

Article

Perceptual Evaluation of Street Quality in Underdeveloped Ethnic Areas: A Random Forest Method Combined with Human–Machine Confrontation Framework Provides Insights for Improved Urban Planning—A Case Study of Lhasa City

Chong Liu , Yang Yu * and Xian Yang 

School of Architecture, Southwest Jiaotong University, Chengdu 611756, China; Incj371@my.swjtu.edu.cn (C.L.); 2020300362@my.swjtu.edu.cn (X.Y.)

* Correspondence: yuyang@swjtu.edu.cn

Abstract: The utilization of street view big data is increasingly being used to uncover visual characteristics and spatial perceptions of urban streets. However, there is a lack of studies that combine street view big data for perceptual evaluation in underdeveloped ethnic areas and better street quality. This study integrates deep learning methods to create a human–computer confrontational model for perception score, with a focus on the central city of Lhasa in Tibet. Pearson correlation analysis was conducted on six dimensions of perception data (beautiful, wealthy, safe, lively, boring and depressing) and visual elements. The streets in the top 20% for both visual elements and perceptual scores were identified to reveal areas with high visual element proportions and high perceptual scores. The spatial distribution characteristics and correlation between visual elements and street perceptions were thoroughly analyzed. The findings of this study reveal that the central city of Lhasa exhibited high percentages of visual elements in buildings (88.23%), vegetation (89.52%), and poles (3.14%). Out of the six perceptions examined, the highest scores were for boring (69.70) and depressing (67.76) perceptions, followed by beautiful (60.66) and wealthy (59.91) perceptions, with lively (56.68) and safe (50.64) perceptions receiving the lowest scores. Visual components like roads (−0.094), sidewalks (−0.031), fences (−0.036), terrain (−0.020), sky (−0.098), cars (−0.016), and poles (−0.075) were observed to have a significant deterring effect on the boring perception, while other visual elements showed a positive influence. This investigation seeks to provide valuable insights for the design and advancement of urban streets in marginalized ethnic localities, addressing a void in perception research of urban streets in such areas.

Keywords: deep learning; human–computer confrontational model; underdeveloped ethnic areas; street quality; perceptual evaluation



Citation: Liu, C.; Yu, Y.; Yang, X. Perceptual Evaluation of Street Quality in Underdeveloped Ethnic Areas: A Random Forest Method Combined with Human–Machine Confrontation Framework Provides Insights for Improved Urban Planning—A Case Study of Lhasa City. *Buildings* **2024**, *14*, 1698. <https://doi.org/10.3390/buildings14061698>

Academic Editor: Derek Clements-Croome

Received: 18 March 2024

Revised: 28 May 2024

Accepted: 3 June 2024

Published: 7 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The problem of unequal urbanization in developing countries has become increasingly apparent during the process of global urbanization [1]. This imbalance is primarily characterized by hindered progress in underdeveloped regions due to various development challenges [2]. From the perspective of development levels in developing nations, it is clear that underdeveloped regions generally exist in the inner periphery of these countries. These regions face numerous difficulties, including unfavorable migration balance, lower living standards, aging populations, limited job opportunities, lower education levels, and reduced regional investments [3]. The conventional model of urbanization encounters hindrances and drawbacks such as ‘risky urbanization’, ‘pseudo-urbanization’, ‘urban diseases’, and ‘urban–rural imbalance’ [3,4]. In the contemporary context, streets are recognized as crucial urban elements that play a significant role in driving economic growth and fostering local development. This is particularly evident in marginalized ethnic regions,

such as Tibetan areas. Therefore, utilizing city streets as a driving force for urbanization presents a favorable option for the progress of underdeveloped ethnic regions [5–7]. In underdeveloped ethnic regions, such as Lhasa City streets, there is a significant presence of Tibetan culture and ambiance. Assessing the quality of streets in Lhasa is essential for the progression of urban areas in underdeveloped ethnic regions and improving the perception of streets in these localities.

Experts have recently conducted extensive investigations in underdeveloped ethnic regions that lack adequate resources. For instance, Zhou [8] employed the TOPSIS entropy weighting method and an integrated coordination model to evaluate the resilience of six cities and states in Sichuan and Tibet between 2015 and 2020. Their research focused on a range of factors, including the economy, society, environment, and infrastructure, with the aim of identifying strategies to enhance urban resilience. They accomplished this by utilizing correlation analysis and gray correlation analysis to identify the primary factors that influence resilience in underdeveloped regions. In another study, Zhang [9] directed their investigation towards the mountainous region located on the western border of Yunnan, which serves as a representative underprivileged area. They utilized a combination of the coupling coordination degree model, spatial gravity model, and panel Tobit model to analyze the temporal and spatial variations in the degree of coupling coordination between the tourism industry and the process of urbanization from 2010 to 2019. Additionally, they explored the factors that influence this coordination. Similarly, Yu [10] employed a GIS-based spatial network analysis method and population distribution data to assess the arrangement of primary healthcare facilities in the built-up area of Lhasa. By implementing the location-configuration model, they were able to optimize the arrangement of these facilities. Existing research on urban environments in underdeveloped ethnic areas primarily focuses on topics such as urban resilience, urbanization, and healthcare. However, there is a lack of research on street perception and the evaluation of urban street environment quality using deep learning methods.

Urban streets play a crucial role in connecting individuals within a city and facilitating daily social interactions among its residents. These streets comprise various elements, including roads, surrounding buildings, and roadside vegetation, all intricately intertwined [11]. The quality of urban streets significantly contributes to enhancing the livability of cities and positively affects the well-being of inhabitants [12]. Previous academic investigations on street quality have primarily focused on comprehending individuals' visual or sensory perceptions of the built environment, employing techniques such as surveys, face-to-face interviews, and onsite observations [13]. Surveys and interviews commonly serve to collect individuals' subjective evaluations of the constructed environment, often employing open-ended queries and numerical rating scales. Nonetheless, concerns have been raised regarding potential biases in participants' responses associated with these methods [14,15]. Personal audits serve as the primary approach to capturing human perceptions and emotional experiences related to urban design across various scales [12]. However, personal audits are characterized by high costs, time constraints, and practical limitations, thereby restricting urban researchers from conducting small-scale empirical studies [14,16].

In recent years, notable advancements have been witnessed in the realm of machine technology. These progressions can be credited to the emergence of mapping services that rely on contributions from the crowd and the wide availability of images with geographic tags. Notably, Baidu Street View images (BSVIs), Google Street View images, and Tencent Street View images are among the convenient services that provide a diverse array of visual data. These data represent the viewpoint of local inhabitants and offer significant insights into the conditions of streets. The influx of such data has made a substantial contribution to achieving a comprehensive understanding and visualization of urban landscapes [17]. The merging of machine technology and research on urban streets has captured the interest of urban planners and researchers. Consequently, there has been a surge in studies that employ advanced learning methods to investigate the perception of street quality and extract valuable information from street view images [18,19]. However, most studies in

this area have focused on developed regions, leaving a gap in research on underdeveloped ethnic regions. These regions have unique characteristics and face more pronounced issues of unbalanced urban development. Therefore, utilizing deep learning to assess the quality of streets in underdeveloped ethnic regions is crucial for promoting balanced development in these areas.

The effectiveness of deep learning in the classification and mapping of urban streets has been well documented. This advanced technology allows for the automatic extraction of semantic data, including pedestrian counts, as well as the categorization of visual elements and representation of scene features in images [20,21]. By understanding how individuals perceive street space, deep learning can provide valuable insights for efficient and thorough urban planning. To demonstrate the capabilities of deep learning in this field, Zhang [22] introduced a data-driven approach that utilizes deep learning to predict human perception of street images. In a similar vein, Zhou [15] employed deep learning techniques to segment and extract physical characteristics from Baidu Maps Street View (BMSV) images in Shenzhen, China, with the goal of evaluating street visuals and promoting healthy urban development. In a groundbreaking study, Zhao [23] developed an innovative deep learning methodology that considers both the street view and subjective factors of human perception to estimate and analyze land value. Moreover, Rui [24] conducted a study in central Düsseldorf, utilizing Google Street View (GSV) and pedestrian views to extract pixel-level semantic information using the DeepLabV3+ segmentation technique. The objective of this research was to identify variations in quantitative measurements of perceptual quality in the street view from different perspectives, thereby facilitating a comprehensive analysis of perceived quality and walking potential to determine priorities for street renewal efforts.

Scientists have expressed a significant interest in investigating the way people perceive the overall quality of urban streets. In a recent study, the researchers [25] delved into this topic by employing both spatial syntax and deep learning techniques. Their focus was specifically on the Binjiang District, which is situated in Hangzhou City, Zhejiang Province, China. In order to assess the quality of the streets, the scientists utilized semantic image segmentation, making use of the SegNet code structure. To facilitate a comparison between the perceptions of humans and machine analysis, a random forest model was implemented. Additionally, the researchers employed a multivariate linear regression approach to understand the relationship between the spatial characteristics of a street and its constituent elements. This study presents a framework that utilizes deep learning, street view images, and iterative feedback mechanisms to assess street perception on an urban scale. The Cityscapes Dataset, considered the most up-to-date, accurate, and comprehensive street view dataset, was utilized for training, validation, and testing using the DeeplabV3+ semantic segmentation method for road images. Visual elements in 53,820 street images from 14,002 segmented sampling points in Lhasa's city center were categorized. This study employed a random forest model within a human-computer confrontational framework to score six dimensions of street perception (beautiful, wealthy, safe, lively, boring, and depressing). Streets with the highest percentage of visual elements and top street perception scores in each dimension were identified and overlaid in GIS. Correlation and spatial distribution analyses were conducted to explore the relationship between visual elements and street view perception in Lhasa's city center. This research sheds light on the spatial heterogeneity of street view perception in underdeveloped ethnic areas, offering valuable insights for urban planning departments in such regions. The findings are particularly relevant for the urban development of underdeveloped ethnic areas, addressing the existing gap in urban street perception studies within these regions.

2. Literature Review

2.1. A Study of City Development and Streets in Underdeveloped Ethnic Areas

At present, many scholars are actively exploring the urban development of underdeveloped ethnic areas. Among them, monitoring changes in land use in coastal regions is crucial for promoting sustainable development worldwide. However, previous studies

have primarily focused on modifications in developed areas, creating a knowledge gap regarding economically undeveloped regions experiencing rapid progress. To bridge this gap, Liang [26] conducted a study in the coastal vicinity of Vietnam. Using the random forest model and the land use alteration index, they assessed spatial variations in land use changes and identified key factors. Their findings offer valuable insights into land use research in economically undeveloped regions undergoing rapid progress. In a separate investigation, Liang [27] examined land use alterations along Highway No. 4 in Cambodia. They analyzed land use data from the GlobeLand30 2020 database for the years 2000, 2010, and 2020. The study focused on the characteristics and drivers of these changes and proposed that leveraging tourism for urbanization is a beneficial strategy in underdeveloped ethnic areas. Expanding upon this research, Zhang [9] studied the mountainous region along the western border of Yunnan, which serves as an exemplary underdeveloped ethnic area. They utilized the coupling coordination degree model, spatial gravity model, and panel Tobit model to investigate the spatio-temporal characteristics of the coupling coordination degree between tourism and urbanization from 2010 to 2019 while identifying influential factors. The primary objective of this study is to investigate the relationship between tourism and urbanization, specifically focusing on understanding the interconnected evolutionary process that connects these two phenomena. Moreover, it examines the advancement of street vending in economically disadvantaged ethnic streets, an area that has attracted considerable academic attention. As described in the research results by Hossam [28], the presence of merchants from various ethnic backgrounds in places like Malmi significantly contributes to the liveliness of public streets and plays a crucial role in enhancing the dynamism of the streets. Additionally, informal settlements in the Greater Cairo region of Egypt exhibit a street arrangement that promotes social interaction among residents. By conducting a comprehensive analysis on an urban scale, it becomes possible to identify specific intersections within the street system of these settlements that may form potential connections with the broader transportation network, thereby influencing regional integration [29].

2.2. Geospatial Big Data and Street Environment Perception

In recent years, there has been a rapid increase in multi-source big geospatial data, especially the rise of multiple image datasets with geographic tags [30]. These images, when combined with location data, contain a plethora of visual information [31] that accurately portrays the visual aspects of daily life [32]. Since landscapes play a crucial role in shaping how city dwellers perceive their surroundings [33], these geotagged image collections offer fresh possibilities for comprehending urban perspectives on a large scale. Street View Imagery (SVI), providing comprehensive views of various streets, has emerged as a valuable asset for deducing urban perceptions. Street view photos are geotagged pictures gathered, processed, and managed by popular mapping services such as Google Maps and Tencent Maps through standard techniques. Captured using specialized gear at multiple locations and landmarks, primarily along city roads [34], these images showcase the physical attributes of urban environments [35,36].

Salesses [37] suggested using street view images to collect human opinions by comparing them in pairs to analyze the impact of urban environments on social and economic results. Expanding on this idea, Dubey [38] utilized online crowdsourcing and machine learning in computer vision to broaden the survey scope to major global cities, generating a detailed dataset of urban perceptions. This method overcomes the challenges of small sample sizes and limitations associated with traditional interview and questionnaire approaches. Prior studies have evaluated city perceptions using data from the MIT Place Pulse program (Place Pulse 1.0 and 2.0) [39–42].

2.3. Deep Learning and Street Quality Evaluation Research

Acquiring a comprehensive understanding of the perception of city streets by users is of immense importance to urban planners and managers. This understanding is crucial for

optimizing the overall living environment and effectively attracting talented individuals, investors, and businesses [43]. The level of perceived quality plays a significant role in the emotional connection residents have with city streets, as indicated by research conducted by Low [44]. Specifically, place attachment refers to the emotional connection individuals establish with a particular street, resulting in increased comfort, security, and a stronger inclination to remain in that area [45]. As a result, in recent years, multiple fields such as geoinformatics, sociology, and urban and rural planning have actively contributed to establishing the relationship between street attributes and residents' perceptions [46].

After AlphaGo, developed by Google, defeated the Go world champion, there has been a remarkable increase in the fascination with deep learning. As a result, scientists have extensively utilized deep learning techniques to extract visual semantic features from images for their research purposes [47]. In the previous era, the extraction of semantic characteristics from images mainly depended on methods centered around object recognition. Nevertheless, these techniques faced certain constraints, including difficulties in distinguishing objects with similar colors or detecting elusive elements within the image [48].

Deep convolutional neural networks have been widely used in recent times for processing visual information found in images. The advancements in this field have been driven by the implementation of deep learning algorithms like FCN, ResNet, and SegNet. These algorithms have shown great effectiveness in identifying various visual elements in images, including roads, buildings, the sky, pathways, trees, and vegetation. This has significantly contributed to enhancing the reliability of research involving urban roads and human perception [49–51]. In recent research conducted by Ma [17], they focused on extracting semantic information at the pixel level and organizing visual components using the SegNet approach for image segmentation. They utilized a dataset consisting of one million panoramic street view images from Shenzhen, China. Additionally, they investigated the spatial characteristics of five perceptual elements and studied how perceptual outcomes differ across different functional streets. This allowed them to determine the overall impact of Urban Renewal Projects (URPs) on street landscape transformations. Donghwan [52] conducted a study that emphasized the computation of the Green View Index (GVI) using FCN semantic segmentation. They specifically focused on pedestrians' viewpoints using Google Street View (GSV) and deep learning methodologies. Another research team led by Wang [53] investigated the spatial quality of streets in Xiamen City. To assess the correlation between greenery variables and the Green View Index (GVI), the researchers' main aim was to utilize conventional correlation analysis and a multiple regression model. Additionally, they conducted a study to understand the relationship between walking duration and green space variables. Various resources, such as Street View Imagery (SVI), Points of Interest (POIs), and data collected from social media comments, were employed to support their analysis. In a separate investigation conducted by William [54], they employed images sourced from Google Street View (GSV) and utilized the SegNet technique for image segmentation. The primary focus of their research was to provide valuable insights for enhancing urban streets. To achieve this objective, they developed a concatenated convolutional neural network.

The field of machine science has increasingly focused on street perception modeling, capitalizing on the well-known DeeplabV3+ learning framework, renowned for its ability to extract extensive and precise information about street scenes. Yu [55] proposed a lightweight technique that applies the DeepLabV3+ with complex-valued computation to conduct semantic segmentation on Polarimetric Synthetic Aperture Radar (PolSAR) images. To tackle the issue of overfitting and improve segmentation accuracy in the analysis of polarized SAR data with limited samples, a specific technique was developed. In order to conduct a comprehensive exploration of the variables influencing the perceived excellence of street views and establish priorities for the revitalization of streets, Rui [24] amassed a total of 5300 images of pedestrian landscapes in the downtown vicinity of Düsseldorf. These images were obtained through a combination of Google Street View

(GSV) and self-captured photographs. To attain comprehensive insights, the advanced segmentation technique known as DeepLabV3+ was utilized to extract highly detailed semantic information at the pixel level. The primary aim of this inquiry centered on assessing diverse perspectives regarding perceived street quality and pedestrian potential, ultimately facilitating an exhaustive assessment that informs decision-making in street refurbishment projects.

3. Data and Method

3.1. Study Areas

Lhasa, situated in southwest China on the Tibetan Plateau, serves as the capital of the Tibet Autonomous Region (TAR). Nestled in the valley plain of the Lhasa River, a tributary of the Yarlung Zangbo River, it stands as the political, economic, cultural, and educational hub of Tibet. As a prime example of a city in an underdeveloped ethnic region, Lhasa is the focus of this study (Figure 1).

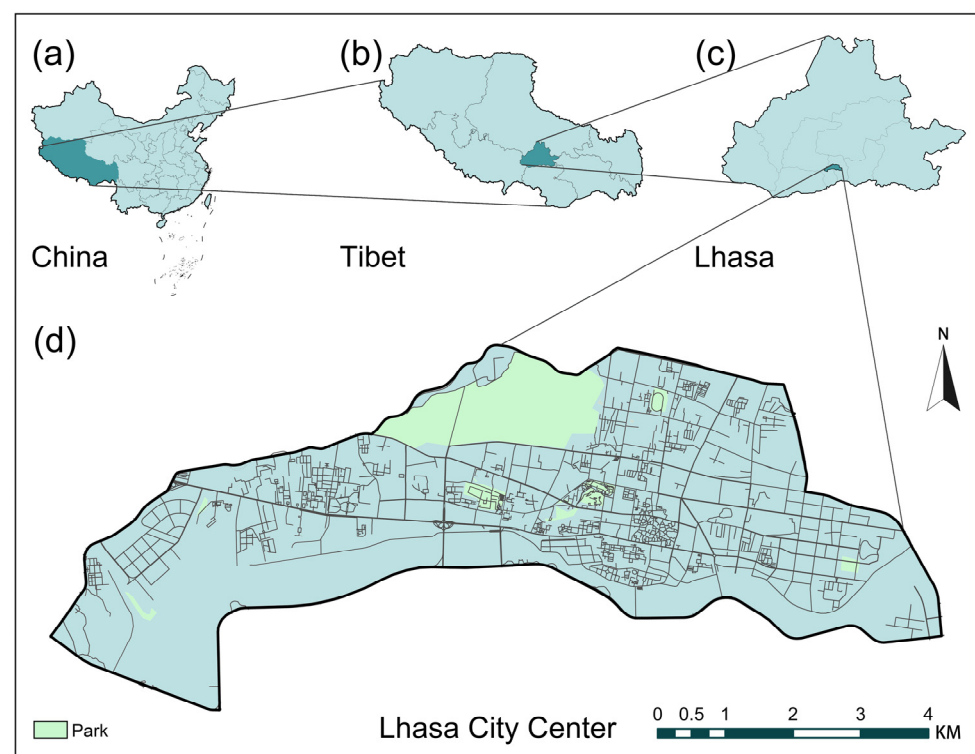


Figure 1. Overview of the study area for the present study: (a) location of Tibet in China, (b) location of Lhasa within Tibet, (c) the central areas of Lhasa City, and (d) the road network in the central areas of Lhasa City.

Located in Lhasa City at an elevation of 3650 m above sea level and spanning 29,640,000 square kilometers, the city had a population of 581,200 as of the conclusion of 2022. With a diverse population of 31 ethnic groups, including Tibetans, Han Chinese, and Hui, recent efforts by the Lhasa Municipal Government have focused on improving urban infrastructure, environmental quality, cityscapes, and resident well-being and safety. Streets are vital components of urban environments, influencing human spatial perception greatly [56]. Assessing the street quality in the central city is crucial for understanding visual transformations linked to street enhancements. The wealth of Baidu Street View images (BSVI) data in Lhasa City increases the precision of research outcomes, making it an ideal setting for the study detailed in this paper.

3.2. Research Framework

The analytical framework and methodology used in this paper are shown in Figure 2.

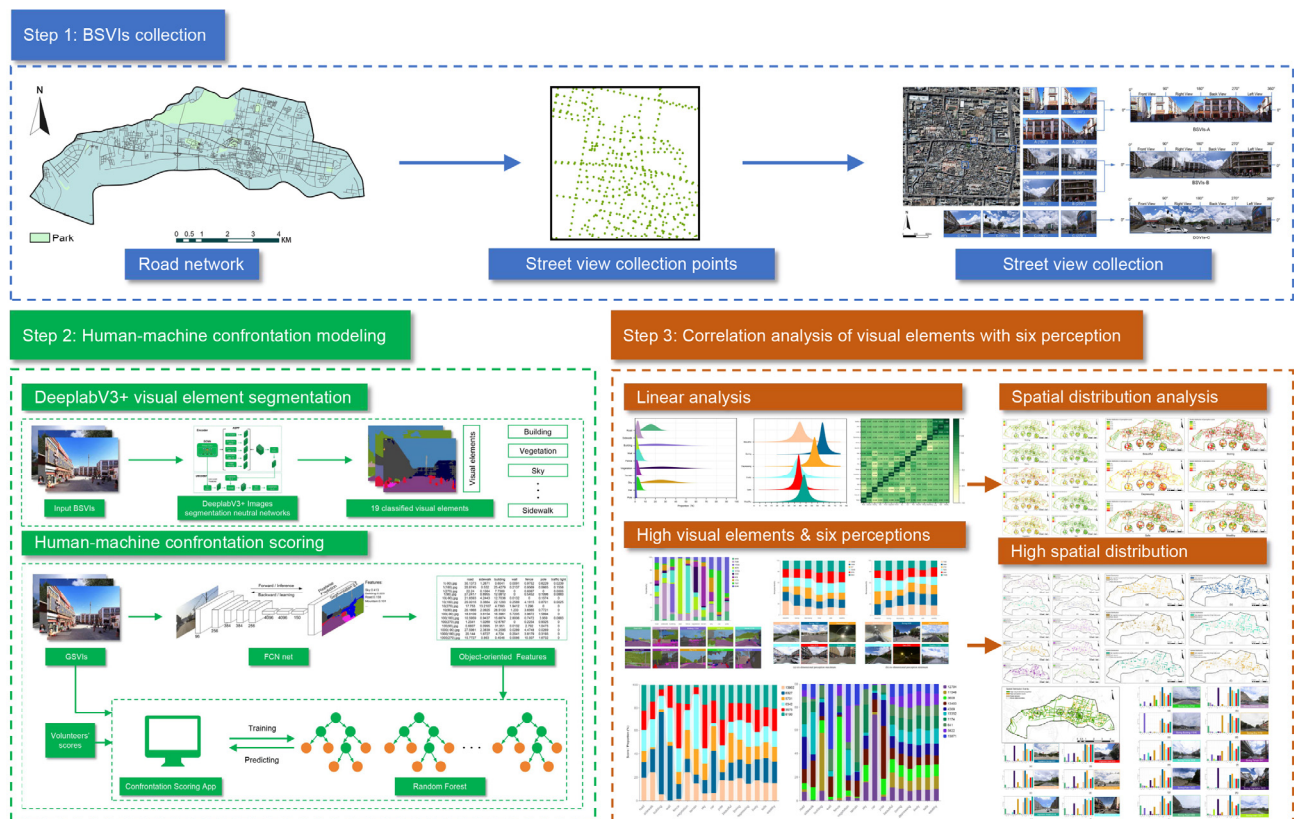


Figure 2. Research framework for this study.

In the first stage, in order to initiate the project, essential data about the research area were gathered using Open Street Map (OSM). Following this, the collected data were fed into ArcGIS to create data points for street view, maintaining a uniform gap of 50 m between each point. Subsequently, the Application Programming Interface (API) of Baidu Street View mapping technology was employed to retrieve Baidu Street View images (BSVIs). Consequently, tailored software was developed to establish a connection with the Baidu Street Map Application Programming Interface (API) and obtain the BSVIs.

In the second stage, our study utilized a Python Torch framework to develop a neural network model for segmenting BSVI data with full-element representation in the second phase. This particular model was trained on the BSVI dataset to establish a confrontational human–computer model for scoring based on perception. Individuals with expertise in the historical and cultural context of the research area were recruited to evaluate the perception of city streets across six dimensions. Utilizing the confrontational model, predictions regarding perception scores were made for the remaining images, aligning with the assessors' scores. Calculation of final spatial perception scores across all six dimensions was achieved through a process of weighting and averaging individual assessors' scores, considering variations from model predictions. To examine the relationships between the perceptual data dimensions and the top visual elements, a Pearson correlation analysis was performed. Following this, the top 20% of images with the highest score across each dimension and the top ten visual elements were selected for a detailed analysis of street view elements.

In the third stage, the following research introduces an innovative method for evaluating the perception of streets in underdeveloped ethnic regions by combining the spatial perception of urban streets with a confrontational analysis framework involving human–computer interaction. Through the integration of visually striking streets with those

exhibiting high scores on six spatial perception metrics, this study seeks to discern spatial connections. The discovery of dimensions associated with the spatial layout and quality of streets in Lhasa's central city, known for its significant visual component, may provide valuable guidance for urban street design and development in underdeveloped ethnic regions.

3.3. BSVI Data Collection

The analysis of city environments and urban streets using street view images has gained significant attention in the realm of urban science [36]. This visual representation primarily emphasizes the human encounters within urban surroundings, enabling a distinctive viewpoint [57]. Numerous platforms dedicated to Street View Imagery provide accessible data and services to delve into this particular kind of visual content.

In order to comprehensively evaluate how residents perceive the central streets, we implemented a data collection system for streets. This system allowed us to gather data every 50 m during the investigation. To generate a comprehensive visualization of the city's perception of the streets, we utilized a Python script to merge the collected images from all four cardinal directions. Specifically focusing on the central district, we obtained a total of 14,002 data collection points from the street network provided by Open Street Map (OSM). By utilizing Python programming and the Baidu Street Map API, we successfully acquired 53,820 images of the street view for subsequent analysis.

To ensure the authentic representation of residents' perspectives, we formulated specific criteria for the Baidu Street View images (BSVIs) derived from the Application Programming Interface (API). These criteria comprise the subsequent elements: the longitude coordinates, which have been converted to the Baidu Street View map BD09; a fixed vertical angle (pitch) of 0° ; and a field of view (fov) width of 90° . In the case of each sampling point, we amalgamate perspectives from four distinctive directions to obtain a comprehensive 360° panoramic depiction of the street. It is noteworthy to mention that the image is confined to a maximum resolution of 400×300 pixels. Figure 3 illustrates the procured street view image. To assure the dependability of the overall evaluation of street quality, we implemented a filtration procedure to eliminate any accessible street view images that were flagged as invalid.

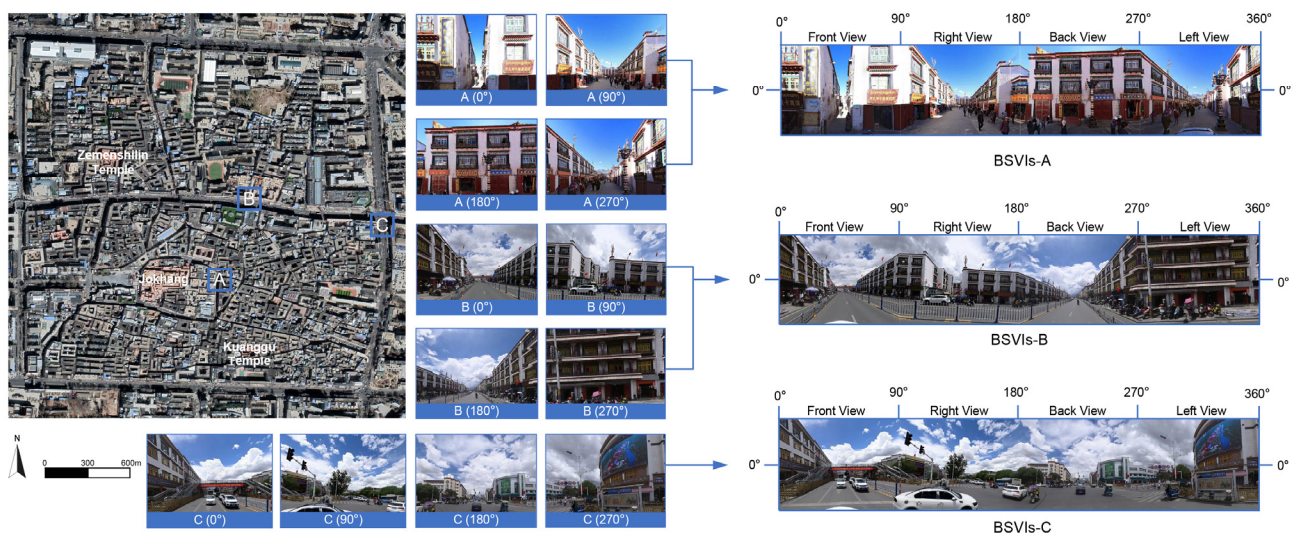


Figure 3. Example of the BSVI collection process in the central areas of Lhasa City.

3.4. Deep Learning-Based Semantic Segmentation and Visual Element Classification for Street View Images

The utilization of the street scene perception model with the DeeplabV3+ learning framework enables the extraction of more comprehensive and accurate street scene in-

formation. In this study, a DeeplabV3+-based semantic segmentation method for road images was implemented using the experimental environment of Windows 11. The training process utilized an NVIDIA GeForce RTX 2080Ti GPU (NVIDIA Corporation, Santa Clara, CA, USA) with 11 GB of graphics memory, and Python Torch was employed as the learning framework. To ensure a robust foundation for the experiments, the Python Torch framework, known for its proficiency in deep learning, was selected. Regarding the dataset employed in this investigation, the Cityscapes Dataset, developed by Mercedes-Benz, was utilized. This dataset serves as a standard to assess the performance of vision algorithms in scene semantic comprehension. It consists of street views from 50 different cities, each with distinctive backgrounds, seasons, and contextual elements, which allows for comprehensive analysis. To effectively categorize various elements, the dataset utilized 19 semantic labels, ranging from 0 to 18. These labels aid in the classification of entities such as roads, buildings, street trees, sky, people, and cars, as illustrated in Table 1 and Figure 4. It is important to note that the dataset was divided into three subsets for accurate evaluation: training, validation, and testing. The training subset consisted of 2975 images, while the validation subset comprised 500 images, and the testing subset encompassed 1525 images. The model xception71_dpc_cityscapes_trainfine, trained by DeepLab-V3+ on the Cityscapes Dataset, was selected for street map recognition in this study due to its algorithm and dataset compatibility.

Table 1. Cityscapes Dataset semantic labeling.

Group	Classes
Flat	Road (0), sidewalk (1)
Buildings	Building (2), wall (3), fence (4)
Object	Pole (5), pole group, traffic light (6), traffic sign (7)
Trees	Vegetation (8), terrain (9)
Sky	Sky (10)
Human	Person (11), rider (12)
Vehicle	Car (13), trunk (14), bus (15), train (16), motorcycle (17), bicycle (18)

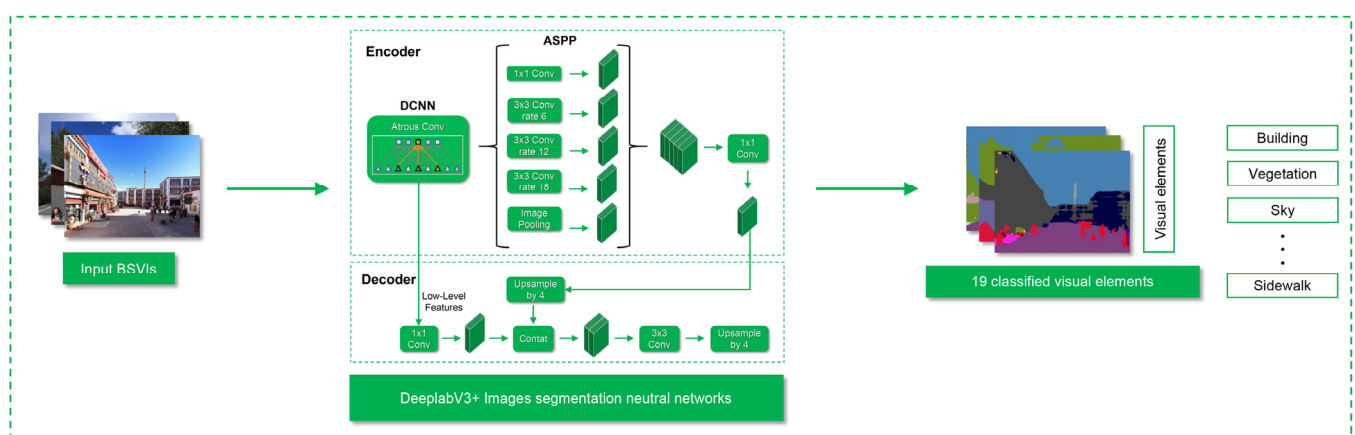


Figure 4. Street view segmentation in central Lhasa City based on DeepLabV3+.

In our scientific investigation, we employed the Cityscapes Dataset for our research investigation. Initially, the training images and labeled files were provided to us in PNG format for the purpose of semantic segmentation. However, to enhance clarity and streamline the training procedure, we purposely decided to convert the format of the training images to JPG. Despite this alteration in format, the critical correspondence between the information conveyed in the training images and the labeled files remains unaltered. Taking into account the specific requirements of the algorithmic criteria and dataset under scrutiny, we intentionally selected the xception71_dpc_cityscapes_trainfine model, which underwent

training utilizing DeepLabV3+ on the Cityscapes Dataset, as our designated model for recognizing street maps.

3.5. Scoring Street Perception Using a Human–Machine Confrontational Scoring Framework

The innate ability of the human brain to recognize scenes gives it a natural advantage in identifying global attributes within images. This advantage serves as the basis for our strategy in the ‘human–machine confrontation’ paradigm, where we leverage this innate ability to assist in human evaluations of street scene perception. The program development and deployment involved predicting perceptions using the human–computer confrontational model algorithm proposed by Yao [58] and deploying it on Tencent cloud servers to collect perceptual data. Training urban environment perception using the MIT Place Pulse dataset [35] aimed to uncover the visual elements that influence how a place is perceived in terms of safety, vibrancy, despondency, and others. The dataset categorizes perceptions into six classes: safe, lively, wealthy, beautiful, boring, and depressing perceptions. Studies have explored the relationship between these perception categories and urban residents’ well-being [59], urban security [22], and the recognition of urban film locations [60]. These studies have shown that these six indicators effectively capture human perception irrespective of cultural backgrounds, income levels, or racial differences among residents [22].

The approach of human–computer confrontational scoring integrates a random forest component to examine the visual attributes of images of urban environments in relation to user assessments. This well-suited model is automatically created once the user rates 50 images, offering suggested ratings for future assessments. The mechanism adapts these recommendations based on user actions. In instances where the suggested ratings notably deviate from user ratings by more than 10 points for over 5 images, the random forest component is retrained to adjust the model settings. The user evaluation procedure concludes when the out-of-bag validation error of the model is under 10 points, leading to the generation of the human–machine confrontational scoring dataset (Figure 5).

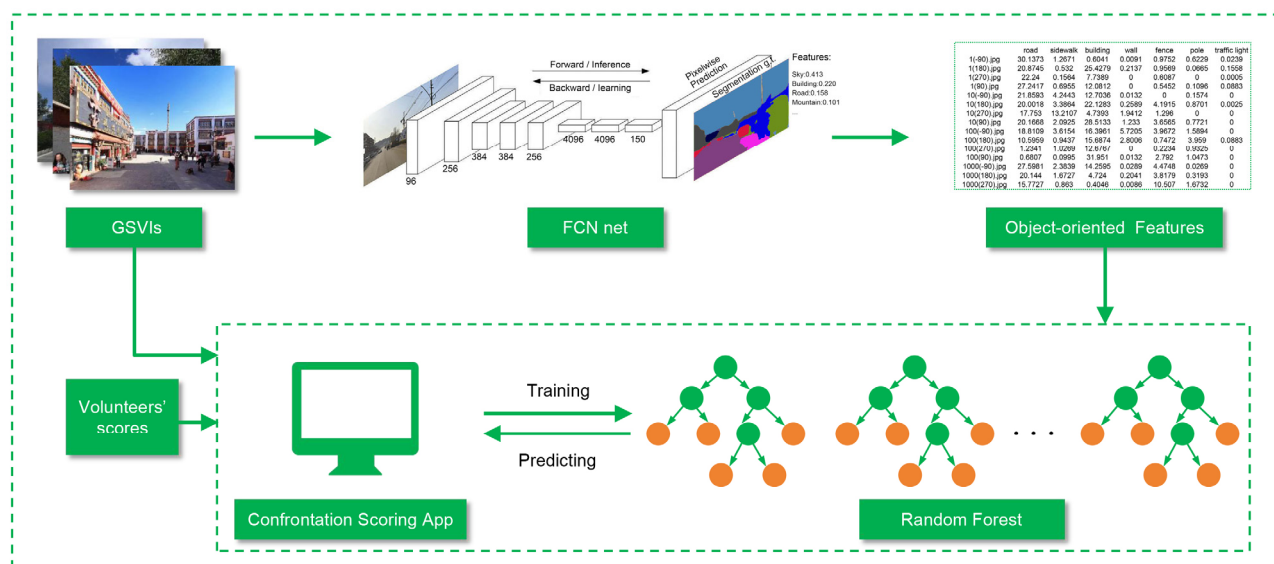


Figure 5. Construction of human–machine confrontation model for street view scoring in central Lhasa.

In the process of training with random forest, the dataset is split randomly into two sets: the training dataset (in-bag) and the testing dataset (out-of-bag). The out-of-bag dataset is used specifically to evaluate the accuracy of the model at each iteration while training. The average validation error from the out-of-bag dataset is a key metric for assessing the effectiveness of the random forest classification model in achieving the highest accuracy

possible. Studies have indicated that using out-of-bag estimation yields better results compared to cross-validation [61].

This research employs various statistical metrics, including the Pearson correlation coefficient (Pearson R), the standardized R-squared (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE), to assess the precision of fitting the human–robot confrontational score to random forests and the RF fitting of city perceptions relative to poi-based city functionalities. The mathematical formulas for Pearson R, standardized R^2 , RMSE, and MAE are delineated in Equations (1) through (4), respectively.

$$\text{Pearson } R = \frac{\sum_{i=1}^n (y_i - \bar{y}_i)(\hat{y}_i - \bar{\hat{y}}_i)}{\sqrt{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}}_i)^2}} \quad (1)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (3)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4)$$

where y_i is the base truth value, $\bar{y}_i$ is equal to $\frac{1}{n} \sum_{i=1}^n y_i$, and $\hat{y}_i$ is the prediction result of the fitted model.

The research methodology employed in this investigation is grounded on the volunteer scoring approach introduced by Wang and colleagues [25], which encompasses six different perception categories for evaluating the quality of urban streets in underdeveloped ethnic regions, specifically concentrating on the central district of Lhasa. To guarantee a thorough assessment of street quality perception, a group of 50 volunteers from various backgrounds including urban planners, architects, sociologists, tourists, college students, and community members well versed in the historical and cultural milieu of the region were recruited. The demographic characteristics of the participants are outlined in Table 2, indicating an average age of 34.60 years, with a majority being male (56.67%). In terms of educational attainment, 26.67% had primary school education or lower, 46.66% had secondary school education, and 26.67% had tertiary education or beyond. The ethnic makeup was predominantly Tibetan (76.44%), with local residents constituting 86.67% of the sample.

Table 2. Statistical table of volunteer information.

Variables	Proportion/Mean (SD)
Gender (%)	
Male	56.67
Female	43.33
Age	34.60 (32.11)
Education (%)	
Primary school or below	26.67
College and above	46.66
High school	26.67
Nation (%)	
Tibetan	76.44
Chinese Han	23.56
Residents (%)	
Local resident	86.67
Non-local resident	13.33

A schematic representation depicting the assessment of urban streets via the human–robot confrontational framework is illustrated in Figure 5. Initially, volunteers evaluated

six criteria: aesthetics, monotony, somberness, vibrancy, wealth, and security. Using the human–computer confrontational framework, the perceived ratings for these criteria were scored on a scale from 1 to 50 to assess the streets’ perceived quality, where a higher score indicated a closer alignment with the specific criterion being evaluated. To aid the scoring process, each volunteer was responsible for subjectively rating the first 2000 street views for each criterion. Following this, the model generated a random forest dataset and offered score recommendations for each criterion based on the relationship between the volunteer’s prior ratings and the visual elements detected in the street views.

4. Results

4.1. Street Quality Analysis Based on BSVIs

By employing cutting-edge DeepLabV3+ technology for image segmentation, this research investigates the spatial arrangement of the top ten visual components found in streets located in the central area of Lhasa (Figures 6 and 7). The central region of Lhasa showcases a greater overall proportion of visual components associated with buildings and vegetation, with buildings constituting 88.23% and vegetation at 89.52%. In contrast, poles make up the tiniest portion at 3.14%. Roads (31.10%), sky (56.27%), terrain (16.27%), and walls (19.49%) display comparable spatial traits in the dispersion of visual components, primarily clustered along the primary thoroughfares. Figure 8 emphasizes the highest percentage of each visual element among the top ten rated streets within Lhasa’s central district. Visual elements like walls, fences, cars, terrain, and poles exhibit a higher ratio of the maximum value, whereas buildings and vegetation closely trail behind. Conversely, roads and sky display a lesser percentage of the maximum value in comparison to the other visual components.

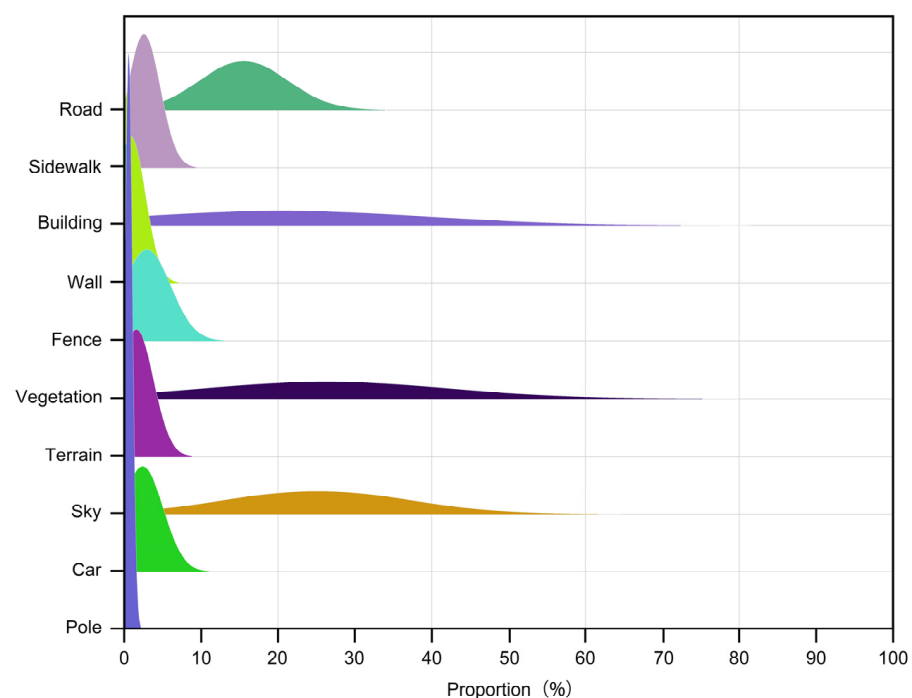


Figure 6. The proportion of the top ten visual elements in Lhasa’s central urban area. The colors in the figure, from top to bottom, represent road, sidewalk, building, wall, fence, vegetation, terrain, sky, car and pole.

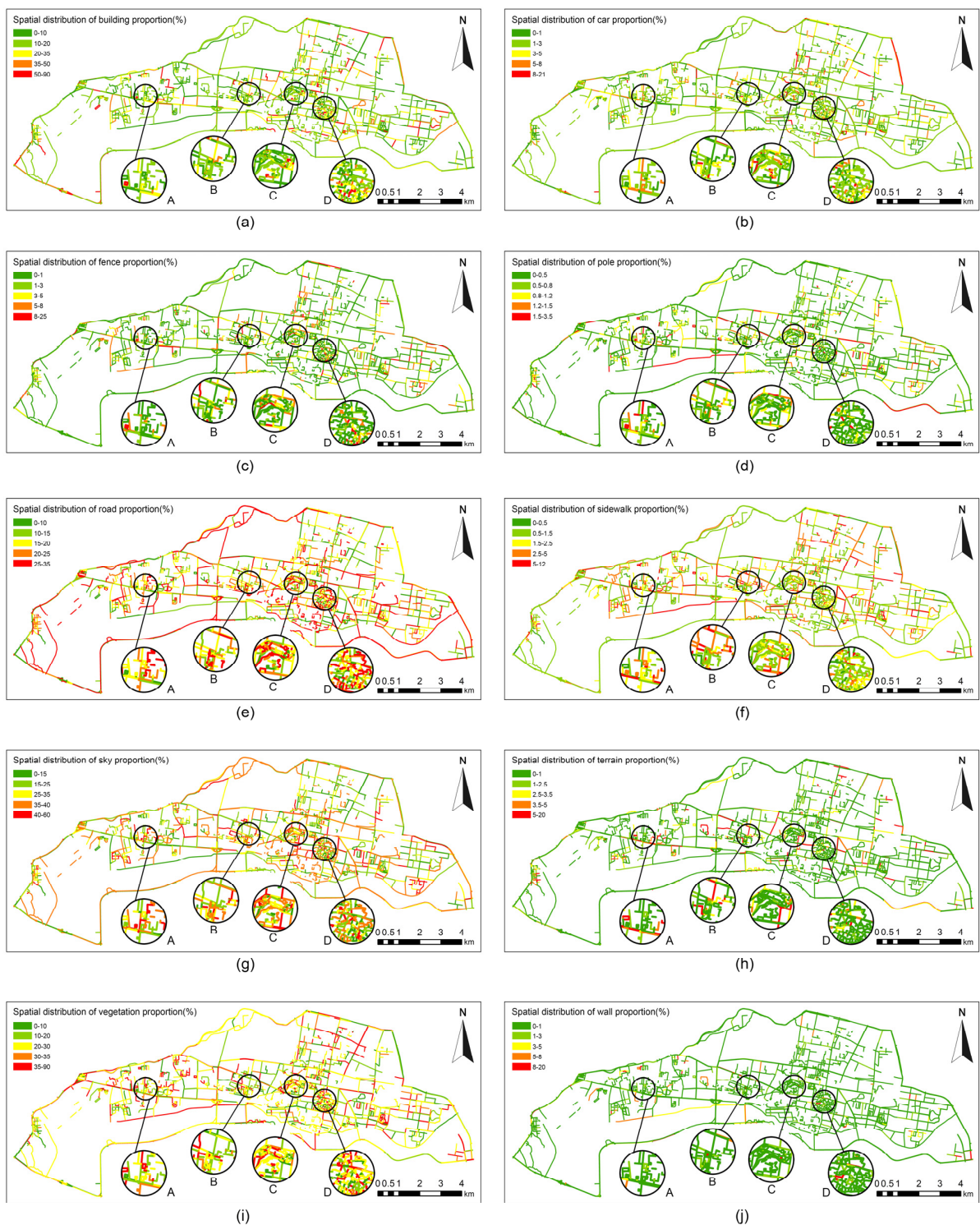


Figure 7. Spatial distribution of the proportion of visual elements in the central urban area of Lhasa: (a) spatial distribution of building proportion, (b) spatial distribution of car proportion, (c) spatial distribution of fence proportion, (d) spatial distribution of pole proportion, (e) spatial distribution of road proportion, (f) spatial distribution of sidewalk proportion, (g) spatial distribution of sky proportion, (h) spatial distribution of terrain proportion, (i) spatial distribution of vegetation proportion, (j) spatial distribution of wall proportion. In the selected area, A represents new residential developments and industrial zones, B represents the central business area, C represents the Potala Palace Scenic Area, and D represents the historic city center, showcases Dazhao Monastery and Barkhor Street.

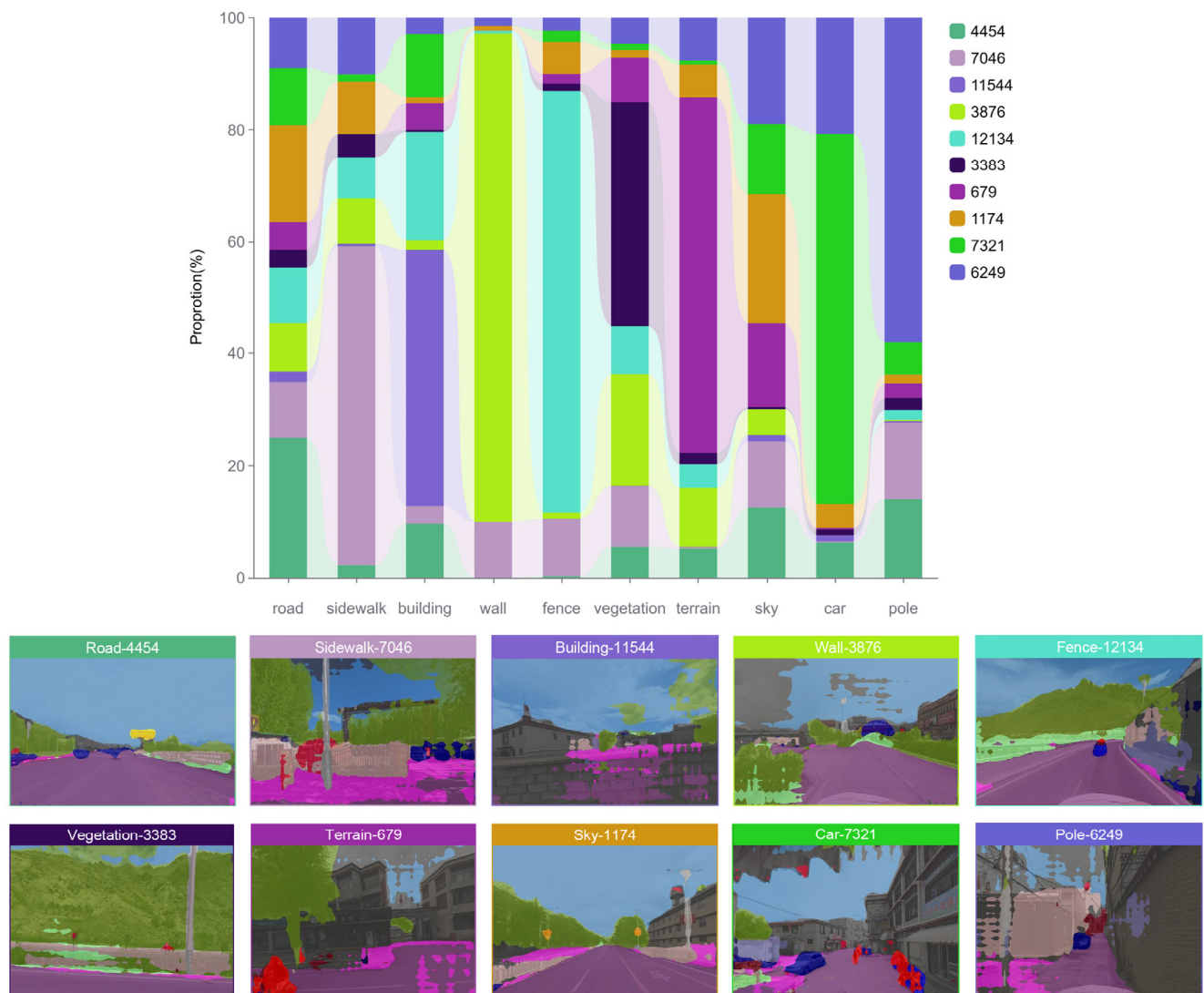


Figure 8. The maximum proportion of each visual element in the top ten urban areas in central Lhasa.

To gain a deeper insight into the spatial distribution characteristics of street visual elements in the central area of Lhasa, this study evaluates four distinct zones. The visual elements of building and vegetation incrementally increase from Zone A to Zone D. Zone A, located near the city boundaries, comprises primarily new residential developments and industrial zones. Zone B serves as the central business area, featuring traditional commercial streets and shopping centers. Zone C is characterized by the presence of the Potala Palace, a significant tourist attraction. Zone D, the historic city center, showcases Dazhao Monastery and Barkhor Street, which exhibit traditional Tibetan street features. Due to the emphasis on Dazhao Monastery and Barkhor Street, Zone D has the largest proportion of road and sky, reflecting Tibetan architectural designs. In contrast, Zone C has a reduced number of visual elements because city planners have restricted building heights to complement the Potala Palace. Zones A and B, being nearer to the suburban areas, are less densely populated with buildings and visual elements. Zones B and C exhibit more topographical variations, whereas Zone D is dominated by walls and roads.

4.2. Street Quality Analysis Based on Six Perceptions

This study examined the perceptual score from 50 volunteers across various dimensions, as illustrated in Figures 9–11, to assess perceptions of the Lhasa city center. The negative dimensions, ‘boring’ (69.70) and ‘depressing’ perceptions (67.76), received the

highest score, whereas the positive dimensions, ‘beautiful’ (60.66) and ‘wealthy’ perceptions (59.91), followed closely. The dimensions of ‘lively’ (56.68) and ‘safe’ perceptions (50.64) were rated the lowest. The central streets of Lhasa City are predominantly characterized by tourist attractions and traditional Tibetan residences, inhabited by a mix of Tibetan locals, tourists, and commercial activities. The lively and safe perceptions were more favorable in the eastern sections and certain central streets, especially those adjacent to Dzungjiao Lukang Park. These streets are enveloped by iconic landmarks such as the Potala Palace and Dazhao Monastery, contributing to a vibrant atmosphere. Additionally, the scores for boring and depressing perceptions were significantly low in Lhasa’s city center. High boring scores in Lhasa are predominantly observed along the city’s principal roads, suggesting that the planning and construction of these pathways overlooked the incorporation of regional characteristics intended to create landscapes with distinct Tibetan features. Rather, there was an emphasis on a homogeneous and standardized approach, which culminated in increased boring perception. On the other hand, the lower score regarding depressing perception might be linked to the relatively older infrastructure of some primary roads in the city center and their expansive design, which does not significantly affect residents in a negative manner.

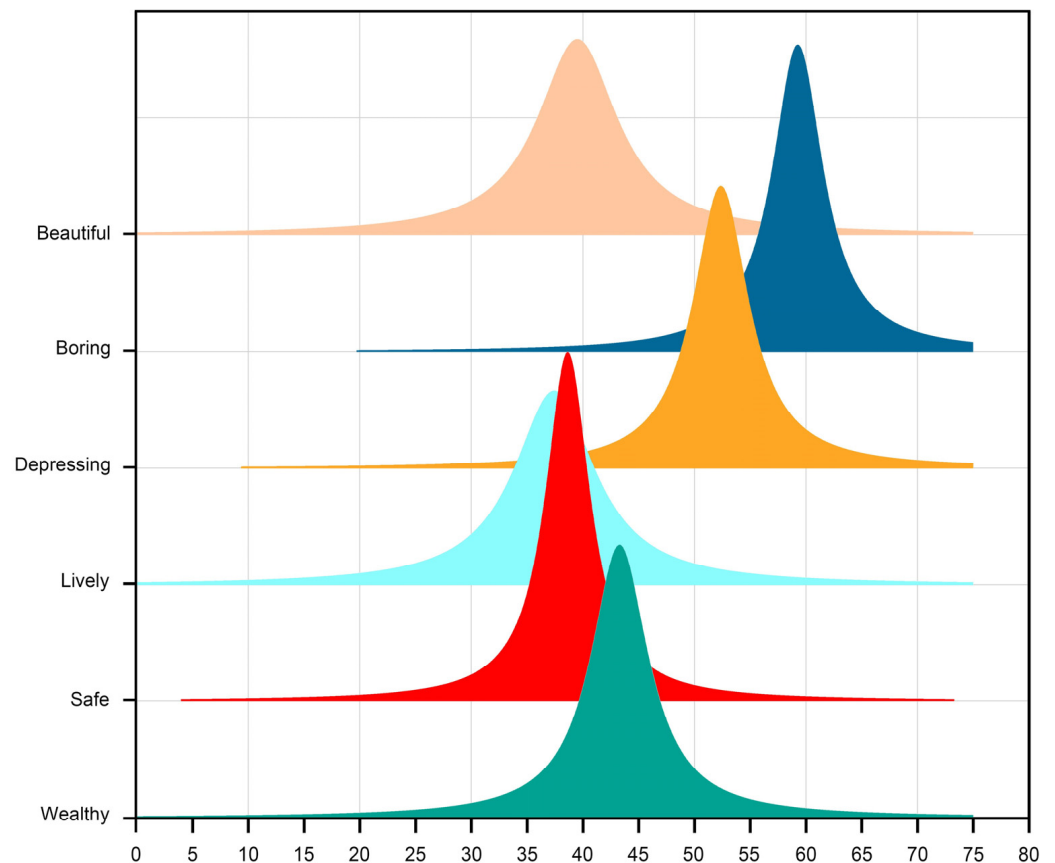


Figure 9. Six types of perception scores in Lhasa’s central urban area.

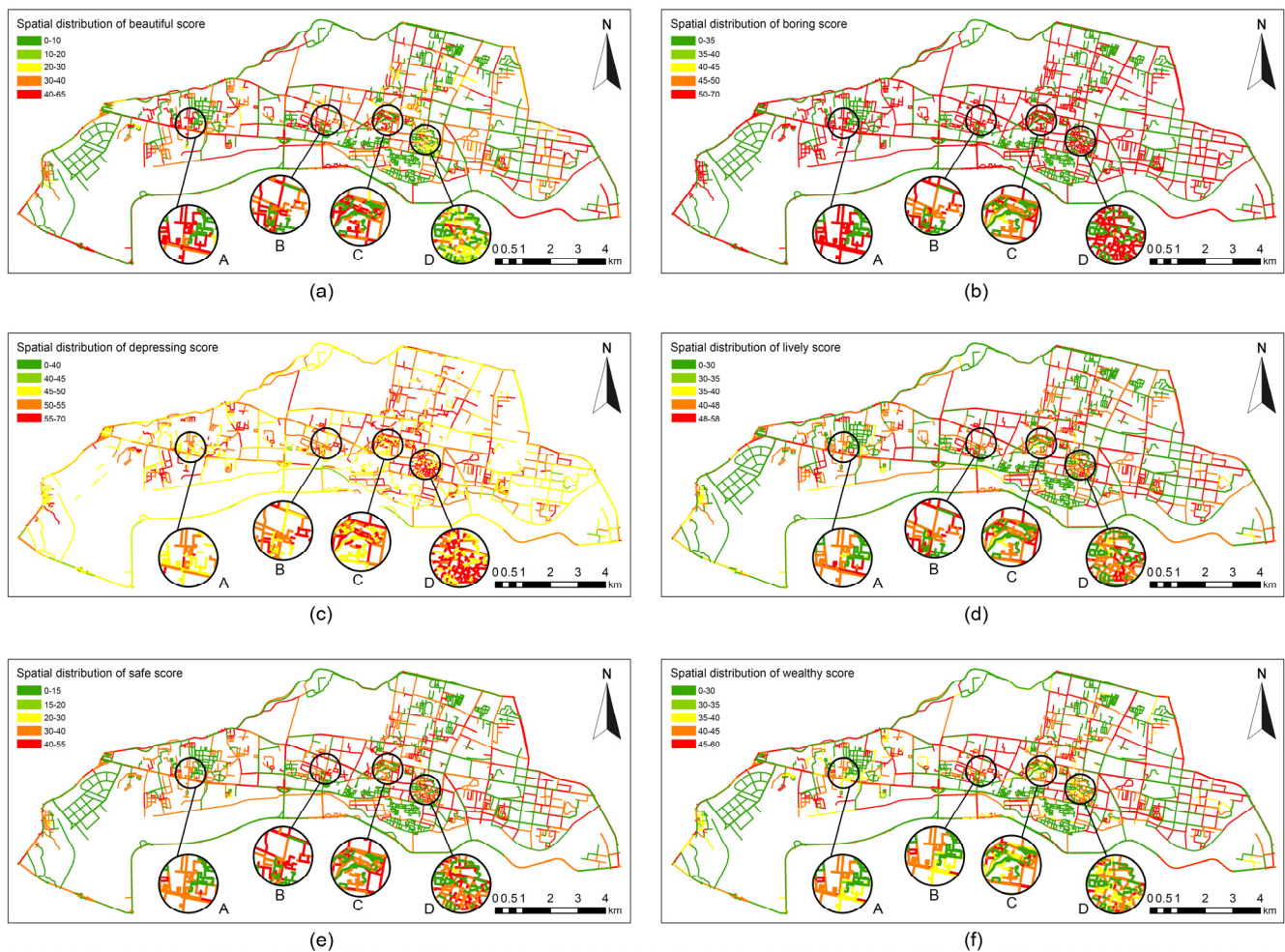


Figure 10. Spatial distribution of six perception scores in Lhasa's central urban area: (a) spatial distribution of beautiful perception, (b) spatial distribution of boring perception, (c) spatial distribution of depressing perception, (d) spatial distribution of lively perception, (e) spatial distribution of safe perception, (f) spatial distribution of wealthy perception. In the selected area, A represents new residential developments and industrial zones, B represents the central business area, C represents the Potala Palace Scenic Area, and D represents the historic city center, showcases Dazhao Monastery and Barkhor Street.

To analyze the spatial distribution of the six perceptions of streets in Lhasa's city center more clearly, this study thoroughly examines the four sub-districts with unique features. Zone A, located on the city outskirts with mainly residential neighborhoods, ranks the highest in terms of beauty and wealth. Zone B, serving as the city's commercial center, features a mix of traditional and modern streets. Zone C is home to the Potala Palace scenic area, with minimal surrounding streets to preserve the landscape. As a result, Zone B and Zone C scored lower compared to Zone A. Zone D, housing the Da Zhao Monastery and Barkhor Street, displays distinct Tibetan street characteristics but scored the lowest for the beautiful and wealthy perceptions due to narrow streets surrounding the old city center. In terms of lively and safe perceptions, Zone B scored the highest as it is situated in the commercial district, showcasing how a suitable scale of commercial streets can enhance street vitality and people's sense of security. On the other hand, Zone D scored highest for boring and depressing perceptions, attributed to its location in an old urban area with narrow, environmentally unfavorable streets. While Zones A, B, and C share similar spatial characteristics, the highest scores are primarily concentrated on the main city streets.

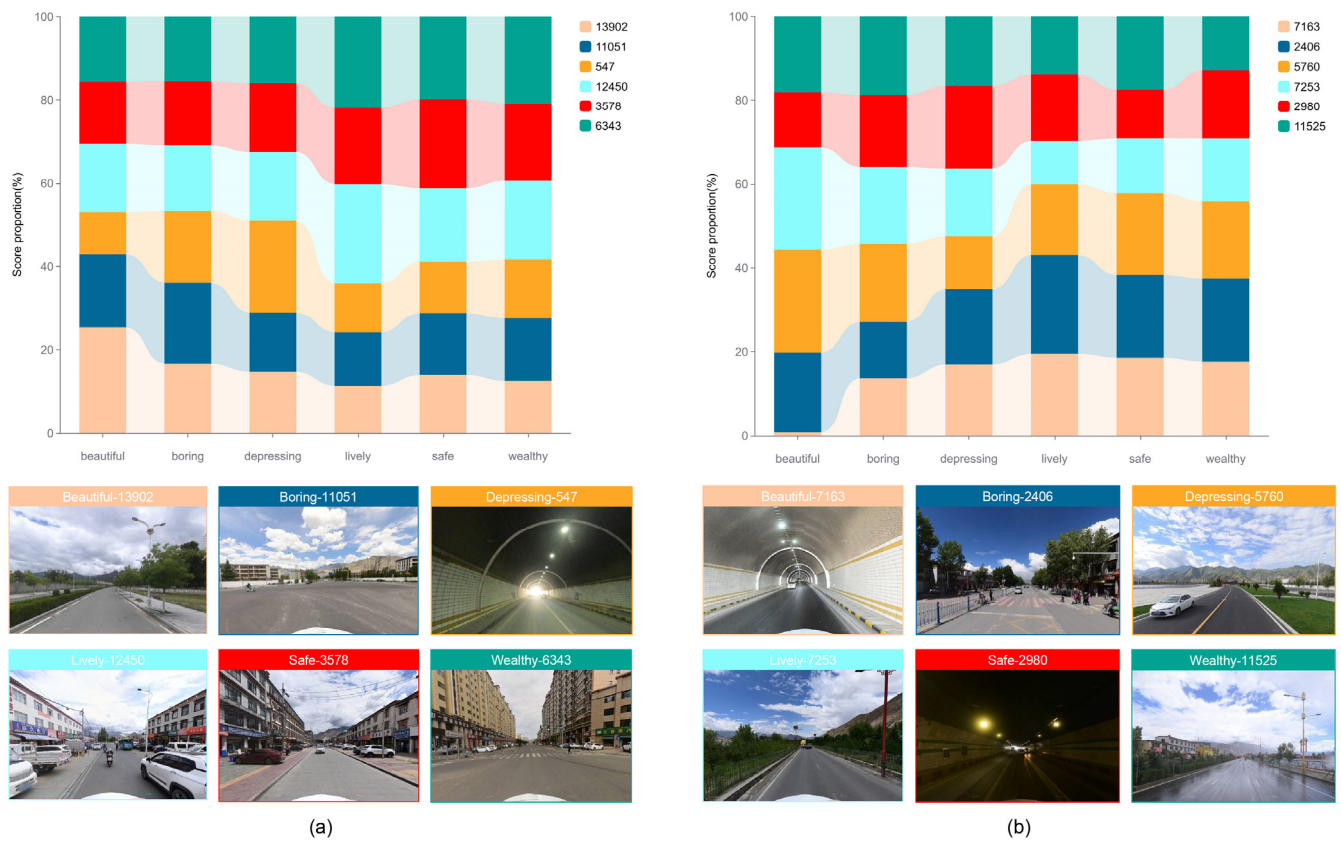


Figure 11. Maximum and minimum value of each perception score in the top ten urban areas in central Lhasa: (a) six-dimensional perception maximum, (b) six-dimensional perception minimum.

4.3. Linear Regression Analysis of Visual Elements and Six Perceptions in Street View Images

In this analysis, multiple linear regression was employed to explore the relationship between the ten primary visual components found in urban landscapes and the six facets of perception within the central area of Lhasa. The presence of a positive correlation signifies a beneficial impact, while a negative correlation indicates a detrimental influence. The data depicted in Figure 12 demonstrate that building aspects exhibit a positive effect (0.030) on the aesthetic aspect but a negative impact (-0.030) on the diversity dimension. This contrast can be traced back to the presence of traditional Tibetan design in the aesthetic aspect, in contrast to the absence of Tibetan attributes in the modern residential areas, which leads to the negative effect in the diversity dimension, subsequently impacting the feelings of safety and liveliness. The subpar quality of features like walls, vegetation, terrain, and cars contributes to decreased perceptions of both safety and diversity among inhabitants, resulting in undesirable outcomes. Concerning adverse spatial reception, roads (-0.094), sidewalks (-0.031), fences (-0.036), terrain (-0.020), sky (-0.098), cars (-0.016), and poles (-0.075) significantly impact the feeling of monotony in a negative manner, while the other visual elements positively influence boredom. Conversely, roads (0.066), sidewalks (0.008), sky (0.078), cars (0.049), and poles (0.028) have a considerable positive impact on boring perception, whereas buildings (-0.009), walls (-0.072), fences (-0.002), vegetation (-0.067), and terrain (-0.018) exhibit a notable negative effect on depressing perception.

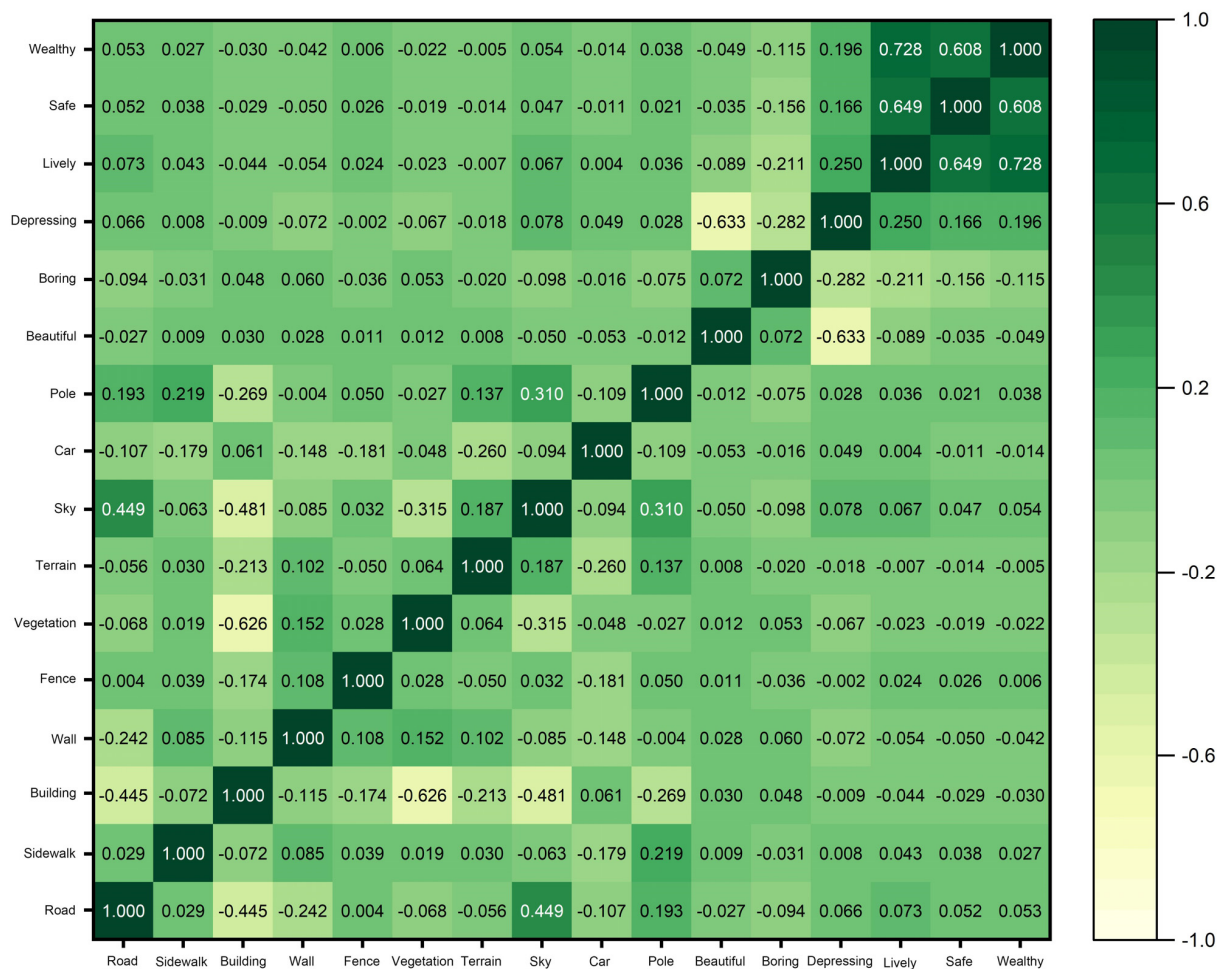


Figure 12. Multiple linear regression analysis results of visual element proportion and perception score.

4.4. Combination Analysis of Street Visual Elements and Six Perceptions

This research analyzes how individuals perceive street space in Lhasa's central city from different perspectives. The goal is to offer valuable insights to urban planners for designing future urban streets in underdeveloped ethnic regions. Streets with the greatest visual elements were categorized as high-visual-element-percentage streets, and those with the top perception score for each aspect were chosen as representatives. High-visual-element-percentage streets were then paired with high-scoring streets throughout all research areas. Areas that scored highly on the lively perception showcase landscape features that enhance the space's energy, promoting a feeling of liveliness and motion. Regions with a strong sense of opulence exhibit elaborate structures, street components, and high-end landscape attributes. On the other hand, regions with a heightened sense of beauty display stunning landscapes, while those with a diminished sense of beauty present dull landscapes and unvarying street designs.

Drawing from previous studies on the six perceptions of streets in the city center of Lhasa, this research revealed that the dimension of boredom received the highest score of 69.70. Figure 13 illustrates the top 20% of streets with high boring scores overlaid with the top 20% of shares of the top ten visual elements for further examination. The overlay highlights specific areas of interest in contrasting colors to denote streets with elevated perception scores and visual element shares. Notably, elements such as vegetation, sidewalks, and poles, which contribute to the boring perception, were primarily concentrated on main streets. Conversely, buildings, fences, and cars led to a higher boring perception score due to their prevalence in the old town and side streets, which generally have a less

favorable urban environment. Terrain and walls displayed similar spatial distribution characteristics with a consistent scatter layout. In contrast, the negative perception of the sky was predominantly observed on secondary streets and small alleys, potentially influenced by the density and height of buildings in these areas.

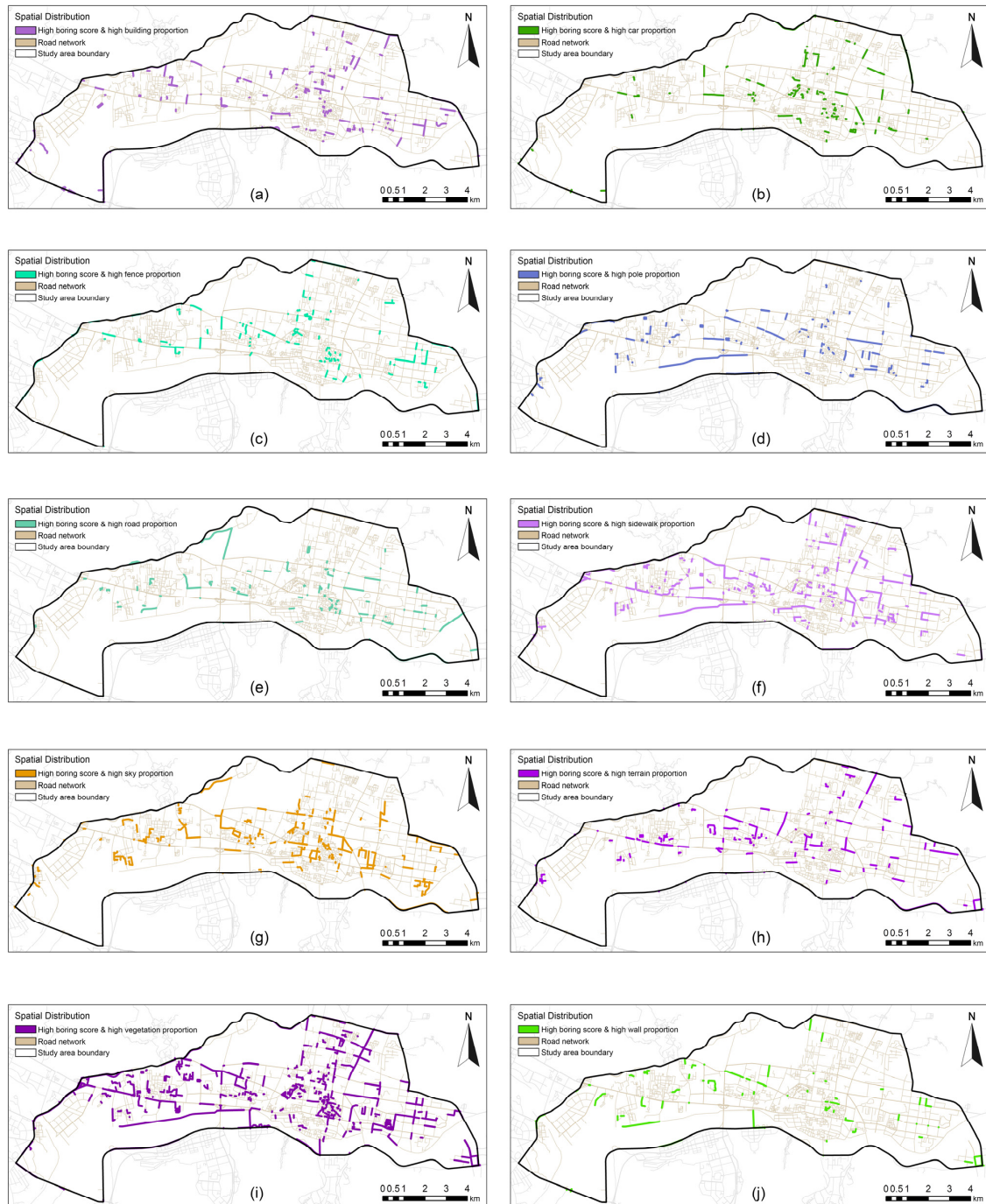


Figure 13. Spatial distribution of high boredom perception (top 20%) and high visual elements (top 20%): (a) high boring score and high building proportion, (b) high boring score and high car proportion, (c) high boring score and high fence proportion, (d) high boring score and high pole proportion, (e) high boring score and high road proportion, (f) high boring score and high sidewalk proportion, (g) high boring score and high sky proportion, (h) high boring score and high terrain proportion, (i) high boring score and high vegetation proportion, (j) high boring score and high wall proportion.

Figure 14 demonstrates the overlay of the highest 20% of vegetation scores with the top 20% scores for each of the six perceptions. The spatial distribution of vegetation displays similarities in beautiful, wealthy, safe, and lively perceptions, showcasing a consistent layout pattern. Some primary roads also display unique attributes. Nevertheless, concerning the boring perception, the dense concentration of vegetation is situated centrally within the historic city, along with other primary roads. This phenomenon can be ascribed to the compact buildings in the old city and the singular alignment of vegetation along primary roads. Additionally, the feeling of melancholy derived from vegetation is predominantly observed in the central historic town and the northern region, where streets are tightly packed and densely distributed. This indicates that a well-designed street layout and vegetation configuration can have a positive impact on the emotional perceptions of the street.

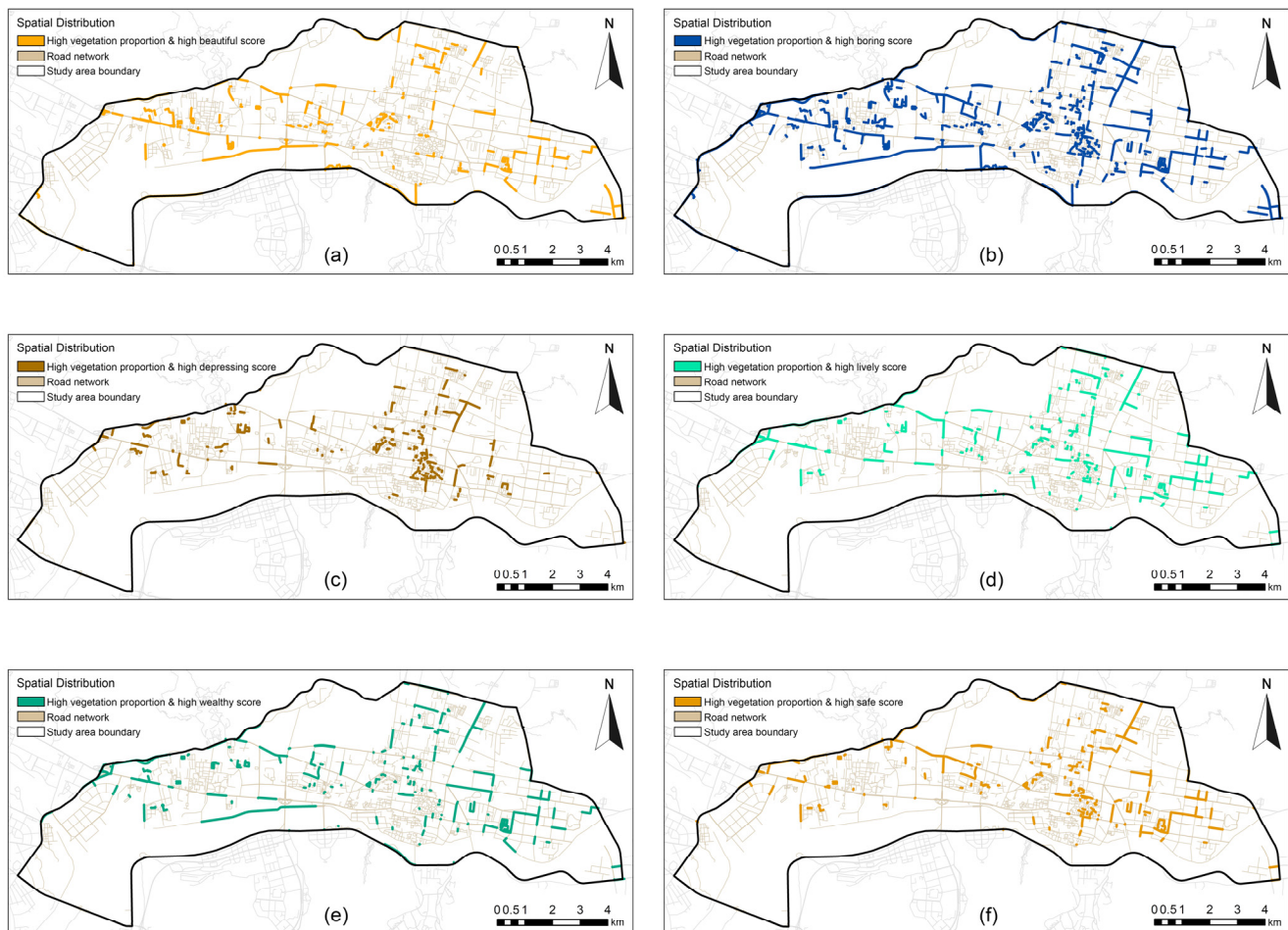


Figure 14. Spatial distribution of high vegetation percentage (top 20%) and high perception score (top 20%): (a) high vegetation proportion and high beautiful score, (b) high vegetation proportion and high boring score, (c) high vegetation proportion and high depressing score, (d) high vegetation proportion and high lively score, (e) high vegetation proportion and high wealthy score, (f) high vegetation proportion and high safe score.

Figures 15 and 16 show various street views comparing levels of boring perception and visual element presence. The results suggest that boring perception is heavily influenced by the overall street environment quality. For example, the perception of visual elements like cars (20.74%), sidewalks (11.18%), sky (57.02%), and fences (24.04%) is closely tied to their placement on major streets. Conversely, roads (30.50%), walls (19.38%), and poles (3.19%) show higher boredom levels in suburban settings. Vegetation (89.51%) is linked

to the historic city center, while the terrain (16.29%) is more impacted by overpasses. In Figures 17 and 18, scenes with abundant greenery and respective perception scores are depicted. The findings reveal that while greenery is generally seen positively, it can also adversely affect street perception. Scores such as boring perception (59.91) are noticeable in tall building neighborhoods, lively perception (58.46) in urban commercial areas, beautiful perception (61.13) in city parks, safe perception (55.55) in older urban neighborhood side streets, and boring perception (69.68) and depressing perception (67.26) on main thoroughfares in the city center.

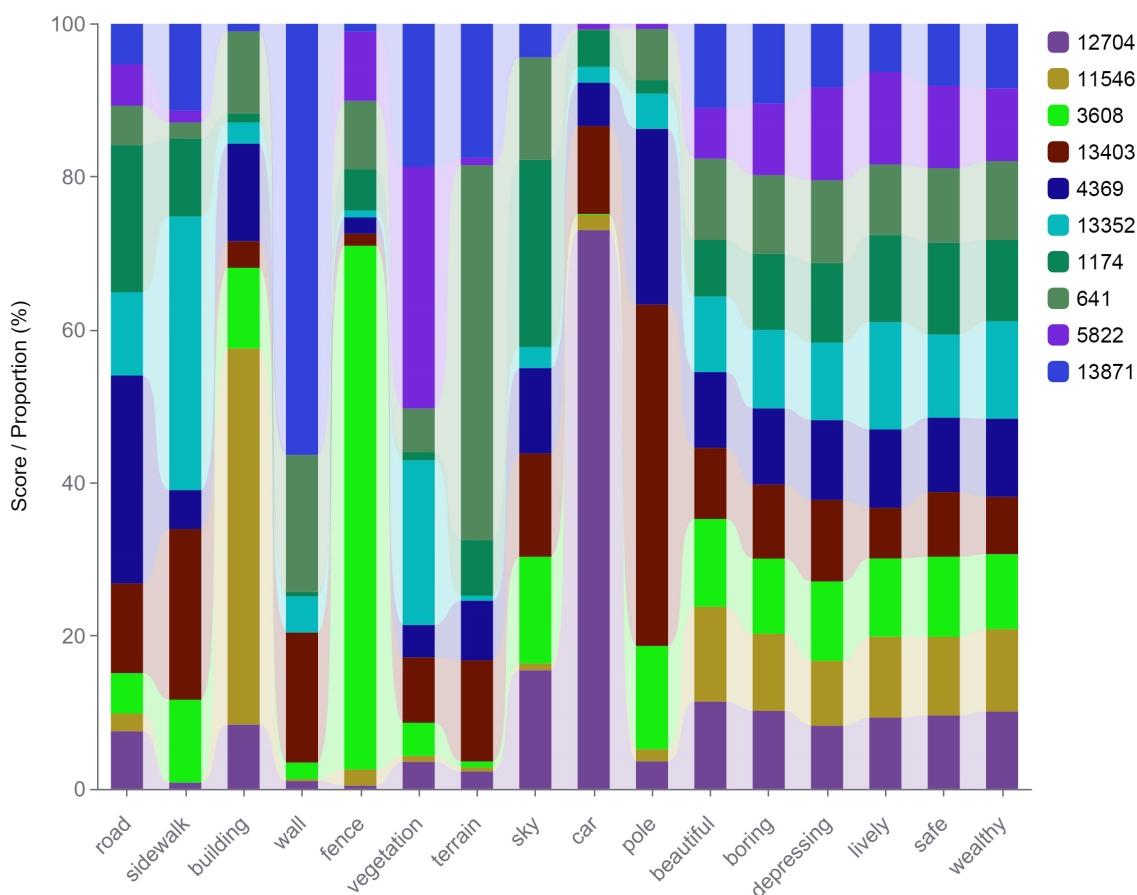


Figure 15. Representative street view relationship distribution of high boring perception (top 20%) and high visual elements (top 20%).

To enhance the analysis of street distribution, this research focuses on the top 20% of regions with high visual element share and perception scores for spatial overlay analysis. Detailed examination is conducted on four distinct regions (See Figure 19). Zone A, situated on the city outskirts, and Zone D display similar distribution patterns, with an equal spread of high visual element share and perception scores. Zone D is bordered by the Da Zhao Monastery and the traditional Tibetan commercial street (Barkhor Street), consisting of numerous alleys, urban villages, a blend of tourists and locals, and intricate traffic and pedestrian movement. In contrast, Zones B and C showcase a distribution highlighted by high perception scores. Zone B is a residential district in Lhasa with consistent building styles and a pleasant street ambiance. Zone C encompasses attractions like the Potala Palace and parks like Dzongkok Lukang Park, underlining the significance of humanistic landscapes in enriching the overall experience for visitors and residents.

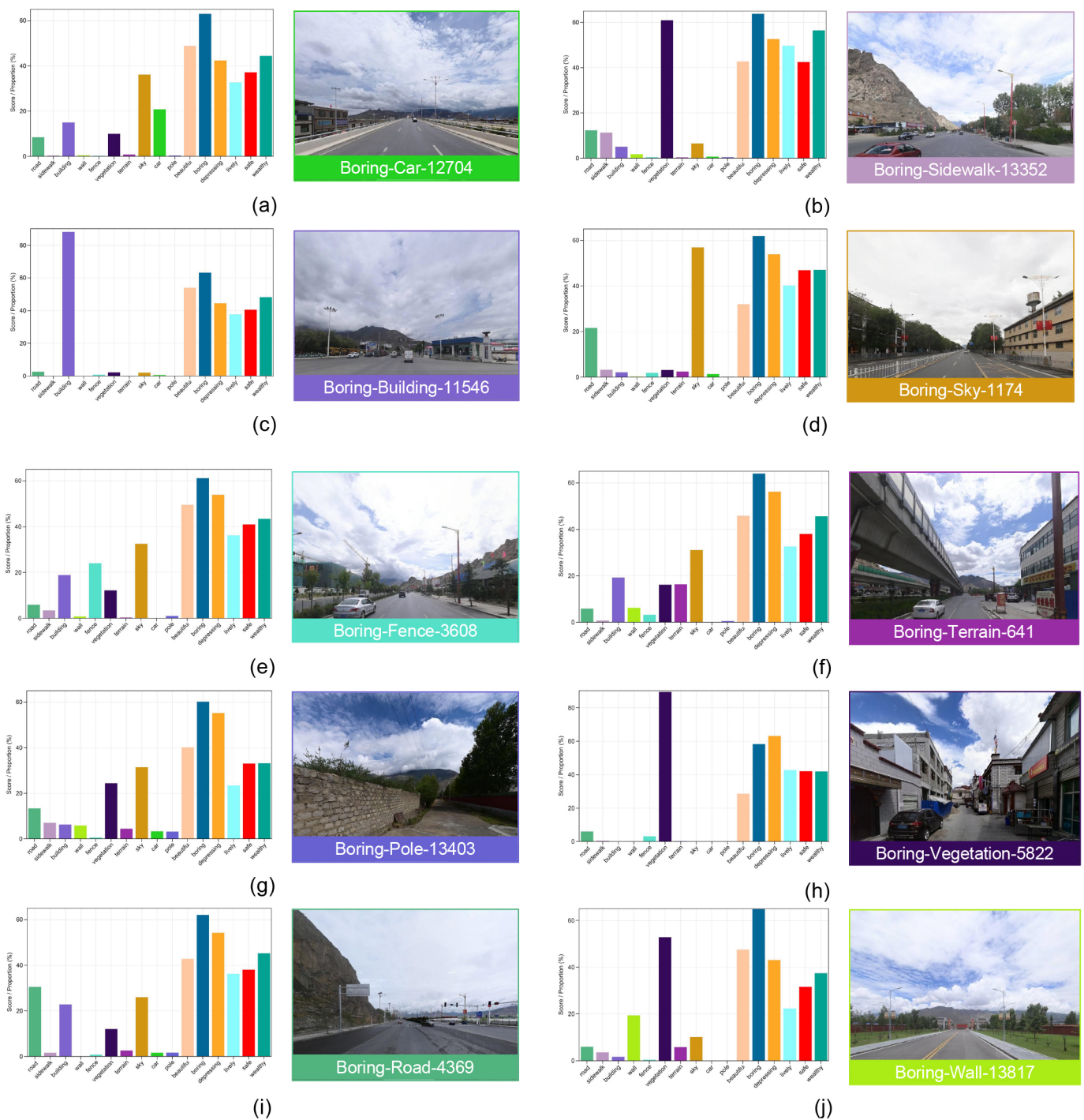


Figure 16. Representative street views with high boring perception (top 20%) and high visual elements (top 20%): (a) high boring score and high car proportion, (b) high boring score and high sidewalk proportion, (c) high boring score and high building proportion, (d) high boring score and high sky proportion, (e) high boring score and high fence proportion, (f) high boring score and high terrain proportion, (g) high boring score and high pole proportion, (h) high boring score and high vegetation proportion, (i) high boring score and high road proportion, (j) high boring score and high wall proportion.

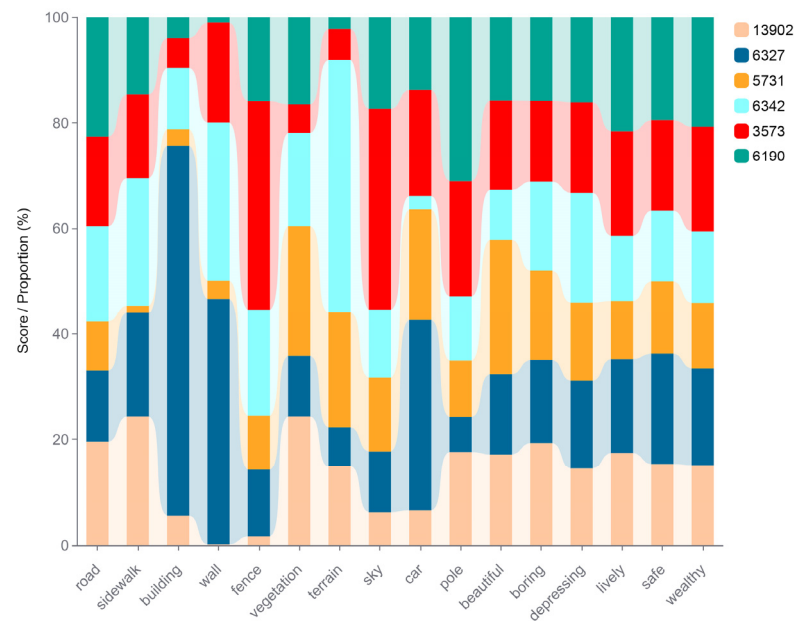


Figure 17. Representative street view relationship distribution between high vegetation proportion (top 20%) and high perception score (top 20%).

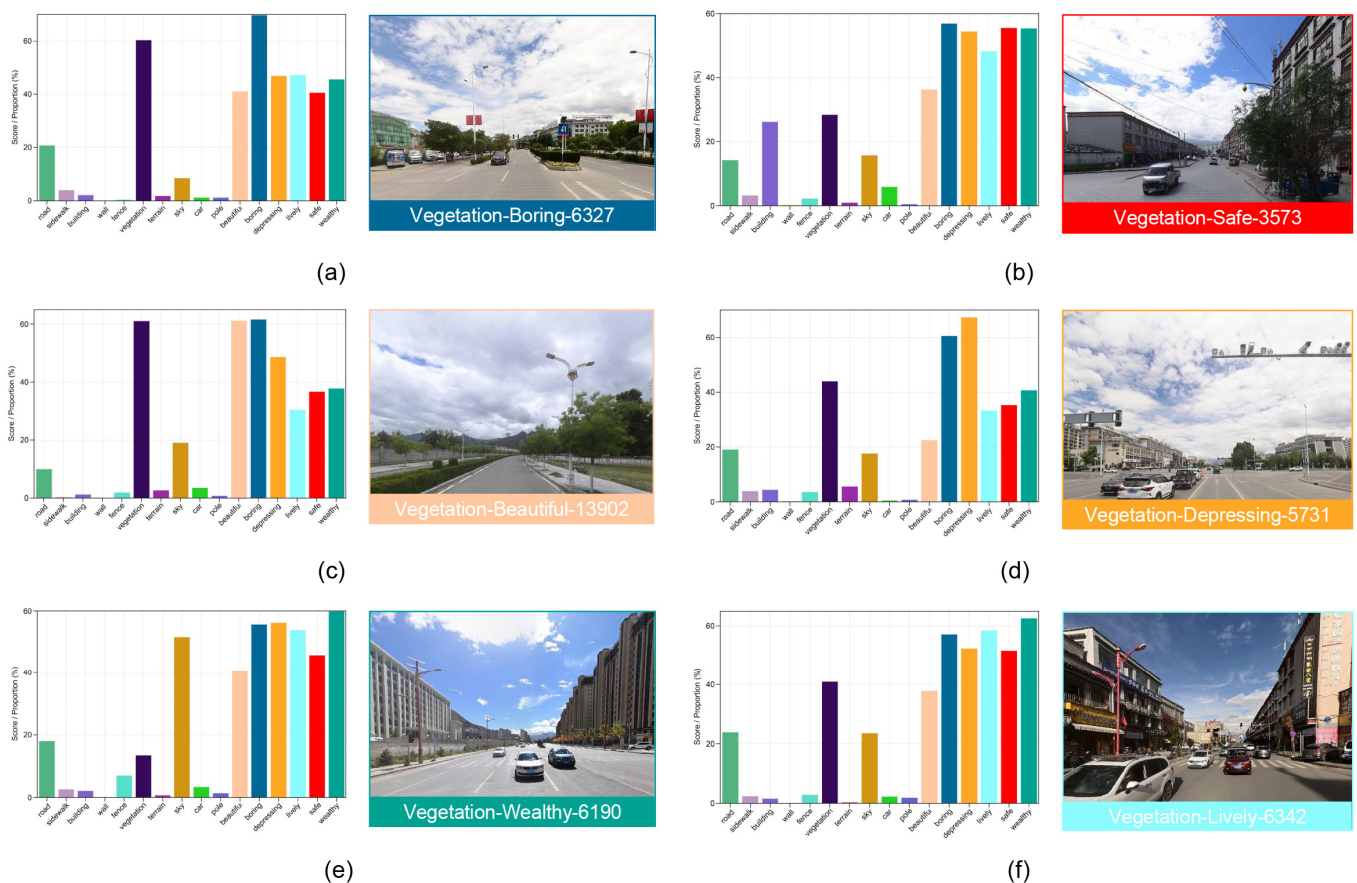


Figure 18. Representative street views with high vegetation proportion (top 20%) and high perception score (top 20%): (a) high vegetation proportion and high boring score, (b) high vegetation proportion and high safe score, (c) high vegetation proportion and high beautiful score, (d) high vegetation proportion and high depressing score, (e) high vegetation proportion and high wealthy score, (f) high vegetation proportion and high lively score.

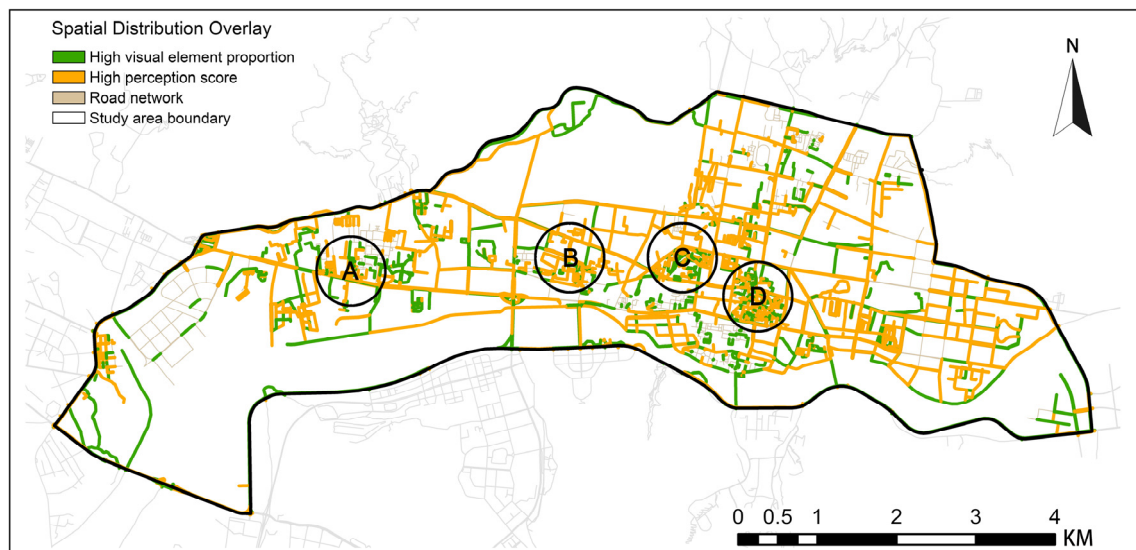


Figure 19. Spatial distribution of high visual element proportion and high perception score. In the selected area, A represents new residential developments and industrial zones, B represents the central business area, C represents the Potala Palace Scenic Area, and D represents the historic city center, showcases Dazhao Monastery and Barkhor Street.

5. Discussion

5.1. The Influence of Visual Elements on the Perception of Street Quality

This research aims to evaluate the perceptual assessment of streets within Lhasa's city center to identify the overlap between areas with a high share of visual elements and elevated spatial perception scores. The results provide vital insights for planning urban street renewals in underdeveloped ethnic regions. Various perceptual dimensions emerge within Lhasa's city center: high-rise neighborhoods are wealthy in perception, business districts exude vitality, city parks are perceived as beautiful, and secondary roads in the old city are felt to be safe, while city center main roads convey boring and depression perceptions. Elements like roads, sidewalks, fences, terrain, sky, cars, and poles significantly impact feelings of boring perception in the old city area and suburbs. Essential visual components such as roads, vegetation, buildings, sky, and sidewalks are critical for enhancing street quality in Lhasa's city center. Future initiatives should focus on these elements to improve street conditions. Moreover, the combination of buildings, vegetation, and roads largely contributes to the monotonous and dreary street atmosphere along Lhasa's central main roads. To improve existing street conditions, the Lhasa municipal government can diversify vegetation, incorporate unique Tibetan architectural features, and enhance the quality of roads, sidewalks, building height, and regional traits for future planning. These findings offer a valuable reference point for improving street quality in other underdeveloped ethnic areas.

5.2. Implications for Urban Development Policy Practices in Underdeveloped Ethnic Areas

The current research also examines the implications of urban policy and design practices, improving the street space quality and assisting city managers in gaining a more comprehensive insight into urban streets. Through the combination of deep learning techniques and overlay analysis of visual elements and perceptions, a more efficient and accurate understanding of the urban landscape in underdeveloped ethnic regions can be attained. This method can support the construction and advancement of cities in these areas, providing scientific backing for city planners. Evaluating the spatial quality of streets in the central city of Lhasa by analyzing a high percentage of visual elements and spatial perception score not only reinforces street perception in underdeveloped ethnic regions but also lays the groundwork for future city development and enhancement of street user

experiences. Additionally, the incorporation of deep learning methods in this study offers a way to assess urban street quality perceptions in underdeveloped ethnic areas, aligning with the real experiences of urban users in these regions. This promotes the improvement of street quality in underdeveloped ethnic regions while preserving the unique cultural characteristics of the Tibetan people.

5.3. Scientific Contribution of Research Methods

In recent years, there has been a surge in interest in analyzing street space perception using street view data within the urban domain. While many studies have employed various methods to gather perception data and assess urban street space quality, the focus has mainly been on developed cities, overlooking underdeveloped ethnic neighborhoods. This research presents a significant scientific advancement by introducing a comprehensive technique to evaluate the perception of urban street space quality in these marginalized ethnic areas through the utilization of deep learning. By integrating street visual components with six perception variables, this approach pinpoints crucial aspects of urban street perception in these regions, offering important benchmarks for evaluating spatial quality and pushing forward research in this area. The developed analytical framework enables the measurement of urban street usage frequency and high street spatial perception in underdeveloped ethnic neighborhoods while also investigating the influence of visual elements on urban street perception. The results of this investigation hold promise for stakeholders in underdeveloped ethnic regions, providing valuable insights for sustainable street planning and urban infrastructure development and introducing a new angle for spatial analysis of urban streets in these locations.

5.4. Research Limitations

This study delves into the perception of urban streets in underdeveloped ethnic areas, highlighting certain issues and gaps that warrant further exploration in future research. By overlaying visual elements and six perceptions, the study aims to pinpoint specific areas within urban street space in underdeveloped ethnic areas that require attention. However, the reliance on a singular type of research data poses challenges in gaining comprehensive insights into the characteristics of the built environment and the evolution of urban street quality in underdeveloped ethnic areas. This limitation hinders the provision of a multi-level explanation for the origins of urban perception outcomes in such areas. To address these limitations, it is crucial to encompass a broader range of underdeveloped ethnic regions, such as Shannan and Linzhi City, in order to evaluate the primary features of visual elements influencing urban street perception in these areas. Furthermore, transcending the constraints of the time dimension by utilizing historical images of BSVIs and Google Street View images can offer valuable insights into the historical progression of urban street quality in underdeveloped ethnic areas. This approach can aid researchers in urban science and anthropology in gaining a deeper understanding of how the built environment influences human perception across different temporal contexts.

6. Conclusions

This research utilizes deep learning methods to evaluate the perception of urban street quality in economically disadvantaged ethnic regions, with a particular emphasis on Lhasa's city center. In Lhasa's city center, the urban area demonstrates a high percentage of visual elements, especially in terms of buildings and vegetation. Specifically, buildings accounted for 88.23% and vegetation for 89.52%, whereas poles represented the lowest percentage at 3.14%. Other visual elements, including roads (31.10%), sky (56.27%), terrain (16.27%), and walls (19.49%), showed similar spatial characteristics, largely concentrated along the main road. Out of the six perceptions evaluated, boring (69.70) and depressing (67.76) perceptions registered the highest scores, followed by beautiful (60.66) and wealthy perceptions (59.91). Lively (56.68) and safe perceptions (50.64) received the lowest scores.

In the analysis of the correlation between visual components and perceptions in architectural design, the results indicated that architecture positively impacted the beautiful perception (0.030) while negatively affecting the wealthy perception (−0.030). In terms of unfavorable spatial impressions, negative effects on boring perception were noted for roads (−0.094), sidewalks (−0.031), fences (−0.036), terrain (−0.020), sky (−0.098), cars (−0.016), and poles (−0.075). Conversely, other visual elements demonstrated a favorable impact on the boring perception. Moreover, roads (0.066), sidewalks (0.008), sky (0.078), cars (0.049), and poles (0.028) were associated with a positive effect on depressing perception. Conversely, buildings (−0.009), walls (−0.072), fences (−0.002), vegetation (−0.067), and terrain (−0.018) exhibited a negative impact on the depressing perception.

The high levels of boring perception are chiefly influenced by the quality of the street environment. Visual aspects such as cars (20.74%), sidewalks (11.18%), sky (57.02%), and fences (24.04%) are closely associated with main thoroughfares. Suburban areas significantly impact the boring perception of roads (30.50%), walls (19.38%), and poles (3.19%). Vegetation (89.51%) is linked with historic towns, whereas terrain (16.29%) is more subject to viaducts. Interestingly, although vegetation is generally a positive visual element, it can also negatively affect street perception. Wealthy perception (59.91%) is predominantly observed in high-rise neighborhoods, while lively perception (58.46%) is more concentrated in urban commercial districts. Beautiful perception (61.13%) is prevalent in urban park areas, safe perception (55.55%) is focused on secondary streets in older urban districts, and boring perception (69.68%) and depressing perception (67.26%) are mainly encountered on main roads in central city areas.

Author Contributions: C.L. and Y.Y. participated in all phases; X.Y. helped with data collection and paper organization. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Natural Science Foundation of China (Grant No. 52178059) for the project titled ‘Research on the evolution mechanism and planning methods of the spatial structure of Tibetan towns along the Sichuan-Tibet Railway’.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author on reasonable request. The data are not publicly available due to privacy policies.

Conflicts of Interest: The authors declare no potential conflicts of interest with respect to the authorship and/or publication of this article.

Nomenclature

FCN
 ResNet
 SegNet
 AlphaGo
 Cityscapes Dataset
 Green View Index (GVI)
 MIT Place Pulse program
 Google Street View (GSV)
 Street View Imagery (SVI)
 Urban Renewal Projects (URPs)
 Baidu Street View images (BSVIs)
 DeeplabV3+ semantic segmentation
 Polarimetric Synthetic Aperture Radar (PolSAR)

References

1. Cao, S.; Lv, Y.; Zheng, H.; Wang, X. Challenges facing China's unbalanced urbanization strategy. *Land Use Policy* **2014**, *39*, 412–415. [\[CrossRef\]](#)
2. Drobnjaković, M.; Panić, M.; Đorđević, J. Traditional undeveloped municipalities in Serbia as a result of regional inequality. *Eur. Plan. Stud.* **2016**, *24*, 926–949. [\[CrossRef\]](#)
3. Hutárová, D.; Kozelová, I.; Špulerová, J. Tourism development options in marginal and less-favored regions: A case study of Slovakia's Gemer Region. *Land* **2021**, *10*, 229. [\[CrossRef\]](#)
4. Fang, C. A theoretical analysis of the mechanism and evolutionary law of urban-rural integration development. *Acta Geogr. Sin.* **2022**, *77*, 759–776.
5. Huang, Z.; Lu, L.; Su, Q.; Zhang, J.; Sun, J.; Wan, X.; Jin, C. Research and development of rural tourism under the background of new urbanization: Theoretical reflection and breakthrough of predicament. *Geogr. Res.* **2015**, *34*, 1409–1421.
6. Pang, X.; Wang, R.; Wang, W. Research on coupling of tourism and urbanization in underdevelopment regions: A case study of Fusong county in Jilin Province. *Geogr. Geo-Inf. Sci.* **2014**, *30*, 130–134.
7. Matlovičová, K.; Kolesárová, J.; Demková, M.; Kostilníková, K.; Mocák, P.; Pachura, P.; Payne, M. Stimulating poverty alleviation by developing tourism in marginalised Roma communities: A case study of the Central Spis Region (Slovakia). *Land* **2022**, *11*, 1689. [\[CrossRef\]](#)
8. Zhou, R.; Yu, Y.; Wu, B.; Luo, X. Quantitative evaluation of urban resilience in underdeveloped regions: A study of six cities in Sichuan & Tibet, China. *Front. Environ. Sci.* **2022**, *11*, 1266487.
9. Zhang, P.; Zhang, L.; Han, D.; Wang, T.; Zhu, H.; Chen, Y. Coupled and Coordinated Development of the Tourism Industry and Urbanization in Marginal and underdeveloped Regions—Taking the Mountainous Border Areas of Western Yunnan as a Case Study. *Land* **2023**, *12*, 640. [\[CrossRef\]](#)
10. Yu, Y.; Zhou, R.; Qian, L.; Yang, X.; Dong, L.; Zhang, G. Supply-demand balance and spatial distribution optimization of primary care facilities in highland cities from a resilience perspective: A study of Lhasa, China. *Front. Public Health* **2023**, *11*, 1131895. [\[CrossRef\]](#)
11. Abusaada, H.; Elshater, A. Effect of people on placemaking and affective atmospheres in city streets. *Ain Shams Eng. J.* **2021**, *12*, 3389–3403. [\[CrossRef\]](#)
12. Tang, J.; Long, Y. Measuring visual quality of street space and its temporal variation: Methodology and its application in the Hutong area in Beijing. *Landsc. Urban Plan.* **2019**, *191*, 103436. [\[CrossRef\]](#)
13. McGinn, A.; Evenson, K.; Herring, A.; Huston, S.; Rodriguez, D. Exploring associations between physical activity and perceived and objective measures of the built environment. *J. Urban Health* **2007**, *84*, 162–184. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Li, X.; Zhang, C.; Li, W.; Ricard, R.; Meng, Q.; Zhang, W. Assessing street-level urban vegetation using Google Street View and a modified green view index. *Urban For. Urban Green.* **2015**, *14*, 675–685. [\[CrossRef\]](#)
15. Zhou, H.; He, S.; Cai, Y.; Wang, M.; Su, S. Social inequalities in street visual walkability: Using street view imagery and deep learning technologies to facilitate healthy city planning. *Sustain. Cities Soc.* **2019**, *50*, 101605. [\[CrossRef\]](#)
16. Wang, R.; Lu, Y.; Zhang, J.; Liu, P.; Yao, Y.; Liu, Y. The relationship between visual enclosure for neighbourhood street walkability and elders' mental health in China: Using street view images. *J. Transp. Health* **2019**, *13*, 90–102. [\[CrossRef\]](#)
17. Ma, X.; Ma, C.; Wu, C.; Xi, Y.; Yang, R.; Peng, N.; Zhang, C.; Ren, F. Measuring human perceptions of street views to better inform urban renewal: A perspective of scene semantic parsing. *Cities* **2021**, *110*, 103086. [\[CrossRef\]](#)
18. Gonzalez, D.; Rueda-Plata, D.; Acevedo, A.; Duque, J.; Ramos-Poll'an, R.; Betancourt, A.; García, S. Automatic detection of building typology using deep learning methods on street level images. *Build. Environ.* **2020**, *177*, 106805. [\[CrossRef\]](#)
19. Nagata, S.; Nakaya, T.; Hanibuchi, T.; Amagasa, S.; Kikuchi, H.; Inoue, S. Objective scoring of street view walkability related to leisure walking: Statistical modeling approach with semantic segmentation of Google Street View images. *Health Place* **2020**, *66*, 102428. [\[CrossRef\]](#)
20. Hu, C.; Zhang, F.; Gong, F.; Ratti, C.; Li, X. Classification and mapping of urban canyon geometry using Google Street View images and deep multitask learning. *Build. Environ.* **2020**, *167*, 106424. [\[CrossRef\]](#)
21. Li, H.; Páez, A.; Liu, D. Built environment and violent crime: An environmental audit approach using Google Street View. *Comput. Environ. Urban Syst.* **2017**, *66*, 83–95.
22. Zhang, F.; Zhou, B.L.; Liu, L.; Liu, Y.; Fung, H.; Lin, H.; Ratti, C. Measuring human perceptions of a large-scale urban region using machine learning. *Landsc. Urban Plan.* **2018**, *180*, 148–160. [\[CrossRef\]](#)
23. Zhao, C.; Ogawa, Y.; Chen, S.L.; Oki, T.; Sekimoto, Y. Quantitative land price analysis via computer vision from street view images. *Eng. Appl. Artif. Intell.* **2023**, *123*, 106294. [\[CrossRef\]](#)
24. Rui, J. Measuring street view perceptions from driveways and sidewalks to inform pedestrian-oriented street renewal in Dusseldorf. *Cities* **2023**, *141*, 104472. [\[CrossRef\]](#)
25. Wang, L.; Han, X.; He, J.; Taeyeol, J. Measuring residents' perceptions of city streets to inform better street planning through deep learning and space syntax. *ISPRS J. Photogramm. Remote. Sens.* **2022**, *190*, 215–230. [\[CrossRef\]](#)
26. Liang, Y.; Zeng, J.; Li, S. Examining the Spatial Variations of Land Use Change and Its Impact Factors in a Coastal Area in Vietnam. *Land* **2022**, *11*, 1751. [\[CrossRef\]](#)
27. Liang, Y.; Zeng, J.; Sun, W.; Zhou, K.; Zhou, Z. Expansion of construction land along the motorway in rapidly developing areas in Cambodia. *Land Use Policy* **2021**, *109*, 105691. [\[CrossRef\]](#)

28. Hossam, H.; Johanna, L. The death and life of Malmi neighbourhood shopping street: Is ethnic retail a catalyst for public life recovery in Helsinki? *Eur. Plan. Stud.* **2021**, *30*, 336–358.
29. Abozied, E.; Vialard, A. Reintegrating informal settlements into the Greater Cairo Region of Egypt through the regional highway network. *Reg. Stud. Reg. Sci.* **2020**, *7*, 333–345. [\[CrossRef\]](#)
30. Zhou, B.; Liu, L.; Oliva, A.; Torralba, A.; Fleet, D.; Pajdla, T.; Schiele, B.; Tuytelaars, T. Recognizing city identity via attribute analysis of geo-tagged images. In Proceedings of the 13th European Conference on Computer Vision (ECCV), Zurich, Switzerland, 6–12 September 2014; pp. 519–534.
31. Xu, G.; Zhu, X.; Fu, D.J.; Dong, J.W.; Xiao, X.M. Automatic land cover classification of geo-tagged field photos by deep learning. *Environ. Model. Softw.* **2017**, *91*, 127–134. [\[CrossRef\]](#)
32. Hu, Y.J.; Gao, S.; Janowicz, K.; Yu, B.L.; Li, W.W.; Prasad, S. Extracting and understanding urban areas of interest using geotagged photos. *Comput. Environ. Urban Syst.* **2015**, *54*, 240–254. [\[CrossRef\]](#)
33. Ulrich, R.S. Visual landscapes and psychological well-being. *Landsc. Res.* **1979**, *4*, 17–23. [\[CrossRef\]](#)
34. Anguelov, D.; Dulong, C.; Filip, D.; Frueh, C.; Lafon, S.; Lyon, R.; Ogale, A.; Vincent, L.; Weaver, J. Google street view: Capturing the world at street level. *Computer* **2010**, *43*, 32–38. [\[CrossRef\]](#)
35. Cheng, L.; Chu, S.S.; Zong, W.W.; Li, S.Y.; Li, M.C. Use of tencent street view imagery for visual perception of streets. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 265. [\[CrossRef\]](#)
36. Gebru, T.; Krause, J.; Wang, Y.L.; Chen, D.Y.; Deng, J.; Aiden, E.L.; Li, F.F. Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 13108–13113. [\[CrossRef\]](#)
37. Zhang, F.; Hu, M.Y.; Lin, H.; Fang, C.Y. Framework for virtual cognitive experiment in virtual geographic environments. *SPRS Int. J. Geo-Inf.* **2018**, *7*, 36. [\[CrossRef\]](#)
38. Salesses, P.; Schechtner, K.; Hidalgo, C.A. The collaborative image of the city: Mapping the inequality of urban perception. *PLoS ONE* **2013**, *8*, e68400. [\[CrossRef\]](#)
39. Dubey, A.; Naik, N.; Parikh, D.; Raskar, R.; Hidalgo, C.A. Deep learning the city: Quantifying urban perception at a global scale. In Proceedings of the 14th European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 11–14 October 2016; pp. 196–212.
40. Ordonez, V.; Berg, T.L. Learning high-level judgments of urban perception. In Proceedings of the 13th European Conference on Computer Vision (ECCV), Zurich, Switzerland, 6–12 September 2014; pp. 494–510.
41. Porzi, L.; Bulò, S.R.; Lepri, B.; Ricci, E. Predicting and understanding urban perception with convolutional neural networks. In Proceedings of the ACM International Conference on Multimedia (ACM Multimedia), Brisbane, Australia, 26–30 October 2015; pp. 139–148.
42. Naik, N.; Kominers, S.D.; Raskar, R.; Glaeser, E.L.; Hidalgo, C.A. Computer vision uncovers predictors of physical urban change. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 7571–7576. [\[CrossRef\]](#)
43. Yin, L. Street level urban design qualities for walkability: Combining 2D and 3D GIS measures. *Comput. Environ. Urban Syst.* **2017**, *64*, 288–296. [\[CrossRef\]](#)
44. Low, S.; Altman, I. (Eds.) Place attachment: A conceptual inquiry. In *Human Behavior & Environment: Advances in Theory & Research*; Plenum Press: New York, NY, USA, 1992; Volume 12, pp. 1–12.
45. Hidalgo, M.; Hernandez, B. Place attachment: Conceptual and empirical questions. *J. Environ. Psychol.* **2001**, *21*, 273–281. [\[CrossRef\]](#)
46. Andrew, L.; Gu, X.; Chen, L.; Perry, H. Predicting perceptions of the built environment using GIS, satellite and street view image approaches. *Landsc. Urban Plan.* **2021**, *216*, 104257.
47. Liu, J.; Li, T.; Xie, P.; Du, S.; Teng, F.; Yang, X. Urban big data fusion based on deep learning: An overview. *Inf. Fusion* **2020**, *53*, 123–133. [\[CrossRef\]](#)
48. He, T.; Li, X. Image quality recognition technology based on deep learning. *J. Vis. Commun. Image Represent.* **2019**, *65*, 102654. [\[CrossRef\]](#)
49. He, N.; Li, G. Urban neighbourhood environment assessment based on street view image processing: A review of research trends. *Environ. Chall.* **2021**, *4*, 50090. [\[CrossRef\]](#)
50. Wang, R.; Liu, Y.; Lu, Y.; Zhang, J.; Liu, P.; Yao, Y.; Grekousis, G. Perceptions of built environment and health outcomes for older Chinese in Beijing: A big data approach with street view images and deep learning technique. *Comput. Environ. Urban Syst.* **2019**, *78*, 101386. [\[CrossRef\]](#)
51. Wu, J.; Feng, Z.; Peng, Y.; Liu, Q.; He, Q. Neglected green street landscapes: A re-evaluation method of green justice. *Urban For. Urban Green.* **2019**, *41*, 344–353. [\[CrossRef\]](#)
52. Donghwan, K.; Sugie, L. Analyzing the effects of Green View Index of street streets on walking time using Google Street View and deep learning. *Landsc. Urban Plan.* **2021**, *205*, 103920.
53. Wang, M.; He, Y.; Meng, H.; Zhang, Y.; Zhu, B.; Joseph, M.; Li, X. Assessing Street Space Quality Using Street View Imagery and Function-Driven Method: The Case of Xiamen, China. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 282. [\[CrossRef\]](#)
54. William, T.; Matthew, N.; Chyi-Lin, L.; Christopher, P. Implementing a deep-learning model using Google street view to combine social and physical indicators of gentrification. *Comput. Environ. Urban Syst.* **2023**, *102*, 101970.
55. Yu, L.; Zeng, Z.; Liu, A.; Xie, X.; Wang, H.; Xu, F.; Hong, W. A Lightweight Complex-Valued DeepLabv3+ for Semantic Segmentation of PolSAR Image. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 930–943. [\[CrossRef\]](#)

56. Kim, J.; Lee, S.; Hipp, J.; Ki, D. Decoding urban landscapes: Google street view and measurement sensitivity. *Comput. Environ. Urban Syst.* **2021**, *88*, 101626. [[CrossRef](#)]
57. Rzotkiewicz, A.; Pearson, A.L.; Dougherty, B.V.; Shortridge, A.; Wilson, N. Systematic review of the use of Google Street View in health research: Major themes, strengths, weaknesses and possibilities for future research. *Health Place* **2018**, *52*, 240–246. [[CrossRef](#)]
58. Yao, Y.; Liang, Z.; Yuan, Z.; Liu, P.; Bie, Y.; Zhang, J.; Wang, R.; Wang, J.; Guan, Q. A human-machine confrontational scoring framework for urban perception assessment using street-view images. *Int. J. Geogr. Inf. Sci.* **2019**, *33*, 2363–2384. [[CrossRef](#)]
59. Li, Y.; Miller, H.J.; Root, E.D.; Hyder, A.; Liu, D.S. Understanding the Role of Urban Social and Physical Environment in Opioid Overdose Events Using Found Geospatial Data. *Health Place* **2022**, *75*, 102792. [[CrossRef](#)]
60. Zhang, G.; Zu, J.; Hu, M.; Zhu, D.; Kang, Y.; Gao, S.; Zhang, Y.; Huang, Z. Uncovering Inconspicuous Places Using Social Media Check-Ins and Street View Images. *Comput. Environ. Urban Syst.* **2020**, *81*, 101478. [[CrossRef](#)]
61. Fernández-Delgado, M.; Cernadas, E.; Barro, S.; Amorim, D. Do we need hundreds of classifiers to solve real world classification problems. *J. Mach. Learn. Res.* **2014**, *15*, 3133–3181.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.