Proceedings of the 2nd International Conference on Computing Innovation and Applied Physics
DOI: 10.54254/2753-8818/18/20230375

Analysis of pressure perception of internal roads in existing settlements based on random forest/deep learning

Siying Li[1,*], Yuntian Me[2], Jinrui Wu[3]

[1]Department of statistics, Zhongnan University of Economics and law, Wuhan, China
[2]Department of Big Data, Shanghai University of Finance and Economic, Shanghai, China
[3]Department of Architecture, Hunan University of Science and Technology,Xiangtan, China

*202123090439@stu.zuel.edu.cn

Abstract. The quality of road space has an influential role in human psychological health. Quantifying residents' perceptions of road spatial quality and exploring the relationship between perceptions and constituent elements have become topics of interest to many parties. In this study, based on the high-resolution remote sensing images acquired by Baidu Maps, the road images of built-up settlements in Beijing in three different periods (1949-1977, 1978-1997, and 1998-2022) were semantically segmented. The large-scale rating data and subjective evaluations of residents' spatial perceptions of roads were analyzed using random forest and multiple linear regression methods. The results show that stress, a negative perception, is positively correlated with buildings and cars, and negatively correlated with vegetation, sky, and sidewalks. The composition of elements affecting residents' stress perceptions in communities of different build-up years was consistent, but the degree of influence of each element was ranked differently. This study attempts to summarize the road characteristics and residential advantages of different built-up neighborhoods, and proposes optimization strategies for upgrading road space in old neighborhoods and countermeasure suggestions for future improvement of the human living environment.

Keywords: Established Settlements, Roadscape, Pressure Perception, Semantic Segmentation, Random Forest.

1. Introduction

According to China's 2022 National Health Insight Report, people will have an average of 4.8 health problems in 2021. The first and foremost of these are negative emotional problems such as anxiety and depression. Greater psychological stress has become a more common social phenomenon. In the post-epidemic era, people are paying more attention to their health and the quality of their living environment. Studies have shown that high quality streets encourage residents to be emotionally active [1] and help to reduce personal stress [2]. Settlements have an irreplaceable and positive role as major life vehicles in terms of leisure, communication and fitness.

DOI: 10.54254/2753-8818/18/20230375

Since the beginning of modern society, the mature cities have preserved many types of settlements in different periods. Existing settlements are large residential areas with relatively independent living environment and complete set of living services in a certain area of the city. In recent years, the quality of the internal environment of settlements has become an important factor for people to consider when purchasing or renting a house. At the same time, as people's demand increases and the city expands and changes its functions, the original residential buildings are facing the problem of reconstruction or renovation. Regardless of the choice, the design concept of "human-centeredness" has been the consensus of the design community. The study of the quality and vitality of street space is an important cornerstone for the improvement of street habitat [3]. Streetscape is often used to measure the safety and emotional perception of street space [4][5]. In recent years, some researchers have proposed "virtual environmental audits" of built environments in cities based on Google Street View data, and virtual audits are increasingly used in the evaluation of urban pedestrian environments because of their efficiency, low cost, and wide applicability [6]. This study combines machine learning and linear regression to explore the connection between the spatial perception of roads and road composition elements in settlements of different built-up eras. As well as the design development trends of settlements in different eras. It is beneficial to scientifically summarize the design strategies for the transformation of old settlements and the environmental improvement of rural building clusters.

2. Methodology

2.1. Research framework

This study attempts to analyze the residential stress experience of residents in different built-up areas with the help of streetscape image maps, random forest models and linear regression models. In turn, it explores the design points in urban renewal. To achieve this goal, the study is divided into four main phases as follows: in the first phase, a program is written to access the Baidu Street View Map Application Programming Interface (API) to collect the internal road images of each of the three built-up residential neighborhoods; in the second phase, the FCN image semantic segmentation model is used to semantically segment the images and organize the percentage of different elements in the images; in the third phase, a manual evaluation is introduced to The pressure perception degree of different streetscape photos is scored and the related subjective evaluations are counted; in the fourth stage, the road characteristics of three different built-up era residential neighborhoods are distinguished and the characteristics and change trends of elements affecting road pressure perception are analyzed in depth.

2.2. Study area classification and data sources

2.2.1. Study area classification. In this study, three phases of completed settlements in Beijing from 1945-1978, 1978-1998, and 1998-2022 were selected, which are hereafter referred to as Time 1, Time 2, and Time 3.

After the founding of New China in 1945, which ended China's colonial and semi-colonial rule and unified management of construction planning and other industries, a large number of residential settlements began to be built in Beijing, and most of the housing in this period was unit compounds, military housing, and silos. The silos were mainly built by factories, schools, etc. to solve the problem of tight housing for employees. The silos had long corridors in the middle and ribbed rooms lined up on the north and south sides. The residents shared the kitchen and bathroom.

The reform and opening up in 1978, the birth of the real estate industry, residential areas towards the development of commercialization of the road. During that period, most of the buildings were 6-story slab buildings with no elevators.

In 1998, the commercialization of housing reform began in Beijing, China, and welfare housing was completely ended. The Chinese central government set the reform direction of marketization, monetization and commercialization of urban housing. Housing consumption and development scale

DOI: 10.54254/2753-8818/18/20230375

increased rapidly. It is becoming popular to build high-rise slab and tower buildings, and the spacing between buildings is becoming larger for fire protection needs.

The rapid development of China's real estate industry as of 2022 has resulted in a more mature and complete design workflow. The existing old neighborhoods in the Beijing area are distinctive and in need of renovation.

2.2.2. Data sources. Download Beijing built-up area boundaries and Beijing road network shp format files through Baidu Map API. The app "Anjuke", which is widely used by Chinese people to buy houses, was used to obtain information such as the number of floors and the year of construction. The information obtained was given to the settlement boundaries through the spatial jion tool in ArcGIS and exported in ArcGIS to classify the settlements in different years of construction in Beijing.

Based on ArcGIS platform, under the settlement layer of each built-up period, the settlement boundary roads are deleted and only the internal road network is retained. Each road was randomly sampled and exported to a csv file with a minimum distance of 20 meters between the two sampling points. The file contains information such as settlement name, latitude and longitude of the sampling points. Write a program to access the Baidu Street View Map Application Programming Interface (API) to collect street view images of the roads within the established settlements.

2.3. Research Tools

2.3.1. FCN image semantic segmentation model. Semantic segmentation of images is performed using an FCN network, which is a deep learning full convolutional network trained by Prof. Qingfeng Guan's team at the School of Information Engineering, China University of Geosciences (Wuhan) based on the ADE_20K dataset [8]. The model has a high accuracy for image segmentation. It can identify the elements in statistical images extremely occupied.

2.3.2. Random forest model. Random forest (RFF) is a classical bagging model with a weak learner as a decision tree model. Firstly, different street photos are manually scored for stress perception. In this study, 20 volunteers were recruited to score the stress perception of the road, 700 pictures in each era and 2100 pictures in total. Among them, 50% were male and 50% were female. The maximum age was 76 years old, the minimum age was 8 years old, and the average age was 35.2 years old. Stress perception has six score choices from 0-5, with higher scores indicating greater and more pronounced stress perception, more depressed, anxious, depressed, etc.; lower scores indicating less stress perception, more pleasant, relaxed, etc.

The data obtained from the semantic segmentation were summarized and organized into a total data set. The data from each of the three chronological periods were then modeled in random forest using chronological periods as classification indicators and stress index scores as dependent variables. Since there are as many as 150 objects involved in this semantic segmentation, and a large proportion of the semantic segmentation objects are zero, the large sparse matrix formed by these high-dimensional data will have a negative impact on the degree of modeling fit. Using the random forest model to output the importance proportion of each factor, the factors with importance greater than 0.015 were selected, and a total of 15 independent variables were encompassed by the statistics. It is worth noting that Id_0 in the semantic segmentation, the meaning indicated is unidentifiable objects, and this category does not have practical significance in the statistical process, so this independent variable is removed and re-normalized, and then the rest of the variables that are not in the category of importance statistics are grouped into other classes, and the output becomes a new data set, and the importance ranking is finally obtained. This process is not only a screening indicator, but also a meaningful statistical evaluation quantity.

2.3.3. Linear regression models for three built eras. The term "regression" was introduced by the British biologist and statistician Galton around 1886. The present study attempts to learn a function

DOI: 10.54254/2753-8818/18/20230375

that predicts through a linear combination of attributes. This function can be used for regression i.e. fitting with a hyperplane. This is also the most common scenario in machine learning - supervised learning: using a computer to learn a model given some data and then using it to predict new data. The expression for a linear regression model is:

$$Y = X\beta + \varepsilon$$

3. Results and Discussion

3.1. FCN model image semantic segmentation results

These three Tables present the several high-ranking visual elements from three different times that had the largest impact on the perception of street pressure in terms of their percentage of coverage within the images. Evidently, the proportion of building element during all of times is maximum , the tree, sky, road occupies a large proportion in these three times when the path, person, floor occupies a lower proportion. Detailly, the building is occupies about 30 percent in these three times, the proportion of tree, sky, car, road, sidewalk, wall, person are similar. Significantly, the plant's occupation is increasing, both the path like this. Particularly, the proportion of the grass is change the most, from 0.4% to 2.4%. But the proportion of the fence and the floor are appear decreasing tendency. Street greening has been used in previous studies as an influence on mental health important factor on both dimensions of mental health, with both direct and indirect on both dimensions of mental health[8]. Greenery is becoming an important consideration in the design and construction of today's plots.

Table 1. Top 14 visual elements identified following segmentation.

Number	visual elements	times	Mean	Max	Min	S.D.
1	building	TIME1	0.352956	0.894492	2.91E-05	0.162488
		TIME2	0.337014	0.829156	0	0.16957
		TIME3	0.286685	0.667996	0	0.145082
2	tree	TIME1	0.137586	0.650807	0	0.130084
		TIME2	0.120511	0.676002	0	0.137264
		TIME3	0.150319	0.681314	0	0.143096
3	sky	TIME1	0.09068	0.427622	0	0.075508
		TIME2	0.106732	0.325055	0	0.070657
		TIME3	0.108676	0.399572	0	0.071751
4	plant	TIME1	0.018833	0.34147	0	0.036914
		TIME2	0.020745	0.350579	0	0.044672
		TIME3	0.044541	0.35384	0	0.058553
5	car	TIME1	0.05904	0.340602	0	0.05918
		TIME2	0.069294	0.336784	0	0.065326
		TIME3	0.062395	0.341412	0	0.067939
6	road	TIME1	0.186524	0.427139	0	0.11231
		TIME2	0.200727	0.437188	0	0.113435
		TIME3	0.178252	0.445393	0	0.117403
7	grass	TIME1	0.004569	0.20856	0	0.016519
		TIME2	0.00654	0.213267	0	0.02149
		TIME3	0.024263	0.377309	0	0.048231
8	sidewalk	TIME1	0.079835	0.405857	0	0.079876
		TIME2	0.081231	0.394146	0	0.082186
		TIME3	0.080351	0.367577	0	0.082733

DOI: 10.54254/2753-8818/18/20230375

Table 1. (continue)

9	path	TIME1	0.002174	0.141057	0	0.011799
		TIME2	0.00638	0.341327	0	0.029818
		TIME3	0.009555	0.24626	0	0.029109
10	ground	TIME1	0.011847	0.346306	0	0.037245
		TIME2	0.011613	0.33798	0	0.037655
		TIME3	0.022374	0.411243	0	0.056824
11	wall	TIME1	0.015156	0.530161	0	0.042521
		TIME2	0.010366	0.492582	0	0.02834
		TIME3	0.009228	0.254968	0	0.022635
12	fence	TIME1	0.008187	0.183654	0	0.021859
		TIME2	0.007784	0.183566	0	0.020819
		TIME3	0.004848	0.149099	0	0.01385
13	person	TIME1	0.002016	0.07536	0	0.007005
		TIME2	0.001572	0.076897	0	0.005857
		TIME3	0.001519	0.056251	0	0.005356
14	floor	TIME1	0.009668	0.366617	0	0.035882
		TIME2	0.005579	0.24747	0	0.023344
		TIME3	0.002613	0.115758	0	0.012218

3.2. Random forest model training and prediction results

Figure 1 shows the ranking of the importance of the elements influencing the perception of road pressure in the settlements in the three periods and in general. The table shows that the types of elements influencing the perception of road pressure are relatively consistent across the three periods. The subjective observation of the streetscape pictures shows that the reason for this result is that the composition of the streetscape itself is relatively homogeneous. In most of the streetscape photos, although there are grasses and trees with vegetation of different heights, the plant species are still slightly monotonous, and the lack of plant species on the roads is more obvious in winter.

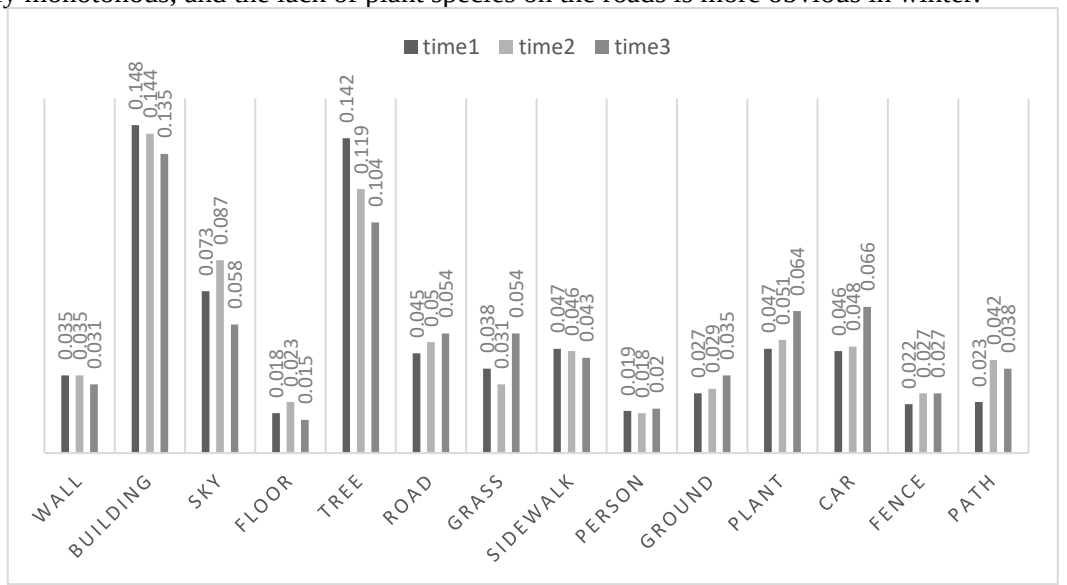

Figure 1. Importance visualization of each time.

DOI: 10.54254/2753-8818/18/20230375

Although buildings and trees have been in the primary position to influence the perception of streetscape pressure, their importance is diminishing with the development of the years. Combined with the content of the interviews with the interviewed volunteers, the appearance of the buildings was an important consideration in their ratings. Some older neighborhoods have issues such as peeling siding and uncomfortable building materials and colors that can lead testers to score low directly.

The external features of buildings can affect mental health, with Jacobs and Brown referring to street windows, continuous storefronts, etc. as the "eyes of the street" [9]. In comparison Time3 has a better level of neighborhood architectural design and more greenery in both. The influential role of buildings and greenery is decreasing and the influence of other streetscape elements is gradually increasing. It is more obvious that the influence role of vehicles is rising, in the original old neighborhood there is a lack of necessary parking spaces and parking lots, and parking is even more unplanned, but in TIme3 there is a large number of centralized parking lots with much less sidewalk parking. The probability of a dense mass of vehicles in a single picture would be such that testers would simply score it low. Thus the role of vehicle influence becomes greater.

The factors influencing the quality of streetscape from Time1 to Time3 show a multifaceted development and rise in tandem, which is the future trend of people's perception of streetscape. More plants and more beautiful buildings are not the same as better quality of space. People expect richer streetscape connotations, less obstructions and vehicles.

3.3. Three linear regression modeling

The data set was rearranged for the variables filtered by the random forest. With the ratings as the dependent variables and all remaining visual elements as independent variables, they were substituted into a multiple linear regression model to model them. The linear expressions about the ratings can be obtained.

P-Value, which is used to represent the degree of significance of the independent variable. The P-values with the degree of significance for each element have been listed in Table 2 for each period. It can be found that the variables sky, trees, grass, and vegetation, placed in all the eras, have a very significant effect on the scores. Combined with the regression coefficients, the higher the percentage of these elements in the graph, the more significant the effect on the respondents' perception of stress. What can be confirmed is that increasing the spacing between buildings and expanding the area of greenery can improve residents' stress perceptions of the streetscape. The salience of the element of roads varies in different eras. In Time 1, the significance of roads is higher, while in other eras it is not significant at all. This may be caused by the influence of urban architectural style on people's pressure perceptions in the early years of the founding of the country. The buildings in the early years of the founding of the country were mostly characterized by low height, and this is when the influence of roads on people's pressure perceptions increased, and the sidewalks below are the same reason as roads.

The regression coefficient $\hat{\beta}$ is the estimated coefficient before the respective variable of the linear model, see Figure 2. to some extent, it reflects the effect of this independent variable on the rating. The distribution of the stress index ranges from 0-5, with larger numbers representing the greater stress brought by the streetscape here to the respondents. Therefore, when the regression coefficient is smaller and obtains a negative number, it can be assumed that the visual element represented by this independent variable after the regression coefficient has a positive effect on reducing people's stress. It is very clear from the figure that the increase in the area of vegetation, grass, and trees has a positive effect on reducing the stress perception of the residents, and this is corroborated with the significance test.

DOI: 10.54254/2753-8818/18/20230375

Table 2. Significance Test Sheet.

	TIME 1		TIME 2		TIME 3	
	P-value	Significance	P-value	Significance	P-value	Significance
intercept	<2E-16	***	1.21E-07	***	5.48E-07	***
wall	0.63213		9.02E-01		4.88E-01	
building	0.298809		2.25E-01		1.26E-01	
sky	0.00000861	***	7.64E-03	**	2.34E-03	**
floor	0.9561		1.80E-01		6.34E-01	
tree	7.94E-12	***	1.19E-05	***	9.13E-05	***
road	0.01021	*	1.38E-01		1.59E-01	
grass	2.17E-11	***	1.01E-07	***	4.24E-10	***
sidewalk	0.00491	**	1.27E-01		2.52E-01	
person	0.39055		1.66E-01		4.78E-01	
ground	0.71383		6.37E-01		5.46E-02	.
plant	2.53E-10	***	1.37E-10	***	5.85E-10	***
car	0.75795		7.07E-01		9.71E-02	.
fence	0.63554		4.27E-02	*	8.74E-01	
path	0.22912		0.00895	**	0.01957	*

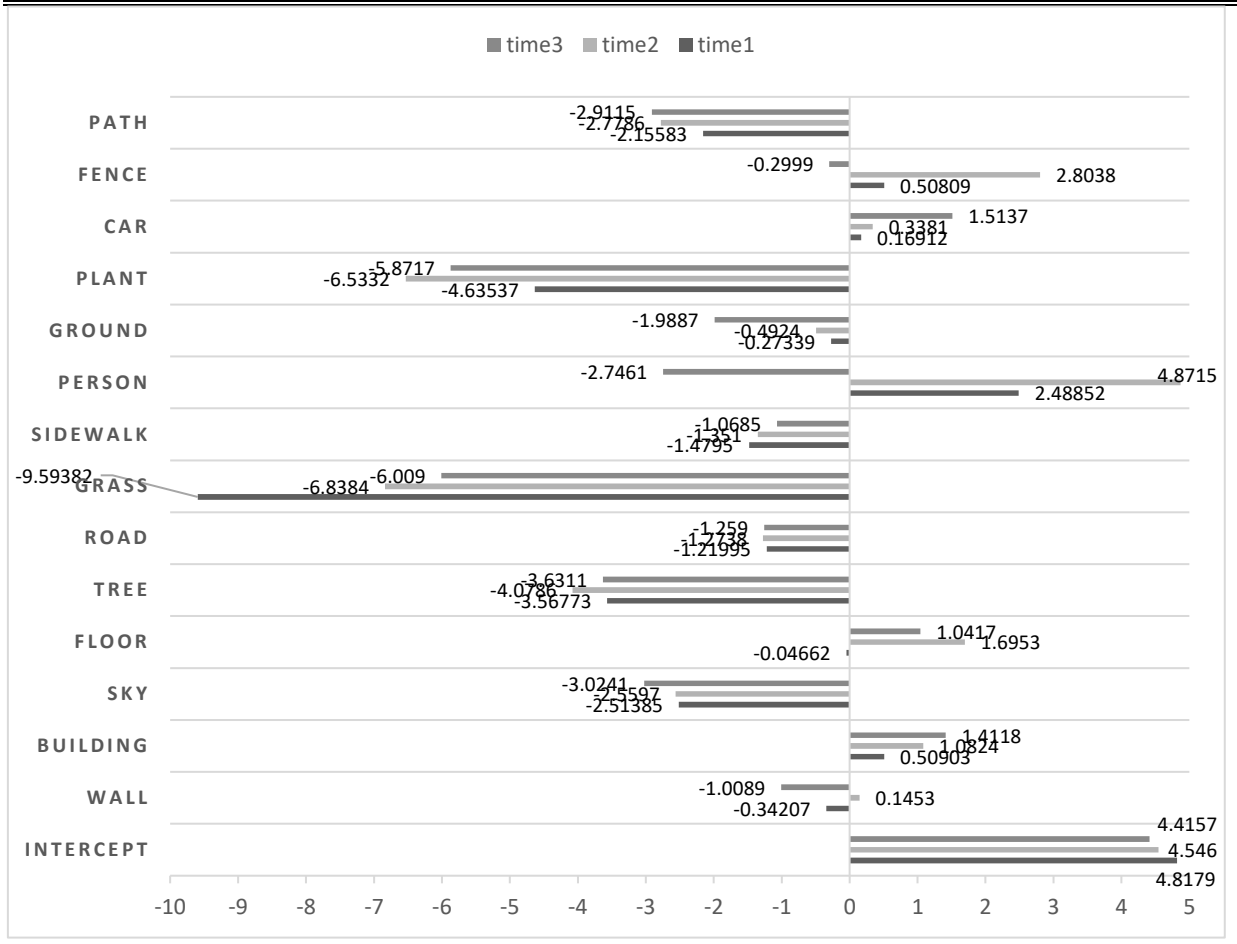

Figure 2. Beta hat of each variable.

DOI: 10.54254/2753-8818/18/20230375

3.4. Statistical tests

After building the models, this study performed the relevant statistical tests for all linear models. The table below shows the adjusted R2 values for the three periods. The goodness-of-fit performance of these three linear equations is relatively excellent.

Table 3. Adjusted R2 of Three Times.

	Time 1	Time 2	Time 3
Adjusted R2	0.6361	0.7642	0.7282

Model diagnostics are performed on these obtained linear equations. Residual plots, QQ plots, Scale-Location plots, and cookie distance plots are output, respectively.

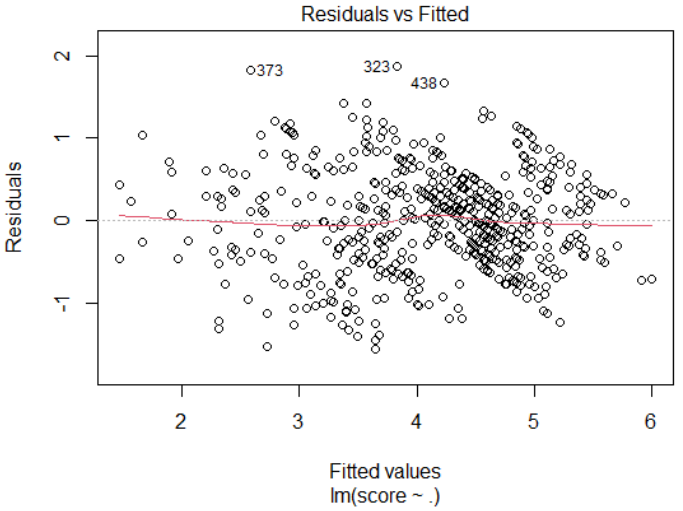

Figure 3. Fitting line.

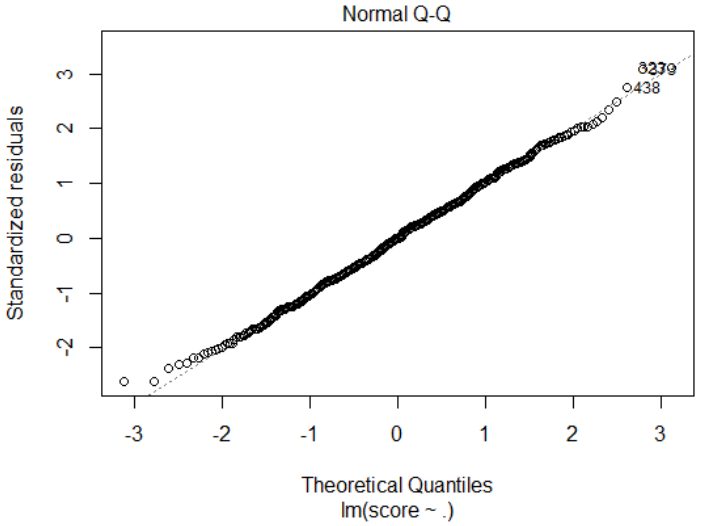

Figure 4. Normal Q-Q.

DOI: 10.54254/2753-8818/18/20230375

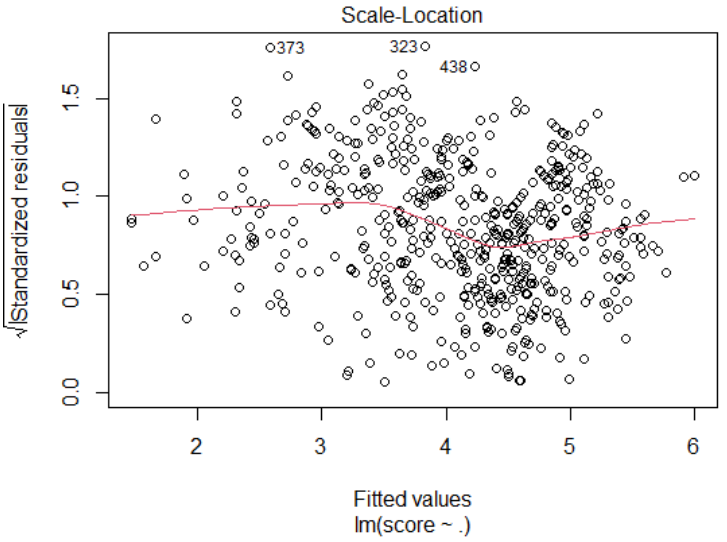

Figure 5. Scale-Location.

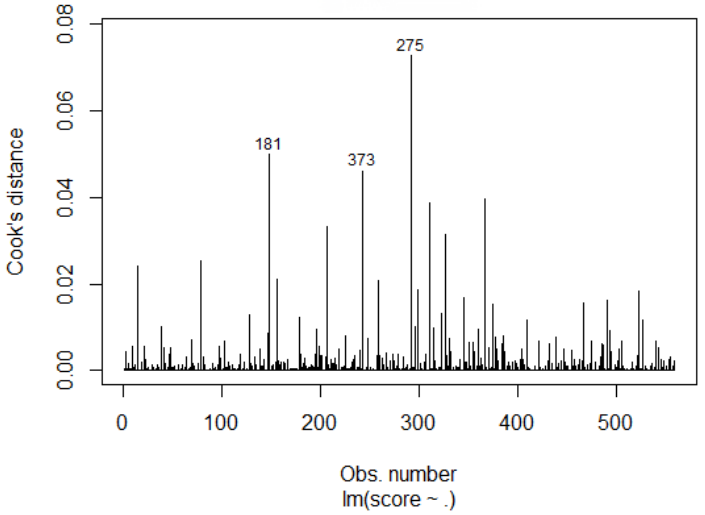

Figure 6. Cook's distance.

Figure 3 shows that the means of the residuals are almost always on a horizontal line, so the estimates of the dependent variables of several linear model models mentioned above are independent of the fluctuations of the residuals; the points in Figure 4 are almost on a straight line, so the assumption of normality of the errors is reasonable and the errors satisfy a normal distribution; the means of the points in Figure 5 remain on a straight line, and the assumption that the variance of this model is constant is reasonable, i.e., it proves the linear regression Figure 6 shows that the Cook distance of each observation is within 0.08, and there are almost no outliers and outliers that can significantly change the regression coefficient of the model.

In summary, the above linear equation model fits well. The basic assumptions are reasonable and there are no outliers, and the model has high credibility.

DOI: 10.54254/2753-8818/18/20230375

4. Discussion

The results showed that road stress perceptions differed among settlements of different build-up ages. In general, the negative perception of pressure is positively correlated with buildings and cars, and negatively correlated with vegetation, sky and sidewalks. Wide views and higher green views are a guarantee of high spatial quality. This also challenges today's high-density neighborhoods. The influence of each element varies over the different built-up eras. Architecture and greenery, while topping the list of influences, tend to wane. Residents are simultaneously considering more of the multiple components of the streetscape. The influence of vehicles and partial shading is increasing, and smooth and level sidewalks are important.

As the economy grows and technology develops, people are demanding higher and higher quality of living environment. Only meeting the needs of survival is a thing of the past. The new living environment needs to take more into account the comfort of human living, functional diversity, greening artistry, etc. Especially in the renovation of old neighborhoods and the improvement of the living environment of rural buildings, more respect for human perception and needs is needed.

There is still room for improvement in this study, and existing studies have difficulty in extracting pixel-level classification of higher-order information. Therefore, new research methods are needed to conduct more in-depth analysis. In addition, the use of visual interpretation and interview dialogues may cause some errors in the accuracy of the analysis results. This needs to be further improved and extended in future studies.

Reference

[1] Hickman, C., 2013. ‘To brighten the aspect of our streets and increase the health and enjoyment of our city’: The National Health Society and urban green space in late- nineteenth century London. Landsc. Urban Plan. 118, 112–119. https://doi.org/ 10.1016/j.landurbplan.2012.09.007.

[2] Carmona, M., Gabrieli, T., Hickman, R., Laopoulou, T., Livingstone, N., 2018. Street appeal: The value of street improvements. Prog. Plan. 126, 1–51. https://doi.org/ 10.1016/j.progress.2017.09.001.

[3] LONG Ying, TANG Jingxian. Progress of research on large-scale quantitative measurement of urban street space quality[J]. Urban Planning,2019,43(06):107-114.

[4] Wang Ruoyu, Liu Ye*, Lu Yi, Yuan Yuan, Zhang Jinbao, Liu Penghua & Yao Yao*. 2019. The linkage between the perception of neighbourhood and physical activity in Guangzhou, China: using street view imagery with deep learning techniques. International Journal of Health Geographics, 18(1), 18.

[5] Wang L, Han X, He J and Jung T (2022) Measuring residents’ perceptions of city streets to inform better street planning through deep learning and space syntax. Front. ISPRS Journal of Photogrammetry and Remote Sensing 190, 215–230. https://doi.org/10.1016/j.isprsjprs.2022.06.011

[6] Steinmets-Wood M,Velauthapillai K,O’Brien G,et al.Assessing the micro-scale environment using Google Street View: the Virtual Systematic Tool for Evaluating Pedestrian Streetscapes (Virtual-STEPS)[J].BMC Public Health, 2019,19: 1246.

[7] Yao, Y., Liang, Z., Yuan, Z., Liu, P., Bie, Y., Zhang, J., … & Guan, Q. (2019). A human-machine adversarial scoring framework for urban perception assessment using street-view images. International Journal of Geographical Information Science, 33(12), 2363-2384.

[8] Liu Y, Wang R, Grekousis G, et al. Neighbourhood greenness and mental wellbeing in Guangzhou, China: What are the pathways?[J]. Landscape & Urban Planning, 2019, 190(1): 28-33.

[9] Brown S C, Mason C A, Lombard J, Let al. The Relationship of Built Environment to Perceived Social Support and Psychological Distress in Hispanic Elders: The Role of "Eyes on the Street"[J]. J Gerontol B Psychol Sci Soc Sci, 2009, 64(2): 234-246.