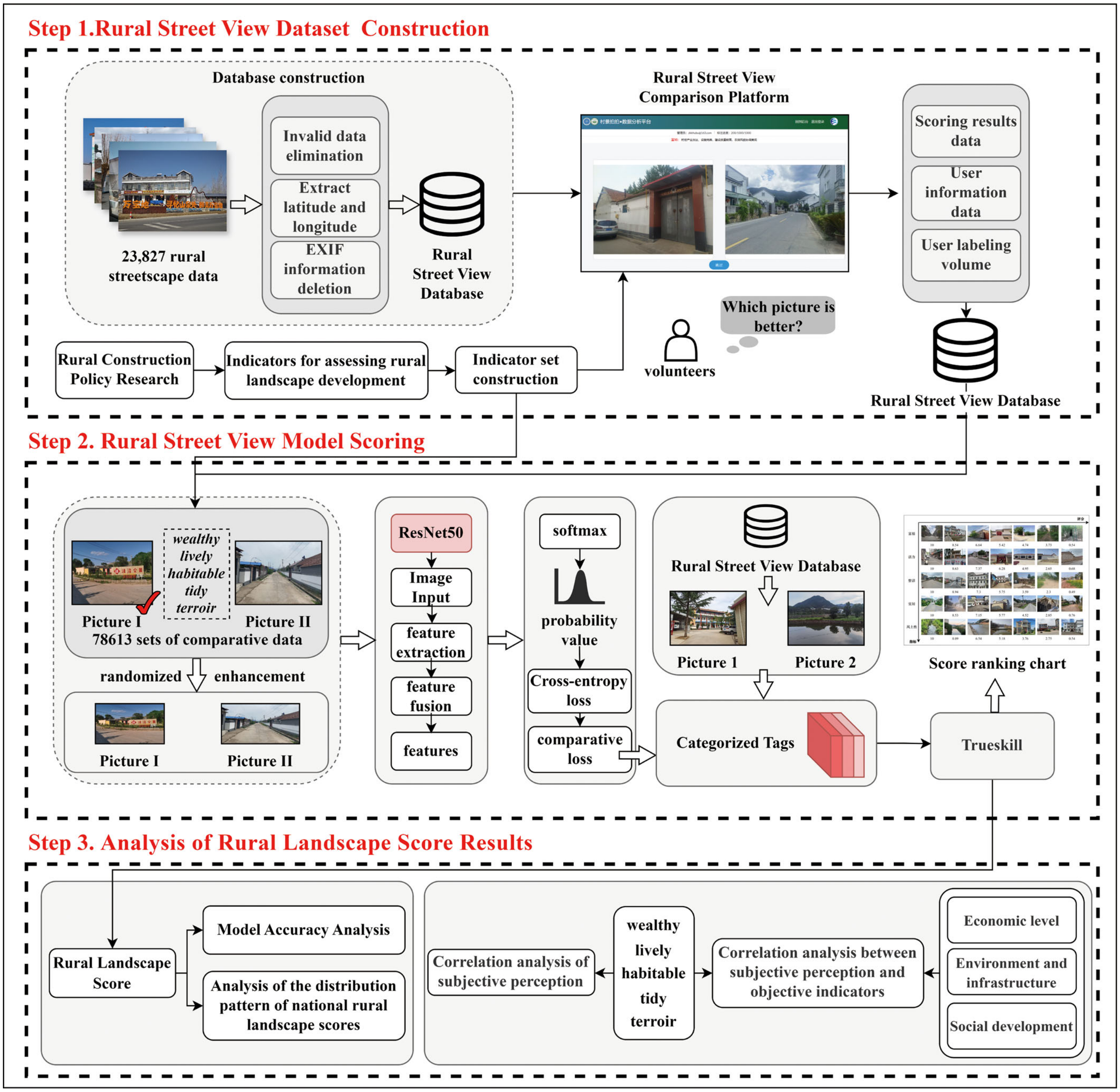

Figure 1. Technical flowchart of the study.

labor costs, and mitigates the errors caused by subjective biases. Consequently, it offers a convenient, accurate, and efficient evaluation method for large-scale rural street view perception. In the following section, we provide a detailed description of the design, training, and extensive application of the comparison model, and further elaborate on how comparison data are used to calculate the score of each rural street view.

Rural Street View Comparison Model Based on PAIR-CNN Network. PAIR-CNN predicts the winner in pairwise comparison tasks by inputting a pair of images (Figure 2). The network's design is inspired by the Siamese network (Chopra, Hadsell, and LeCun 2005), primarily comprising two subnetworks: feature extraction and feature fusion. In the feature extraction stage, the two images share a network with identical weights. We use ResNet50

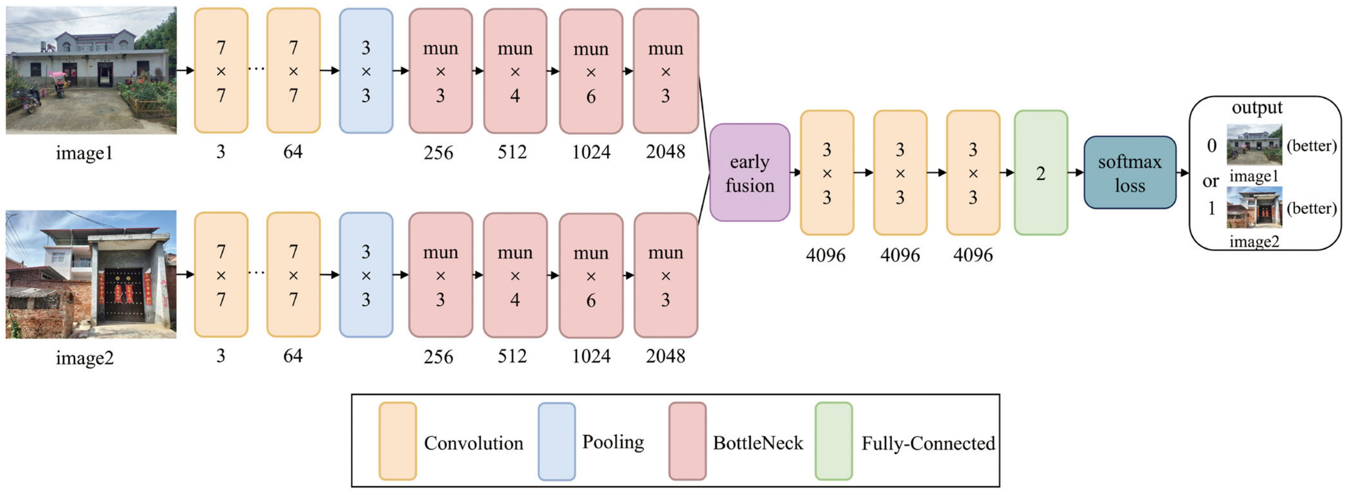

Figure 2. Schematic diagram of PAIR-CNN network structure.

pretrained on Places2 (http://places2.csail.mit) for this process. ResNet50 is a deep learning model based on residual networks, which effectively mitigates the vanishing gradient problem by introducing residual connections, thus exhibiting outstanding performance in large-scale image classification tasks (F. Zhang et al. 2018). Places2 is a large-scale image database containing 10 million images and encompassing 434 types of scenes (B. Zhou et al. 2018). By pretraining on Places2, the model's generalization capability will be significantly improved, and it can capture various scene features, thereby more effectively addressing the challenges of scene diversity and complexity in rural street view images.

In the feature fusion stage, an early fusion method is employed, concatenating the two features along the vector dimension to capture the correlated features between the two images (Barnum, Talukder, and Yue 2020). Subsequently, a set of convolutional layers and fully connected layers output softmax loss to achieve a binary classification task. If the output is 0, the perceived score of the first image is higher than that of the second image; if the output is 1, the opposite is true. Therefore, the cross-entropy loss function is used to calculate the softmax:

$$L = \sum_{(i,j,y)\in P} \sum_{k}^{K} -1|y = k|\log\left(g_k(x_i, x_j)\right) \tag{1}$$

where (i, j, y) represents the sample, y is the true label (with a value of 0 or 1), $K=2$, and g is the softmax of the final activation layer.

It is worth noting that the collection of rural crowdsourced street view data predominantly relies on smartphone photography. The parameters of the devices are difficult to standardize, however, resulting in issues such as low light or distortion in the images. To mitigate the impact of these issues on model performance, it is necessary to enhance the saturation of village scene images before training and use GridMask to improve the robustness of the network learning (P. Chen et al. 2020). Finally, we use the Accuracy score to measure the performance of the model:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

where $Accuracy$ denotes the accuracy of the model, TP is the number of correctly predicted positive samples, TN is the number of correctly predicted negative samples, FP is the number of incorrectly predicted positive samples, and FN is the number of incorrectly predicted negative samples.

During the optimization stage, exponential learning rate (ExponentialLR) is used to adjust the learning rate to enable the model to converge quickly. The initial learning rates for the five models—wealthy, tidy, lively, habitable, and terroir—are all set to 0.085, with momentum set to 0.90, and decay factors set to 0.50, 0.25, 0.25, 0.20, and 0.65, respectively. On validation, the model performs best under this parameter setting (Tables S.1–S.5).

Rural Street View Scoring Model Based on TrueSkill. TrueSkill is a Bayesian theory-based multiplayer ranking algorithm (Graepel, Minka, and Herbrich 2007). In this study, two images in a sample set are regarded as two contestants. Initially, each image has the same score. Then, the scores of the two images are adjusted simultaneously according to the competition results, with the winning image scoring higher. When the number of comparisons for each image reaches twenty-four to thirty-six times, a stable ranking score can be obtained (Dubey et al. 2016; Gong et al. 2023; Qiu et al. 2023).

Given the large scale of data in this study, we randomly selected forty images to pair with each rural street view image, ensuring no duplicates in the comparison records. This setting ensures the stability and convergence of the ranking scores and avoids score fluctuations that might be caused by insufficient comparison times, enhancing the generalization ability of the model's comparison results. Ultimately, a $20{,}000 \times 40$ comparison data set was generated and input into the PAIR-CNN model to obtain comparison results of the large-scale rural street view data.

This study uses the TrueSkill algorithm to analyze the large-scale rural street view comparison data. All images are initially assigned a score of zero and given a Gaussian distribution with a mean variance of 0.0833 and a draw probability of 0.0001. This configuration accommodates diverse image comparison scenarios. After all comparison results have converged, to make score differences more significant and facilitate subsequent analysis, we normalize the initial scores to a range of zero to ten as follows:

$$S = 10 \times \frac{Score - Score_{\min}}{Score_{\max} - Score_{\min}} \tag{3}$$

where $Score_{\min}$ is the lowest score in this perception dimension, $Score_{\max}$ is the highest score in this perception dimension, and *Score* is the original score.

Analysis of Rural Landscape Score Results

This study uses a correlation matrix to show the correlation and significance levels between each indicator. We use SPSS's correlation analysis method to analyze the scores of 20,000 street view images based on the PAIR-CNN network and TrueSkill model in five indicators: wealthy, lively, habitable, tidy, and terroir. Then, we calculated the Pearson correlation coefficients between the five indicators and used a two-tailed test for significance testing.

To further explore the relationship between the subjective perception of the rural environment and objective socioeconomic characteristics, this study investigated existing literature and selected twelve representative variables from three categories: economic level, environment and infrastructure, and social development (Hu and Wang 2020; Guo et al. 2023; H. Zhang, Liu, and Guo 2024). These variables comprehensively cover key areas affecting the rural living environment. Then, the scores of all street view images in each city on five subjective perception dimensions were calculated, and the average of these scores was used as the composite index of the corresponding dimension for that city. Finally, we calculated the rank correlation coefficients between the five subjective perception dimensions and various objective indicators.

The economic level indicators include regional gross domestic product (GDP; X11), per-capita GDP (X12), fiscal revenue (X13), and farmers' per-capita disposable income (X14). These indicators directly reflect the level of regional economic development and are important factors in measuring the wealth of villages and the quality of life of residents (Brambert and Kiniorska 2018). The environment and infrastructure indicators include the prevalence rate of public toilets (X21), rural greening rate (X22), road hardening rate (X23), and the proportion of cultivated land and water conservancy facilities used land (X24). These indicators directly affect the "hard environment" of rural residents' daily lives (Sharp et al. 2002). Among them, the proportion of cultivated land and water conservancy facilities used land (X24) measures the proportion of traditional agricultural landscapes such as cultivated land and canals in the land-use structure, and these landscapes are an important part of the rural environment. Social development indicators include the urban–rural income ratio (X31), the proportion of rural residents' cultural and entertainment expenditure (X32), the number of rural cultural stations (X33), and the number of traditional villages (X34). These indicators reflect the distribution of urban–rural income differences and social and cultural resources (Guo et al. 2023). Among them, the number of rural cultural stations (X33) refers to the number of cultural activity venues in rural areas, reflecting the accessibility and richness of cultural facilities. Refer to the Supplemental Material for a detailed description of the data sources for objective indicators.

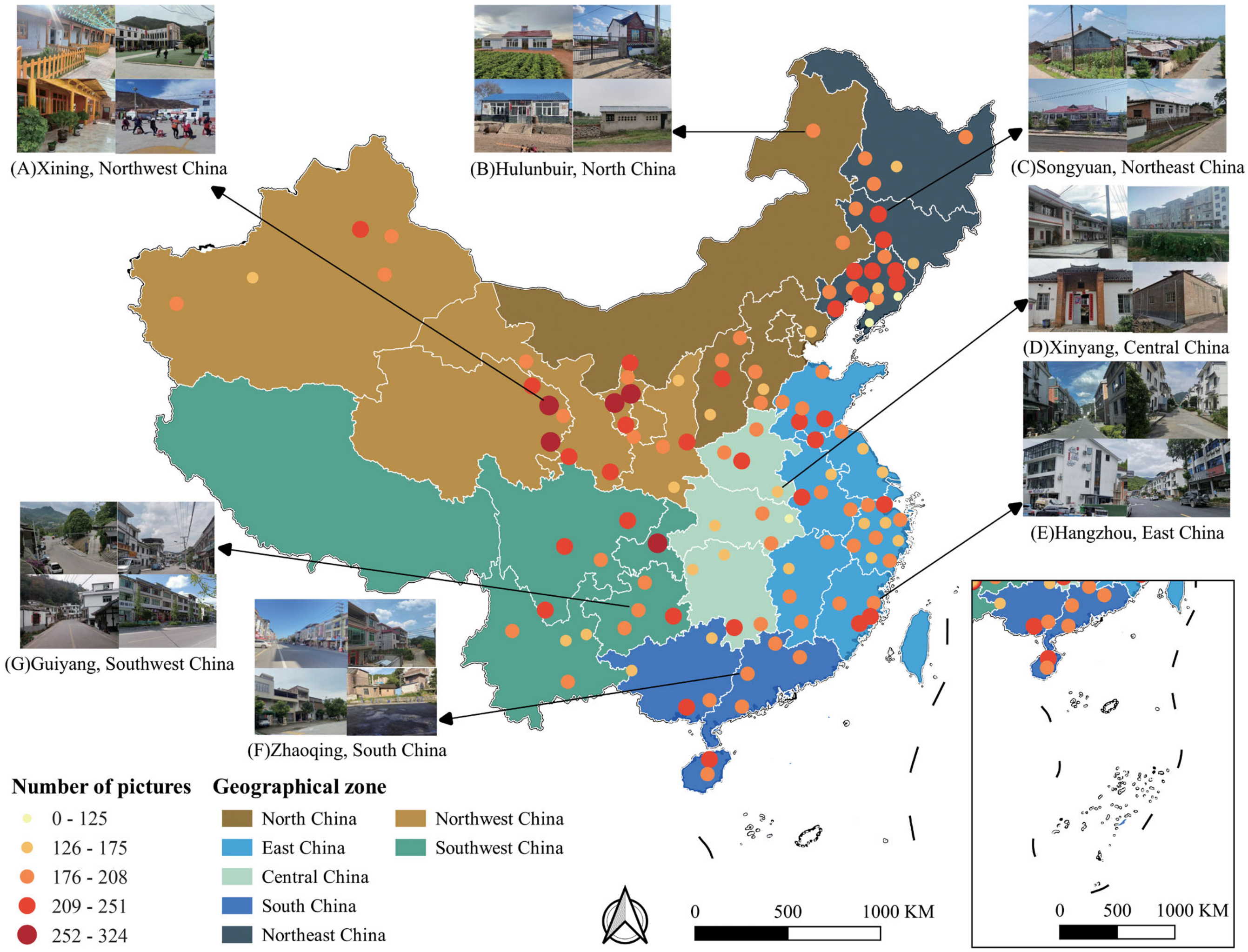

Figure 3. Spatial distribution of collected rural street view images across seven geographical zones in China, where (A) to (G) show the typical village scenes of the representative cities in each geographical zone.

Results

Data and Volunteer Scoring Results

This study enlisted volunteers to collect street view images of rural areas across the nation, and the data collection time was uniformly arranged from April to October 2023. This time period covered three seasons of spring, summer, and autumn, which could fully display the diversity and seasonal characteristics of rural scenery in different regions, and also effectively avoid the possible impacts of climate anomalies or long-term changes between different years on data consistency (Biljecki and Ito 2021; Qiu et al. 2023). Ultimately, 23,827 rural street view images were gathered, covering twenty-eight provinces, 118 prefecture-level cities, and 121 county-level regions, forming a large-scale rural street view data set. On average, each county contains 197 images, ensuring the diversity and representativeness of the data.

The spatial distribution of the data is illustrated in Figure 3, which shows the sampling quantity of prefecture-level cities nationwide against the backdrop of seven geographical divisions. In addition, Figure 3 showcases the typical features of rural street views in various regions, including the modernized villages of the southeast (e.g., Hangzhou), pastoral landscapes of the south (e.g., Zhaoqing), farmyards of the north (e.g., Hulunbuir), traditional village dwellings in the northwest (e.g., Xining) and northeast (e.g., Songyuan), as well as regionally distinct rural street views in the southwest (e.g., Guiyang). The results indicate that the rural street view data

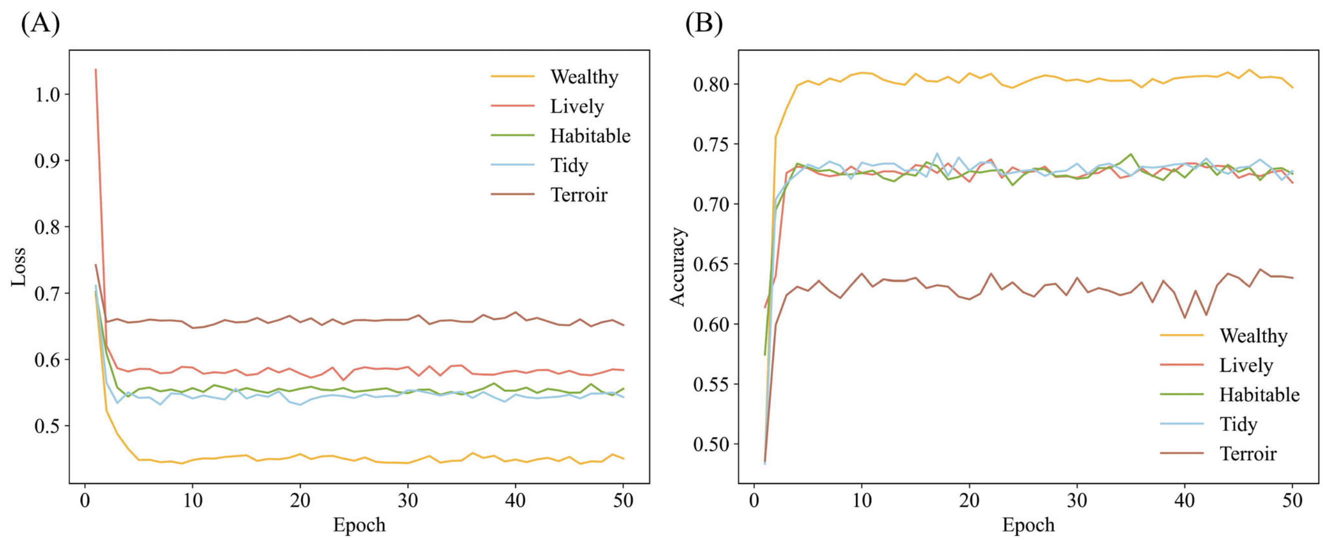

Figure 4. Loss and precision variations during the model's training process. (A) The loss curves during fifty epochs of model training under different metrics. (B) The accuracy curves during fifty epochs of model training under different metrics.

collected in this study cover a wide range of distribution and reflect the diversity of Chinese rural landscapes in terms of architectural style, natural environment, and cultural characteristics, providing a solid data foundation for subsequent research.

Although the data collected in this study cover a wide range, there is still some imbalance in the collection quantity between various regions. To reduce the potential impacts of this difference on the results, we randomly selected 3,500 rural street view images from the entire rural street view data set, with 500 images from each geographical division. This sampling strategy effectively balanced the number of images in each region, minimizing the bias that might be brought by the difference in collection quantity (Howell et al. 2020; Hou et al. 2024). Subsequently, we conducted a manual quality check on the selected images to ensure that all images could clearly and truly reflect the characteristics of rural landscapes in each region, laying a reliable data foundation for subsequent scoring and analysis.

We invited eighty volunteers to score these images using a human–machine collaboration scoring system to gather subjective perception data. The volunteers, between twenty and forty years old, were all students and faculty members specializing in geographic science, human geography, and urban–rural planning. They possessed professional knowledge in geographic information systems (GIS) applications and urban–rural planning theory, effectively mitigating perception assessment bias (Yao et al. 2019). Each volunteer scored 1,000 pairs of images. In total, 78,613 valid ratings were collected, providing ample high-quality labeled samples for subsequent model training. Each image was compared an average of forty-one times, closely aligning with the comparison frequency in Yao et al. (2021), ensuring the adequacy of the scoring.

It should be noted that although the experience of local villagers has important reference value for the optimization of specific villages, the subjective perception evaluation on a national scale must prioritize ensuring the cross-regional consistency of evaluation standards. Therefore, this study selected a group of volunteers with homogeneous professional backgrounds to avoid the impact of the limitations of villagers' perspectives or regional cognitive biases on the comparability of results (S. Liu et al. 2022). This method has been verified to be feasible in macro perception studies (Lu and Chen 2024).

Rural Street View Image Scoring Results Based on Human–Machine Collaboration

The study divided the 78,613 volunteer scoring results data into a training set and a test set at a ratio of 7:3 (Z. Wang, Ito, and Biljecki 2024; Yu et al. 2024). After fifteen rounds of training, the accuracy and loss of the model both tended to stabilize. Figure 4 illustrates the model's training process

Figure 5. Examples of typical rural street view images under each indicator. The indicator axis represents the five perceptual dimensions, and the score axis represents the scores of the rural street view images.

across the five perception dimensions. The results indicate that the accuracy for the wealthy dimension reached 80 percent, and the lively, habitable, and tidy dimensions all achieved accuracies above 70 percent. These results are comparable to the accuracy of most existing street view studies (e.g., 71.8 percent in Z. Li et al. [2022] and 72 percent in M. Sun et al. [2022]), demonstrating that our model is accurate and reliable. The accuracy for the terroir dimension was the lowest (64.6 percent), but this result is consistent with the range of [0.60, 0.70] reported by Min et al. (2020), suggesting that the model has a certain level of usability in this dimension as well. The poorer performance in the terroir dimension compared to the other four dimensions reflects the significant variation in people's perception of unique rural features. This variability could arise from the complex nature of the terroir concept, which involves numerous factors that are difficult to quantify, making it challenging for the model to predict and evaluate accurately (Shucksmith 2018).

We applied the model to a data set of 20,000 rural street view images to obtain large-scale comparison results. Using the TrueSkill algorithm, we calculated the scores of each image across five perceptual dimensions. To visually demonstrate the reliability of the model's evaluation results, Figure 5 presents typical examples with different scores under the five indicators. These images reflect the score variations across each dimension. For instance, in the wealthy dimension, high-scoring images (above eight points) display tidy and modern buildings and infrastructure, whereas low-scoring ones (below two points) show dilapidated houses and rudimentary environments. In the lively dimension, high-scoring images (above eight points) depict bustling streets and vibrant community activities, whereas low-scoring ones (below two points) reflect desolation and lack of vitality.

Figure 6(A) shows the score distribution across the five perception dimensions, with all dimensions exhibiting a clear normal distribution trend. The peak values are predominantly concentrated between

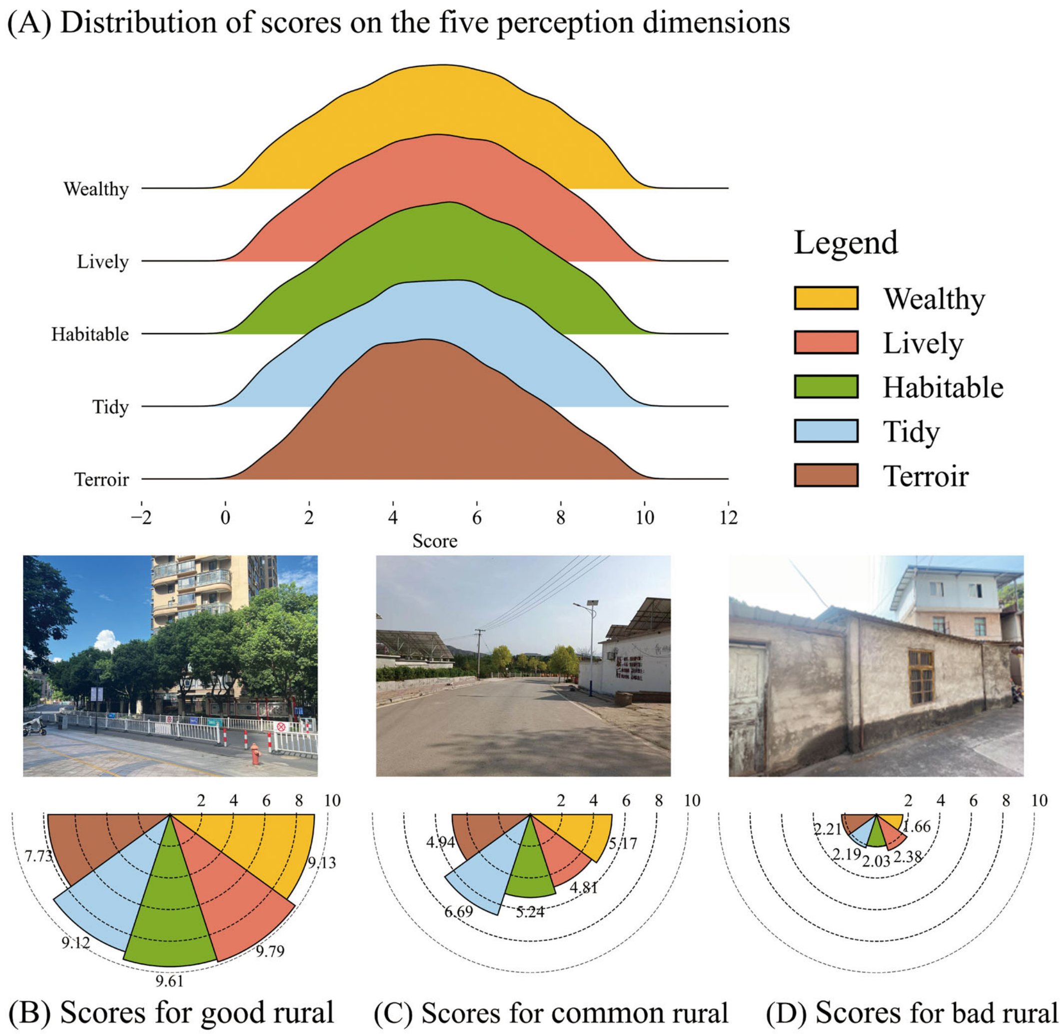

Figure 6. Score distribution and typical areas of rural street views, where (A) shows the distribution of scores on the five perception dimensions, and (B), (C), and (D) present representative images with high, medium, and low average scores, respectively, along with the scores for each dimension.

six and eight points, reflecting high consistency in volunteers' evaluations of rural street views. Notably, the score distribution in the terroir dimension is relatively concentrated, mainly ranging from four to eight points. This pattern indicates that most villages perform relatively well in maintaining local characteristics. To further demonstrate the reliability of the scores, we selected three representative rural street view images from the sample, each corresponding to high (above seven points), medium (four to six points), and low (below three points) average scores, respectively (Figure 6B–D). These scores were found to align with subjective perceptions. High-quality (above seven points) rural environments are clean and tidy, with comprehensive service facilities and no signs of excessive urbanization (characterized by a high green-looking ratio and low-density residential areas). Medium-scoring (four to six points) rural environments perform relatively average, with infrastructure and environmental quality at a general level. In contrast, poor (below three points) rural environments appear dilapidated and desolate, with a strong sense of oppression.

Rural Landscape Score Results

To more intuitively illustrate the spatial heterogeneity of the five evaluation indicators across China, we divided the country into seven geographical regions and calculated their respective scores for each indicator, as shown in Figure 7. The results reveal that the east China region outperformed other

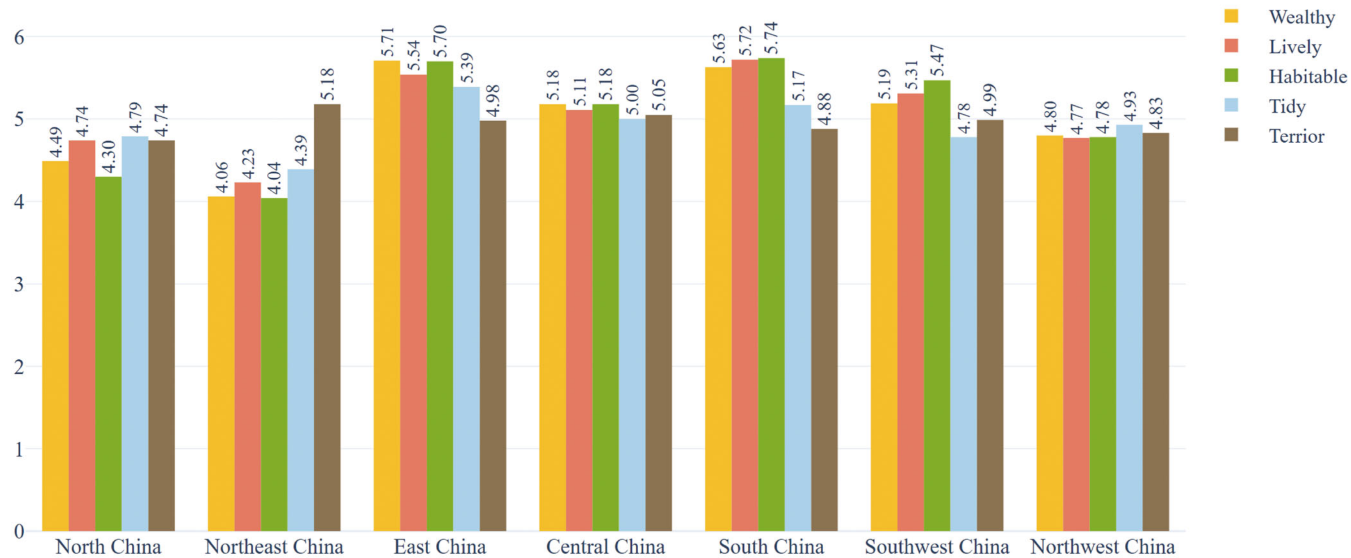

Figure 7. Rural street view scores in seven geographical regions across five evaluation metrics.

regions overall, whereas the northeast region exhibited the lowest overall scores. Specifically, east China scored notably higher in the wealthy (5.71), lively (5.54), habitable (5.70), and tidy (5.39) dimensions, reflecting its significant advantages in economic development, quality of life, and environmental sanitation. In contrast, the northeast region scored the lowest in wealthy (4.06), lively (4.23), and habitable (4.04) dimensions, indicating room for improvement in these areas. The northeast region, however, excelled in the terroir dimension with a score of 5.18, demonstrating its success in preserving local characteristics in rural development. Additionally, although the northwest region's overall scores were relatively low, its performance across each indicator was balanced with minimal fluctuation: wealthy (4.80), lively (4.77), habitable (4.78), tidy (4.93), and terroir (4.83). This suggests that although the overall development level of the rural environment in the northwest is limited, its performance across various aspects is stable without significant shortcomings.

To further investigate the performance differences among specific cities within various regions, we analyzed the score distribution of prefecture-level cities across the country (Figure 8). Overall, cities in the eastern region outperform those in other regions across multiple dimensions. For instance, Zhejiang Province leads in scores for the four indicators: wealthy (6.47), lively (6.17), habitable (6.50), and tidy (5.78). These scores showcase Zhejiang's significant advantages in economic development, quality of life, and environmental hygiene.

In contrast, northern regions such as Jilin have relatively lower scores in wealthy (3.94), lively (4.02), habitable (3.91), and tidy (4.22). The relatively low scores suggest that these areas still have considerable room for improvement in economic development and living environment enhancement. In the terroir indicator, however, the northern regions excel, particularly Heilongjiang, which scores the highest in terroir (5.37), indicating that it has better preserved local characteristics. These interregional differences highlight the uneven development of China's rural environment across different regions.

Correlation Results Between Subjective Perception Dimensions

We conducted Pearson correlation analysis on the five perceptual dimensions, and Table 1 shows the Pearson correlation coefficients between them. The results indicate that all Pearson correlation coefficients among the dimensions achieve statistical significance ($p < 0.01$), suggesting a significant interrelationship exists between each dimension. There is a strong positive correlation between wealthy and lively (0.84), habitable (0.86), and tidy (0.74). Lively is significantly positively correlated with

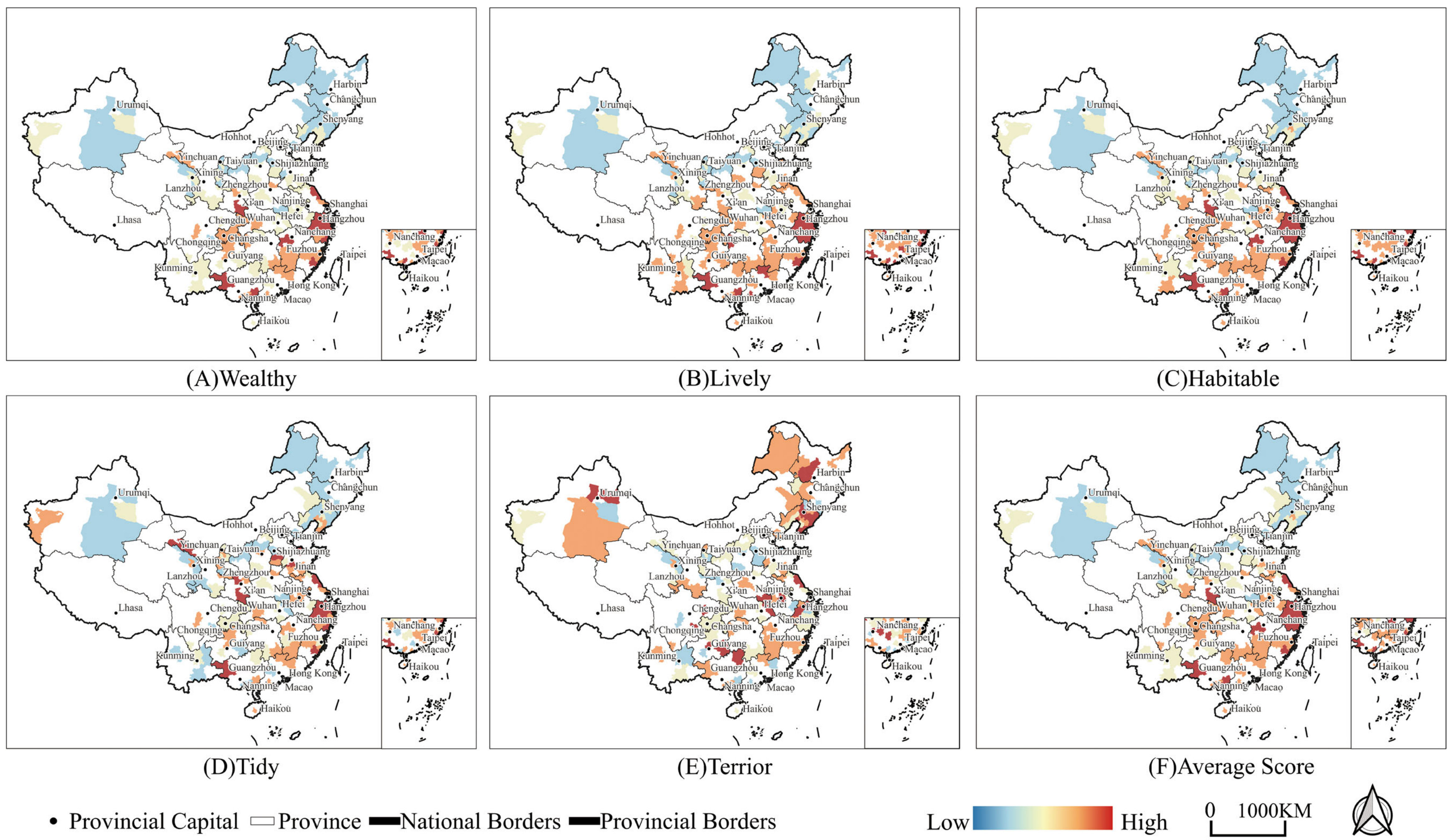

Figure 8. The distribution of rural street view scores at the prefecture level nationwide. (A) through (E) show the score distributions for the five perceptual dimensions, and (F) shows the distribution of the average scores across the five dimensions.

habitable (0.80) and tidy (0.70). Habitable is significantly positively correlated with tidy, with a correlation coefficient of 0.68. Terroir, however, is significantly negatively correlated with wealthy (–0.31), lively (–0.25), habitable (–0.14), and tidy (–0.24). The negative correlation is relatively weak, making it a relatively independent indicator.

Results of Correlation Between Subjective Perception and Objective Indicators

There is obvious heterogeneity in economic levels, infrastructure construction, social development, and so on, in different regions of China, which could affect the subjective perception of the rural environment and its association with objective indicators (Guo et al. 2023). This study investigated the relationships between five dimensions of subjective perception and three types of objective indicators across national, eastern, and western scopes through Spearman correlation analysis (Table 2). On a national scale, economic level indicators are significantly positively correlated with wealthy, habitable, and lively, confirming the important role of economic development in rural environmental perception (Hu and Wang 2020). Most of the environment and infrastructure indicators are positively correlated with subjective perception, but X24, the proportion of cultivated land and water conservancy facility land, is significantly negatively correlated with wealthy (–0.401), habitable (–0.483), and lively (–0.439), reflecting the lack of convenience and economic vitality in agriculture-dominated areas. Among the social development indicators, X34, the number of traditional villages, exhibits the strongest correlation with terroir (0.282), indicating that traditional villages are key factors in shaping the perception of local characteristics.

In the eastern region, economic level indicators exert the most significant impact on perception. Among them, X14, farmers' per-capita disposable income, is highly correlated with wealthy (0.673), lively (0.673), and tidy (0.508), indicating that economic characteristics dominate the influence on rural perception in the eastern region. In contrast, the environment and infrastructure indicators, and social development indicators mostly exhibit a

Table 1. Pearson correlation coefficient between each perceptual dimension

Perceptions	Wealthy	Lively	Habitable	Tidy	Terroir
Wealthy	1.00	0.84**	0.86**	0.74**	–0.31**
Lively	0.84**	1.00	0.80**	0.70**	–0.25**
Habitable	0.86**	0.80**	1.00	0.68**	–0.14**
Tidy	0.74**	0.70**	0.68**	1.00	–0.24**
Terroir	–0.31**	–0.25**	–0.14**	–0.24**	1.00

*$p < 0.05$.
**$p < 0.01$.

Table 2. Spearman correlation coefficient between subjective perception and objective indicators

		Economic level				Environment and Infrastructure				Social development			
Region	Perception	X11	X12	X13	X14	X21	X22	X23	X24	X31	X32	X33	X34
Nationwide	Wealthy	0.365**	0.341**	0.350**	0.521**	0.209*	0.209*	0.228*	–0.401**	–0.245**	0.210*	0.216*	0.246**
	Lively	0.285**	0.242**	0.295**	0.431**	0.200*	0.177	0.225*	–0.483**	–0.225*	0.206**	0.214**	0.326**
	Habitable	0.369**	0.329**	0.358**	0.488**	0.215*	0.200*	0.236*	–0.439**	–0.243**	0.222*	0.224*	0.307**
	Tidy	0.148	0.274**	0.169	0.457**	0.220**	0.251**	0.234**	–0.165	–0.257**	0.220*	0.220*	0.020
	Terroir	0.492	0.101	0.462	0.061	0.142	0.105	0.138	0.128	–0.146	0.150	0.141	0.282**
Eastern	Wealthy	0.319**	0.411**	0.500**	0.673**	0.313	0.137	0.191	–0.524**	0.259	0.374*	0.237	0.196
	Lively	0.155**	0.232**	0.360*	0.462**	0.266	0.136	0.195	–0.660**	0.227	0.354	0.184	0.279
	Habitable	0.333**	0.413**	0.504**	0.673**	0.283	0.139	0.156	–0.588**	0.236	0.310	0.193	0.240
	Tidy	0.193*	0.266**	0.451**	0.508**	0.321	0.069	0.263	–0.147	0.255	0.413*	0.309	0.025
	Terroir	0.425	0.348	0.460**	0.377*	0.234	0.313	0.123	–0.015	0.207	0.005	0.149	0.055
Western	Wealthy	0.106	0.299	0.065	–0.026	0.551**	0.568**	0.577**	–0.158	–0.555**	0.570**	0.555**	0.023
	Lively	0.040	0.335*	0.020	0.049	0.477**	0.482**	0.517**	–0.279	–0.476**	0.519**	0.493**	0.142
	Habitable	0.164	0.236	0.136	0.017	0.558**	0.568**	0.588**	–0.154	–0.563**	0.590**	0.565**	0.172
	Tidy	–0.359*	0.249	–0.378*	–0.060	0.351*	0.382*	0.374*	–0.012	–0.354*	0.360*	0.348*	–0.218
	Terroir	0.827**	0.193	0.800**	0.159	0.169	0.184	0.217	0.056	–0.179	0.198	0.180	0.202

*$p < 0.05$.
**$p < 0.01$.

weaker correlation with subjective perception. This could be attributed to the relatively well-developed infrastructure and balanced social development in the eastern region, which diminishes the marginal effects on subjective perception.

In the west, the correlation between economic level indicators and subjective perception is generally weak. Only X11, regional GDP, and X13, fiscal revenue, show a certain correlation with tidy and terroir. The environment and infrastructure indicators (e.g., X21, the prevalence rate of public toilets; X22, rural greening rate; and X23, road hardening rate) and social development indicators (e.g., X32, the proportion of rural residents' cultural and entertainment expenditure, and X33, the number of rural cultural stations) are significantly positively correlated with wealthy, lively, habitable, and tidy, however. This indicates that the improvement of infrastructure and investment in rural culture and entertainment can effectively enhance rural perception in the western region.

In addition, the impact of urban–rural income gap (X31) on subjective perception shows regional differences. At the national and western levels, X31 is significantly negatively correlated with most perception dimensions, indicating that the widening urban–rural income gap has a significant negative impact on the subjective perception of the rural environment (Ding et al. 2021; Huang et al. 2024). In the eastern region, however, X31 is weakly positively correlated with subjective perception and does not reach a significant level. This might be because the overall economic development level in the eastern region is relatively high, and the gradient difference in the urban–rural income gap among different cities is not significant, thereby weakening the negative impact of X31 on subjective perception.

Discussion

Scientific Contributions

With the implementation of the rural revitalization strategy, subjective perceptions of the rural environment have increasingly become a crucial dimension for assessing the level of rural environment construction. Challenges in data acquisition and the limitations of evaluative methods, however, have meant that existing research is predominantly confined to small-scale, specific area assessments. This limitation hinders the understanding of environmental perception characteristics on a regional scale. This study introduces a method for evaluating subjective perceptions of the rural living environment using rural street view images and deep learning models, addressing this research gap. The primary contribution of our study is the demonstration of the feasibility of using deep learning and street view images for large-scale rural perception assessments. This method provides a versatile tool for the rapid and precise quantification of subjective perceptions in rural environments.

This study uncovers the regional disparities in perceptions of the rural environment and identifies the multidimensional complexity of its underlying formation mechanisms. The east China region excels across dimensions such as wealthy, lively, habitable, and tidy, showcasing characteristics of systematic development. This advantage is related to the region's economic foundation and reflects the synergistic improvement of infrastructure, public services, and the living environment over long-term development (Fang et al. 2022). In contrast, the northeast region exhibits distinctive developmental characteristics: notable strength in the terroir dimension, yet comparatively weaker in other areas. This imbalance highlights the potential tension between the preservation of traditional features and modernization and indicates that improvement in a single dimension is insufficient for the overall enhancement of the rural environment. These findings overcome the limitations of traditional research that overly emphasizes a single factor, offering a more comprehensive explanatory framework for understanding regional differences in rural environment perception.

This study reveals the intricate relationship between subjective perceptions and objective indicators, highlighting significant regional heterogeneity in these associations. Such heterogeneity can be attributed to differences in regional development stages and resource endowments. In the east, after sustained development, a rural construction model driven by economic strength has evolved. Improvements in economic status are directly reflected in building quality, environmental management, and the systematic enhancement of infrastructure, thus translating more readily into subjective perceptions. Additionally, because the eastern region's infrastructure and social service systems are relatively well-developed, further marginal improvements have a diminishing influence on residents' perceptions, resulting in a relatively subdued impact of environmental and social development indicators. Conversely, the western region, characterized by generally lower economic levels and a lack of resources for comprehensive rural development (Li et al. 2021), tends to display "point-like" development features. Improvements in individual infrastructure and public services, such as greening rates, public toilet accessibility, and the presence of cultural facilities, can lead to noticeable enhancements in local environments (X. Sun et al. 2023). Such enhancements are more visibly reflected in the villages, significantly influencing how volunteers rate aspects like "habitable" and "tidy." This indicates that in areas with limited resources, advancing infrastructure and social services can effectively mitigate the constraints that insufficient economic development places on perception enhancement.

Practical Values

In practice, by quantifying the subjective perception of the rural environment in numerous cities nationwide, this study can provide precise rural environment perception maps and perception reports for local governments. These results can assist local governments in promptly identifying areas with poor perception, thus enabling more efficient allocation of limited resources and preventing "one-size-fits-all" policies and resource wastage.

Furthermore, this study introduces new perspectives and technical support for village classification. By evaluating the perceptions across various dimensions of village environments and integrating subjective perception characteristics into the classification system, planners can more precisely identify the distinct features of villages, providing a foundation for quick screening of different types of villages. For instance, based on the perception combination of high terroir–low wealthy, traditional villages with strong local characteristics but lagging economic development can be identified, thereby formulating more targeted planning schemes.

Additionally, local residents can gain insight into external evaluations of their village environments through perception score maps, thereby increasing their concern and sense of identity regarding village development. This, in turn, encourages more active participation in rural environment improvement, cultural preservation, and community governance, collectively driving the sustainable development of villages.

Limitations and Future Works

This study also has some limitations that need further improvement in future work. First, the existing model performs relatively poorly in the terroir dimension. In the future, consideration can be given to introducing multimodal data (e.g., cultural background information) and improving volunteer annotation methods to enhance the model's evaluation accuracy in the terroir dimension (Yao et al. 2019). In addition, the specific meaning and influencing factors of scores in each dimension are still insufficiently explained, so strengthening the interpretability analysis of the scoring results will be an important direction for future research.

Second, there are certain issues with the image quality and coverage of rural street view data. Because crowdsourced data relies on volunteers to capture images, discrepancies in the photographic skills and equipment quality of different volunteers might affect the consistency of image quality (Antequera et al. 2020; Biljecki and Ito 2021). Future research could improve the consistency of crowdsourced image quality by standardizing shooting protocols, providing equipment support, or conducting photography training. In addition, although we invited as many volunteers as possible to participate in capturing images, the number of participants is still limited, leading to differences in image coverage rates across regions. To address this, we call on major map service providers to further expand data coverage in rural areas, and emphasize the need to continue promoting the acquisition of more comprehensive rural street view data through crowdsourcing (Sánchez and Labib 2024). Furthermore, with the continuous accumulation of image data and volunteer evaluation data, future research can further carry out fine-grained perception studies on small-scale areas (e.g., villages or communities) to reveal more microscopic differences in rural environment perception characteristics.

Finally, subjective perception also has certain inherent limitations. The individual characteristics of volunteers, such as age, gender, and rural living experience, could influence their subjective perception of the rural environment, thereby introducing bias (Fang et al. 2022; S. Liu et al. 2022). Future studies should aim to expand the scale and diversity of volunteer groups, including recruiting rural residents from diverse regions, to achieve a more comprehensive and relevant understanding of the rural environment (Ito and Biljecki 2021; Lu and Chen 2024). Furthermore, this study's image collection period was concentrated between late spring and early autumn. Seasonal variations in rural street view characteristics, such as the growth condition of vegetation, might influence perception results (Chiang et al. 2023; Z. Wang, Ito, and Biljecki 2024;). Future efforts can involve adding timestamps to street view images and giving priority to scoring and analyzing images from the same season, thus effectively reducing bias introduced by seasonal characteristic differences (Qiu et al. 2023).

Conclusion

Evaluating the subjective perception of the rural living environment is crucial for optimizing rural revitalization strategies. This study introduces an innovative approach that integrates deep learning algorithms with rural street view images to quantitatively evaluate the subjective perception of the rural environment across 118 cities in China. The results show that our proposed method is highly reliable and practical in evaluating the subjective perception of rural environments.

Our study makes three primary contributions. First, we confirm the feasibility of using deep learning technology and street view images to evaluate large-scale subjective perceptions of the rural environment, offering new perspectives for future rural environment research in diverse countries and regions. Second, the study uncovers the characteristics of regional disparities in rural environment perception and their complex formation mechanisms. We find that rural environment perception is shaped by a comprehensive set of factors, such as economic foundation and protection of traditional features, with noticeable imbalances observed across regions. Third, the study identifies the complex association between subjective perception and objective indicators and reveals notable heterogeneity in the eastern and western regions. Thus, it provides a scientific basis for rural environment improvement strategies in different regions. Overall, this study substantially supports the quantitative evaluation of rural environment perception and lays a theoretical foundation for developing more precise and effective rural revitalization strategies.

Acknowledgments

We are very grateful for the comments and suggestions from the editors and anonymous reviewers. We also greatly appreciate the scoring volunteers who assisted with this research.

Disclosure Statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the National Natural Science Foundation of China (42171466, 42471491), the CUG Scholar Scientific Research Funds at China University of Geosciences (Wuhan; 2022034), the Guangdong–Hong Kong–Macau Joint Laboratory Program (2020B1212030009), and the Open Fund of Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources (KF-2023-08-04).

Supplemental Material

Supplemental data for this article can be accessed on the publisher's site at: https://doi.org/10.1080/24694452.2025.2505682XXX

ORCID

Yatao Zhang http://orcid.org/0000-0001-5701-2836
Xiaoqin Yan http://orcid.org/0009-0000-3664-1465
Yao Yao http://orcid.org/0000-0002-2830-0377
Qingfeng Guan http://orcid.org/0000-0002-7392-3709
Xun Li http://orcid.org/0000-0002-7190-0853

Data and Codes Availability Statement

The demo data and codes that support the findings of this study are available on Figshare at. https://doi.org/10.6084/m9.figshare.27087652

References

Antequera, M. L., P. Gargallo, M. Hofinger, S. R. Bulò, Y. Kuang, and P. Kontschieder. 2020. Mapillary planet-scale depth dataset. In *Computer vision – ECCV 2020*, ed. A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, 589–604. Cham, Switzerland: Springer International. doi: 10.1007/978-3-030-58536-5_35.

Barnum, G., S. Talukder, and Y. Yue. 2020. On the benefits of early fusion in multimodal representation learning. arXiv. doi: 10.48550/ARXIV.2011.07191.

Biljecki, F., and K. Ito. 2021. Street view imagery in urban analytics and GIS: A review. *Landscape and Urban Planning* 215 (November):104217. doi: 10.1016/j.landurbplan.2021.104217.

Brambert, P., and I. Kiniorska. 2018. Changes in the standard of living in rural population of Poland in the period of the EU membership. *European Countryside* 10 (2):263–79. doi: 10.2478/euco-2018-0016.

Chen, J., L. Chen, Y. Li, W. Zhang, and Y. Long. 2023. Measuring physical disorder in urban street spaces: A large-scale analysis using street view images and deep learning. *Annals of the American Association of Geographers* 113 (2):469–87. doi: 10.1080/24694452.2022.2114417.

Chen, P., S. Liu, H. Zhao, X. Wang, and J. Jia. 2020. GridMask data augmentation. *arXiv*. doi: 10.48550/ARXIV.2001.04086.

Chen, S., S. Lin, Y. Yao, and X. Zhou. 2024. Urban public space safety perception and the influence of the built environment from a female perspective: Combining street view data and deep learning. *Land* 13 (12):2108. doi: 10.3390/land13122108.

Chiang, Y.-C., H.-H. Liu, D. Li, and L.-C. Ho. 2023. Quantification through deep learning of sky view factor and greenery on urban streets during hot and cool seasons. *Landscape and Urban Planning* 232 (April):104679. doi: 10.1016/j.landurbplan.2022.104679.

Chopra, S., R. Hadsell, and Y. LeCun. 2005. Learning a similarity metric discriminatively, with application to face verification. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*,Vol. 1, ed. C. Schmid, S. Soatto, and C. Tomasi, 539–46. San Diego, CA: IEEE. doi: 10.1109/CVPR.2005.202.

Ding, J., J. Salinas-Jiménez, and M. del Mar Salinas-Jiménez. 2021. The impact of income inequality on subjective well-being: The case of China. *Journal of Happiness Studies* 22 (2):845–66. doi: 10.1007/s10902-020-00254-4.

Dubey, A., N. Naik, D. Parikh, R. Raskar, and C. A. Hidalgo. 2016. Deep learning the city: Quantifying urban perception at a global scale. In *Computer Vision—ECCV 2016*, ed. B. Leibe, J. Matas, N. Sebe, and M. Welling, 196–212. Cham, Switzerland: Springer International. doi: 10.1007/978-3-319-46448-0_12.

Fang, C., H. Ma, C. Bao, Z. Wang, G. Li. S. Sun, and Y. Fan. 2022. Urban–rural human settlements in China: Objective evaluation and subjective well-being. *Humanities and Social Sciences Communications* 9 (1):1–14. London: Palgrave. doi: 10.1057/s41599-022-01417-9.

General Office of the State Council of the Communist Party of China. 2021. The five-year action program to improve the rural living environment (2021–2025). https://www.gov.cn/zhengce/2021-12/05/content_5655984.htm.

Gong, W., X. Huang, M. White, and N. Langenheim. 2023. Walkability perceptions and gender differences in urban fringe new towns: A case study of Shanghai. *Land* 12 (7):1339. doi: 10.3390/land12071339.

Goodchild, M. F. 2007. Citizens as sensors: The world of volunteered geography. *GeoJournal* 69 (4):211–21. doi: 10.1007/s10708-007-9111-y.

Graepel, T., T. Minka, and R. Herbrich. 2007. A Bayesian skill rating system. *Advances in Neural Information Processing Systems* 19:569–76.

Guo, Y., P. Chen, Y. Zhu, and H. Zhang. 2023. Study on comprehensive evaluation of human settlements quality in Qinghai Province, China. *Ecological Indicators* 154 (October):110520. doi: 10.1016/j.ecolind.2023.110520.

Ho, H. C., H. Y. Man, M. S. Wong, Y. Shi, and B. B. Walker. 2020. Perceived differences in the (re)production of environmental deprivation between subpopulations: A study combining citizens' perceptions with remote-sensed and administrative data. *Building and Environment* 174 (May):106769. doi: 10.1016/j.buildenv.2020.106769.

Hou, Y., M. Quintana, M. Khomiakov, W. Yap, J. Ouyang, K. Ito, Z. Wang, T. Zhao, and F. Biljecki. 2024. Global streetscapesA comprehensive dataset of 10 million street-level images across 688 cities for urban science and analytics. *ISPRS Journal of Photogrammetry and Remote Sensing* 215 (September):216–38. doi: 10.1016/j.isprsjprs.2024.06.023.

Howell, C. R., W. Su, A. F. Nassel, A. A. Agne, and A. L. Cherrington. 2020. Area based stratified random sampling using geospatial technology in a community-based survey. *BMC Public Health* 20 (1):1678. doi: 10.1186/s12889-020-09793-0.

Hu, Q., and C. Wang. 2020. Quality evaluation and division of regional types of rural human settlements in China. *Habitat International* 105 (November):102278. doi: 10.1016/j.habitatint.2020.102278.

Ito, K., and F. Biljecki. 2021. Assessing bikeability with street view imagery and computer vision. *Transportation Research Part C: Emerging Technologies* 132 (November):103371. doi: 10.1016/j.trc.2021.103371.

Ji, H., L. Qing, L. Han, Z. Wang, Y. Cheng, and Y. Peng. 2021. A new data-enabled intelligence framework for evaluating urban space perception. *ISPRS International Journal of Geo-Information* 10 (6):400. doi: 10.3390/ijgi10060400.

Krupowicz, W., A. Czarnecka, and M. Grus. 2020. Implementing crowdsourcing initiatives in land consolidation procedures in Poland. *Land Use Policy* 99 (December):105015. doi: 10.1016/j.landusepol.2020.105015.

Kruse, J., Y. Kang, Y.-N. Liu, F. Zhang, and S. Gao. 2021. Places for play: Understanding human perception of playability in cities using street view images and deep learning. *Computers, Environment and Urban Systems* 90 (November):101693. doi: 10.1016/j.compenvurbsys.2021.101693.

Li, X., H. Yang, J. Jia, Y. Shen, and J. Liu. 2021. Index system of sustainable rural development based on the concept of ecological livability. *Environmental Impact Assessment Review* 86 (January):106478. doi: 10.1016/j.eiar.2020.106478.

Li, Y., S. Cheng, Z. Li, H. Song, M. Guo, Z. Li, H.-P. Mang, Y. Xu, C. Chen, D. Basandorj, et al. 2021. Using system dynamics to assess the complexity of rural toilet retrofitting: Case study in eastern China. *Journal of Environmental Management* 280 (February):111655. doi: 10.1016/j.jenvman.2020.111655.

Li, Y.-R., L.-Z. Cao, P.-Y. Wang, and G.-J. Chang. 2022. Rural living environment improvement and rural revitalization. *Journal of Natural Resources* 37 (1):96. doi: 10.31497/zrzyxb.20220107.

Li, Z., Z. Chen, W.-S. Zheng, S. Oh, and K. Nguyen. 2022. AR-CNN: An attention ranking network for learning urban perception. *Science China Information Sciences* 65 (1):112104. doi: 10.1007/s11432-019-2899-9.

Liu, L., B. Zhou, J. Zhao, and B. D. Ryan. 2016. C-IMAGE: City cognitive mapping through geo-tagged photos. *GeoJournal* 81 (6):817–61. doi: 10.1007/s10708-016-9739-6.

Liu, R., L. Zhang, Y. Tang, and Y. Jiang. 2024. Understanding and evaluating the resilience of rural human settlements with a social-ecological system framework: The case of Chongqing municipality, China. *Land Use Policy* 136 (January):106966. doi: 10.1016/j.landusepol.2023.106966.

Liu, S., J. Ge, M. Bai, M. Yao, L. He, and M. Chen. 2022. Toward classification-based sustainable revitalization: Assessing the vitality of traditional villages. *Land Use Policy* 116 (May):106060. doi: 10.1016/j.landusepol.2022.106060.

Liu, Y. 2018. Research on the urban–rural integration and rural revitalization in the new era in China. *Acta Geographica Sinica* 73 (4):637–50.

Liu, Y., X. Liu, S. Gao, L. Gong, C. Kang, Y. Zhi, G. Chi, and L. Shi. 2015. Social sensing: A new approach to understanding our socioeconomic environments. *Annals of the Association of American Geographers* 105 (3):512–30. doi: 10.1080/00045608.2015.1018773.

Lu, Y., and H.-M. Chen. 2024. Using Google street view to reveal environmental justice: Assessing public perceived walkability in Macroscale City. *Landscape and Urban Planning* 244 (April):104995. doi: 10.1016/j.landurbplan.2023.104995.

Lv, Q., X. Xie, and Y. Li. 2019. The effects of economic benefits on place affect: The mediating role of perceived values and non-material life satisfaction. *Asia Pacific Journal of Tourism Research* 24 (1):1–13. doi: 10.1080/10941665.2018.1515091.

Min, W., S. Mei, L. Liu, Y. Wang, and S. Jiang. 2020. Multi-task deep relative attribute learning for visual urban perception. *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society* 29:657–69. doi: 10.1109/TIP.2019.2932502.

Moor, N. J. A., K. Hamers, and M. Mohammadi. 2022. Ageing well in small villages: What keeps older adults happy? Environmental indicators of residential satisfaction in four Dutch villages. *International Journal of Environmental Research and Public Health* 19 (7):3922. doi: 10.3390/ijerph19073922.

Ogawa, Y., T. Oki, C. Zhao, Y. Sekimoto, and C. Shimizu. 2024. Evaluating the subjective perceptions of streetscapes using street-view images. *Landscape and Urban Planning* 247 (July):105073. doi: 10.1016/j.landurbplan.2024.105073.

Pang, Y., W. Zhang, and H. Jiang. 2024. A socio-spatial exploration of rural livability satisfaction in megacity Beijing, China. *Ecological Indicators* 158 (January):111368. doi: 10.1016/j.ecolind.2023.111368.

Qi, Z., J. Li., Z. He, and X. Yang. 2024. The influence of urban streetscape color on tourists' emotional perception based on streetscape images. *Journal of Geo-information Science* 26:514–29.

Qiu, W., W. Li, X. Liu, Z. Zhang, X. Li, and X. Huang. 2023. Subjective and objective measures of streetscape perceptions: Relationships with property value in Shanghai. *Cities* 132 (January):104037. doi: 10.1016/j.cities.2022.104037.

Rossetti, T., H. Lobel, V. Rocco, and R. Hurtubia. 2019. Explaining subjective perceptions of public spaces as a function of the built environment: A massive data approach. *Landscape and Urban Planning* 181 (January):169–78. doi: 10.1016/j.landurbplan.2018.09.020.

Salesses, P., K. Schechtner, and C. A. Hidalgo. 2013. The collaborative image of the city: Mapping the inequality of urban perception. PLoS *ONE* 8 (7):e68400. doi: 10.1371/journal.pone.0068400.

Sánchez, I. A. V., and S. M. Labib. 2024. Accessing eye-level greenness visibility from open-source street view images: A methodological development and implementation in multi-city and multi-country contexts. *Sustainable Cities and Society* 103 (April):105262. doi: 10.1016/j.scs.2024.105262.

Sharp, J. S., K. Agnitsch, V. Ryan, and J. Flora. 2002. Social infrastructure and community economic development strategies: The case of self-development and industrial recruitment in rural Iowa. *Journal of Rural Studies* 18 (4):405–17. doi: 10.1016/S0743-0167(02)00011-6.

Shi, R., J. Xu, and J. Ma. 2023. Analysis of the goal and path of rural living environment upgrade under the strategy of rural revitalization. *Journal of Agricultural Resources and Environment* 40 (5):1104–10. doi: 10.13254/j.jare.2023.0256.

Shucksmith, M. 2018. Re-imagining the rural: From rural idyll to good countryside. *Journal of Rural Studies* 59 (April):163–72. doi: 10.1016/j.jrurstud.2016.07.019.

Sjaf, S., Sampean, A., Aulia Arsyad, L. Elson, A. R. Mahardika, L. Hakim, S. Anom Amongjati, R. Gandi, Z. Ardinal Barlan, I. Made Godya Aditya, et al. 2022. Data Desa Presisi: A new method of rural data collection. *MethodsX* 9:101868. doi: 10.1016/j.mex.2022.101868.

Sun, M., F. Zhang, F. Duarte, and C. Ratti. 2022. Understanding architecture age and style through deep learning. *Cities* 128 (September):103787. doi: 10.1016/j.cities.2022.103787.

Sun, X., D. Chu, Q. Li, Y. Shi, L. Lu, and A. Bi. 2023. Dynamic successive assessment of rural human settlements environment in China. *Ecological Indicators* 157 (December):111177. doi: 10.1016/j.ecolind.2023.111177.

Wang, F., and M. Prominski. 2020. Landscapes with locality in urban or rural areas. Indoor and Built Environment 29 (8):1047–52. doi: 10.1177/1420326X20938321.

Wang, P., X. Qin, and Y. Li. 2021. Satisfaction evaluation of rural human settlements in northwest China: Method and application. *Land* 10 (8):813. doi: 10.3390/land10080813.

Wang, Y., S. Shi, Z. Zhou, S. Guo, and B. Zhao. 2024. Air infiltration rate distribution across Chinese five climate zones: A modelling study for rural residences. *Building and Environment* 252 (March):111284. doi: 10.1016/j.buildenv.2024.111284.

Wang, Y., Y. Zhu, and M. Yu. 2019. Evaluation and determinants of satisfaction with rural livability in China's less-developed eastern areas: A case study of Xianju County in Zhejiang Province. *Ecological Indicators* 104 (September):711–22. doi: 10.1016/j.ecolind.2019.05.054.

Wang, Z., K. Ito, and F. Biljecki. 2024. Assessing the equity and evolution of urban visual perceptual quality with time series street view imagery. *Cities* 145 (February):104704. doi: 10.1016/j.cities.2023.104704.

Xiao, C., J. Zhou, X. Shen, J. Cullen, S. Dobson, F. Meng, and X. Wang. 2022. Rural living environment governance: A survey and comparison between two villages in Henan Province of China. Sustainability 14 (21):14136. doi: 10.3390/su142114136.

Xing, Y., and J. Leng. 2024. Evaluation of public space in traditional villages based on eye tracking technology. *Journal of Asian Architecture and Building Engineering* 23 (1):125–39. doi: 10.1080/13467581.2023.2229410.

Yang, N., Z. Deng, F. Hu, Q. Guan, Y. Chao, and L. Wan. 2024. Urban perception assessment from street view images based on a multifeature integration encompassing human visual attention. *Annals of the American Association of Geographers* 114 (7):1424–42. doi: 10.1080/24694452.2024.2363783.

Yao, Y., Z. Liang, Z. Yuan, P. Liu, Y. Bie, J. Zhang, R. Wang, J. Wang, and Q. Guan. 2019. A human–machine adversarial scoring framework for urban perception assessment using street-view images. *International Journal of Geographical Information Science* 33 (12):2363–84. doi: 10.1080/13658816.2019.1643024.

Yao, Y., J. Wang, Y. Hong, C. Qian, Q. Guan, X. Liang, L. Dai, and J. Zhang. 2021. Discovering the homogeneous geographic domain of human perceptions from street view images. *Landscape and Urban Planning* 212 (August):104125. doi: 10.1016/j.landurbplan.2021.104125.

Yu, X., J. Ma, Y. Tang, T. Yang, and F. Jiang. 2024. Can we trust our eyes? Interpreting the misperception of road safety from street view images and deep learning. *Accident; Analysis and Prevention* 197 (March):107455. doi: 10.1016/j.aap.2023.107455.

Yuan, M., J. Yang, J. Gong, Y. Wang, L. Wang, and Y. Sun. 2024. Strategies for enhancing rural vitality from the perspective of comprehensive land consolidation: Integrating production, living, ecology, and efficiency enhancement. Land 13 (12):2092. doi: 10.3390/land13122092.

Yuan, Y., R. Wang, T. Niu, and Y. Liu. 2023. Using street view images and a geographical detector to understand how street-level built environment is associated with urban poverty: A case study in Guangzhou. *Applied Geography* 156 (July):102980. doi: 10.1016/j.apgeog.2023.102980.

Yusheng, K., and A. H. Ntarmah. 2021. Developing a rural ecological assessment index for Jiangsu Province, China. *Journal for Nature Conservation* 64 (December):126093. doi: 10.1016/j.jnc.2021.126093.

Zhang, F., B. Zhou, L. Liu, Y. Liu, H. H. Fung, H. Lin, and C. Ratti. 2018. Measuring human perceptions of a large-scale urban region using machine learning. *Landscape and Urban Planning* 180 (December):148–60. doi: 10.1016/j.landurbplan.2018.08.020.

Zhang, H., J. Liu, and Q. Guo. 2024. Evaluation of rural human settlements in China based on the combined model of DPSIR and PLS-SEM. *Environmental Impact Assessment Review* 108 (September):107617. doi: 10.1016/j.eiar.2024.107617.

Zhang, Y., S. Li, R. Dong, H. Deng, X. Fu, C. Wang, T. Yu, T. Jia, and J. Zhao. 2021. Quantifying physical and psychological perceptions of urban scenes using deep learning. *Land Use Policy* 111 (December):105762. doi: 10.1016/j.landusepol.2021.105762.

Zhang, Y., L. Wang, R. Dong, H. Deng, X. Fu, B. Huang, Z. Niu, and F. Chen. 2023. Understanding the effects of urban perceptions on housing rent using big data and machine learning. *International Journal of Sustainable Development & World Ecology* 30 (8):964–80. doi: 10.1080/13504509.2023.2234332.

Zhao, X., Y. Lu, and J. Wang. 2024. Multiscale scoring model for enhanced urban perception evaluation. In *ICASSP 2024–2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, ed. J. Hansen, 2935–39. Piscataway, NJ: The Institute of Electrical and Electronics Engineers, Inc. doi: 10.1109/ICASSP48485.2024.10447909.

Zheng, J., J. He, and H. Tang. 2023. The vitality of public space and the effects of environmental factors in Chinese suburban rural communities based on tourists and residents. *International Journal of Environmental Research and Public Health* 20 (1):263. doi: 10.3390/ijerph20010263.

Zhou, B., A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba. 2018. Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40 (6):1452–64. doi: 10.1109/TPAMI.2017.2723009.

Zhou, H., L. Liu, J. Wang, K. Wilson, M. Lan, and X. Gu. 2024. A multiscale assessment of the impact of perceived safety from street view imagery on street crime. *Annals of the American Association of Geographers* 114 (1):69–90. doi: 10.1080/24694452.2023.2249975.

Zhou, M., S. Chu, and X. Du. 2019. Safeguarding traditional villages in China: The role and challenges of rural heritage preservation. *Built Heritage* 3 (2):81–93. doi: 10.1186/BF03545729.

Zhou, Y., Y. Li, and C. Xu. 2020. Land consolidation and rural revitalization in China: Mechanisms and paths. *Land Use Policy* 91 (February):104379. doi: 10.1016/j.landusepol.2019.104379.

KUNKUN ZHU is a Graduate Student at the School of Geography and Information Engineering, China University of Geosciences, Wuhan 430078, China. E-mail: zhukun2023@cug.edu.cn. His research interests focus on geospatial big data mining and geospatial artificial intelligence (GeoAI).

YU GU is a PhD Candidate at the School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China. E-mail: guyu29@mail2.sysu.edu.cn. His research explores rural development studies, geo-intelligence systems, and big data applications in geographical analysis.

YATAO ZHANG is a Doctoral Candidate at the Future Resilient Systems Program, Singapore-ETH Center, ETH Zurich, Singapore 138602. E-mail: yatzhang@ethz.ch. His research interests lie in GeoAI, context-based spatiotemporal analysis, and urban computing.

YANDI SONG is a Graduate Student at the School of Geography and Information Engineering, China University of Geosciences, Wuhan 430078, China. E-mail: songyandi@stu.sicnu.edu.cn. His research addresses social perception computing, urban visual intelligence, and GeoAI methodologies.

ZIHAO GUO is a Graduate Researcher at the School of Geography and Information Engineering, China University of Geosciences, Wuhan 430078, China. E-mail: gzh19@cug.edu.cn. His work integrates geospatial big data mining with data-driven urban modeling techniques.

XIAOQIN YAN is a PhD Candidate at the Institute of Remote Sensing and Geographical Information Systems, Peking University, Beijing 100871, China. E-mail: yanxiaoqin@stu.pku.edu.cn. His studies focus on spatiotemporal big data computing frameworks and social sensing applications.

YAO YAO is a Professor at China University of Geosciences (Wuhan), Hitotsubashi University, and Reitaku University. E-mail: yaoy@cug.edu.cn. His research interests include spatiotemporal big data mining, social geographic computing, and urban geographic information systems.

QINGFENG GUAN is a Professor at China University of Geosciences, Wuhan, China, 430078. E-mail: guanqf@cug.edu.cn. His research interests include high-performance spatial intelligence computation and urban computing.

XUN LI is a Professor at the School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China. E-mail: lixun@mail.sysu.edu.cn. His research examines urban planning innovation mechanisms and economic geography spatial patterns.