https://doi.org/10.15302/J-LAF-0-020024 Received Date | 2023-11-02 Accepted Date | 2024-04-05

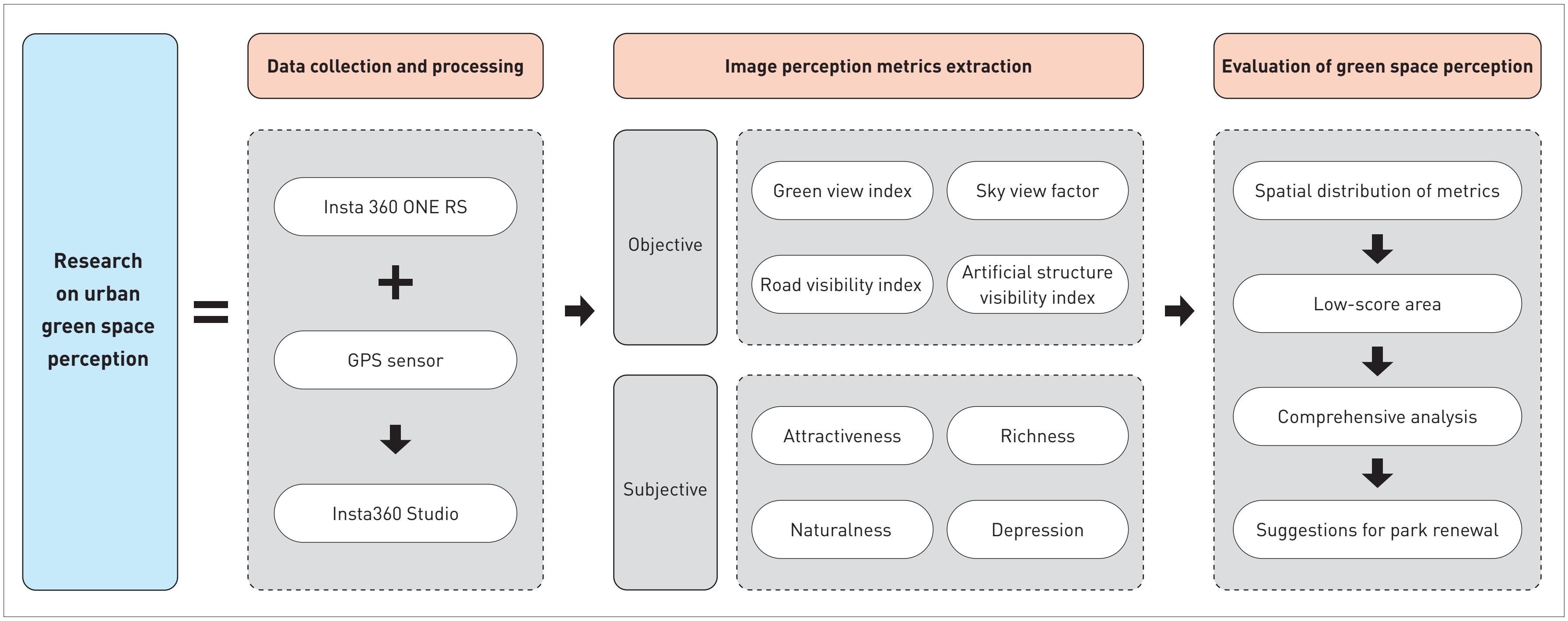
Research on the Perception Evaluation of Urban Green Spaces Using Panoramic Images and Deep Learning: A Case Study of Zhujiang Park in Guangzhou

Xukai ZHAO[1,2], **Guangsi LIN**[1,2,3,*]

1 State Key Laboratory of Subtropical Building and Urban Science, South China University of Technology, Guangzhou 510641, China
2 Department of Landscape Architecture, School of Architecture, South China University of Technology, Guangzhou 510641, China
3 Guangzhou Key Laboratory of Landscape Architecture, South China University of Technology, Guangzhou, 510641, China

*CORRESPONDING AUTHOR
Address: No. 381 Wushan Road, Tianhe District, Guangzhou 510641, Guangdong Province, China
Email: asilin@126.com

GRAPHICAL ABSTRACT



ABSTRACT

Visual quality assessment of urban green spaces is a major topic in landscape architecture research, yet traditional methods face limitations in practice. The rapid development of artificial intelligence and street-view big data offers opportunities for advancing green space perception studies. However, the lack of full street view image coverage of green spaces in China poses challenges for related research. Focusing on public landscape perception evaluation, this research took Zhujiang Park in Guangzhou, China as a case study. The research team utilized a convenient image collection method by panoramic camera and an effective processing workflow, and then employed the Segformer-B5 semantic segmentation model and the ViT-base-p16 image classification model to calculate four objective evaluation metrics (green view index, sky view factor, road visibility index, and artificial structure visibility index) and four subjective evaluation metrics (attractiveness, richness, naturalness, and depression) for visual quality assessment. Based on the spatial distribution results of these metrics, comprehensive analyses were conducted and low-score areas were identified. Research results indicate that vegetation and water features significantly enhance park attractiveness and positive perceptions, while excessive sky and artificial structures produce negative effects; oppressive artificial

landscapes and constrained architectural views also lower overall landscape quality. The image collection and visual perception evaluation methods proposed in this study provide a scientific basis for the renovation and management of urban green spaces.

KEYWORDS

Landscape Perception Evaluation; Visual Landscape Assessment; Panoramic Camera; Artificial Intelligence; Urban Green Space; Semantic Segmentation; Image Classification

HIGHLIGHTS

- Explores a convenient image collection and processing workflow using panoramic cameras for urban green spaces
- Develops a deep-learning-based evaluation method for park landscape visual quality, enabling unbiased analysis
- Applies quantitative computation and statistical approaches to rapidly identifying areas needing optimization measures by integrating subjective and objective evaluation metrics

RESEARCH FUNDS

- Project of "A study on the Inclusive Design in the Recreational Place of Urban Green Space in Response to Passive and Active Exclusion," National Natural Science Foundation of China (No. 52378054)
- Project of "Research on Green Space Supply Evaluation Methods Based on Public Perception," Fundamental Research Funds for the Central Universities (No. CGPY202410)
- Project of "Research and Application of Deep Learning-Driven Park Perception Evaluation Methods," South China University of Technology Step Climbing Program (No. j2tw202402095)

EDITED BY Yuting Gao, Xidong Ma, Tina Tian

1 Introduction

Urban green spaces—natural or semi-natural land uses within cities—are essential components of landscapes, offering residents a broad range of ecosystem services and opportunities for connection with nature and recreational activities[1]. Visual perception is one of the significant ways through which people perceive the environment[2], and assessing the visual quality of urban landscapes—also known as visual landscape assessment—constitutes a major topic of landscape research[3]. Such assessments provide valuable insights for researchers and governments into urban landscape quality. Traditional visual landscape assessment methods primarily employ scenic beauty estimation[4] and questionnaire surveys[5]. Although these methods effectively collect people's preferences on specific landscapes, they still have numerous limitations: they heavily rely on costly manual judgments from experts or respondents regarding images, making the process labor and resource-intensive, complex to implement, and limited in data sources[6][7]. Furthermore, the complexity of real-world landscapes often makes it challenging to apply findings from generalized studies into new scenarios.

The recent development of artificial intelligence (AI) technologies offers solutions to address these issues. AI has demonstrated tremendous potential in research on intelligent built environments and is widely recognized as highly promising in the fields of sustainable smart cities and landscape planning[8][9]. Among these advancements, street view images (SVIs) have emerged as a new form of crowd-sourced data that offers a realistic depiction of urban environments and supports feedback based on genuine human perceptions, which serves as a high-quality data source for evaluating the visual quality of urban built environments[6][10]. Related research spans five key areas: landscape design and environmental assessment, thermal environment, neighborhood morphology, environmental perception of neighborhood, and socioeconomic factor analysis[7].

With the increasing availability of satellites and coverage of street view services, satellite imagery and SVIs have become vital data sources for understanding large-scale urban landscapes. However, these sources also have certain limitations. For instance, satellite imagery fails to reflect the human-eye perception. Currently, most green spaces, communities, and educational institutions in many Chinese cities are not accessible to map service providers for image collection; certain roads also lack street view service coverage. Additionally, the lack of timely updates is a major shortcoming of publicly available data[11]. Consequently, scholars attempt to use wearable cameras, drones, or other devices to manually collect images as a supplement or replacement for SVIs. For instance, Yan Li et al. used GoPro cameras mounted on a car to collect street images of Xining City, China and developed a vacancy estimation model using object detection techniques to infer the storefront vacancy rate[12]. Similarly, Junjie Luo et al. employed drones to establish an oblique dataset for the river landscape visual evaluation of a section of the Grand Canal in Tianjin, China[11]. Despite these efforts, no studies have specifically focused on manually collecting images of parks as a substitute for SVIs. Given the fact that certain routes or terrains within parks (e.g., staircases, stepping stones) are unsuitable to collect images by riding or driving with a GoPro and drones cannot replicate human perspectives, further exploration of image collection devices and their application in visual landscape assessment is necessary.

Simultaneously, as quality of life improves, there is a growing demand for high-quality green spaces, requiring urban planners to accurately identify and improve low-quality areas within these spaces. However, as existing research is hard to be applied in practice, relevant assessments rely primarily on designers' personal experience and subjective determination, often overlooking the public's actual needs and preferences. AI technology has the potential to address this challenge by effectively simulating public perceptions and conducting visual evaluations on environmental images[6][7][10]. Nevertheless, AI-based public perception evaluation methods specific to parks have yet to be developed.

This study aims to establish an intelligent perception framework for urban green spaces based on urban park image collection and deep learning techniques. The goal is to enable rapid, accurate, and comprehensive evaluation of park visual quality, identify low-quality areas, and so as to inform spatial renewal and improvement plans. Specifically, this study focuses on the following questions. How to collect green space images more conveniently? How can AI-based algorithmic systems be developed to accurately reflect public perceptions and preferences of parks, thereby identifying low visual quality spaces? And in what ways can the perception evaluation results from such a system support theories related to visual landscape assessment? These explorations aim to promote the development of quantitative and evidence-based research on landscape perception and provide effective decision-making guidance for the renewal of urban green spaces.

2 Materials and Methods

2.1 Study Area

This study focused on Zhujiang Park, located in the Tianhe District of Guangzhou, Guangdong Province, China, which is an ecological park that integrates ecological, recreational, and cultural functions. It features diverse space types for various activities, covering an area of approximately 28 hm^2. The park is highly popular and serves as a representative green space in the subtropical region of China.

2.2 Technical Framework

First, this study adopted a convenient approach to collecting park images using a panoramic camera, and verified its feasibility with on-site operations. Subsequently, the Seformer-B5 model trained on the ADE20K dataset was used to automatically identify 150 categories of objects in the collected images and calculated four objective evaluation metrics: green view index (GVI), sky view factor (SVF), road visibility index (RVI), and artificial structure visibility index (ASVI). Additionally, four subjective evaluation metrics—attractiveness, richness, naturalness, and depression—were employed. A public perception dataset was established through pairwise comparisons of the images, classifying each image into high or low values for the four subjective dimensions. The ViT-base-p16 model was trained on this dataset to enable effective prediction of the subjective metrics. Next, the spatial distribution of both objective and subjective evaluation metrics was visualized, enabling the identification of areas associated with low-scoring images. Finally, correlations between objective and subjective metrics were analyzed to provide insights for park renovations (Fig. 1).

2.3 Data Collection and Processing

The images were collected on July 6, 2023, between 9:00 and 13:00, under clear weather conditions with temperatures around 30°C. A collector walked along all paths in the park, with the Insta360 ONE RS at a height of approximately 1.7 m. A handheld GPS sensor (Garmin eTrex 221x) recorded the location of each shooting point. According to previous experience, the research team captured images at road intersections, turning points, the midpoints of two turning points, and landmarks (e.g., buildings, pavilions, sculptures) to provide comprehensive visual information and ensure high efficiency. In this study, the interval distance between two collection points was no more than 40 m (about 50 walking steps). A total of 275 panoramic images were captured, all located along the centerline of the paths (Fig. 2).

Collected images were processed with Insta360 Studio and all were clear enough for the study. Then, the research team extracted perspectives at 0° and 180° in flat mode, generating 550 images that represent the surroundings at each point. The images were subsequently matched with GPS spatial data using ArcMap 10.6.

2.4 Deep Learning-based Image Evaluation Methods

2.4.1 Objective Evaluation Metrics Extraction With Semantic Segmentation Model

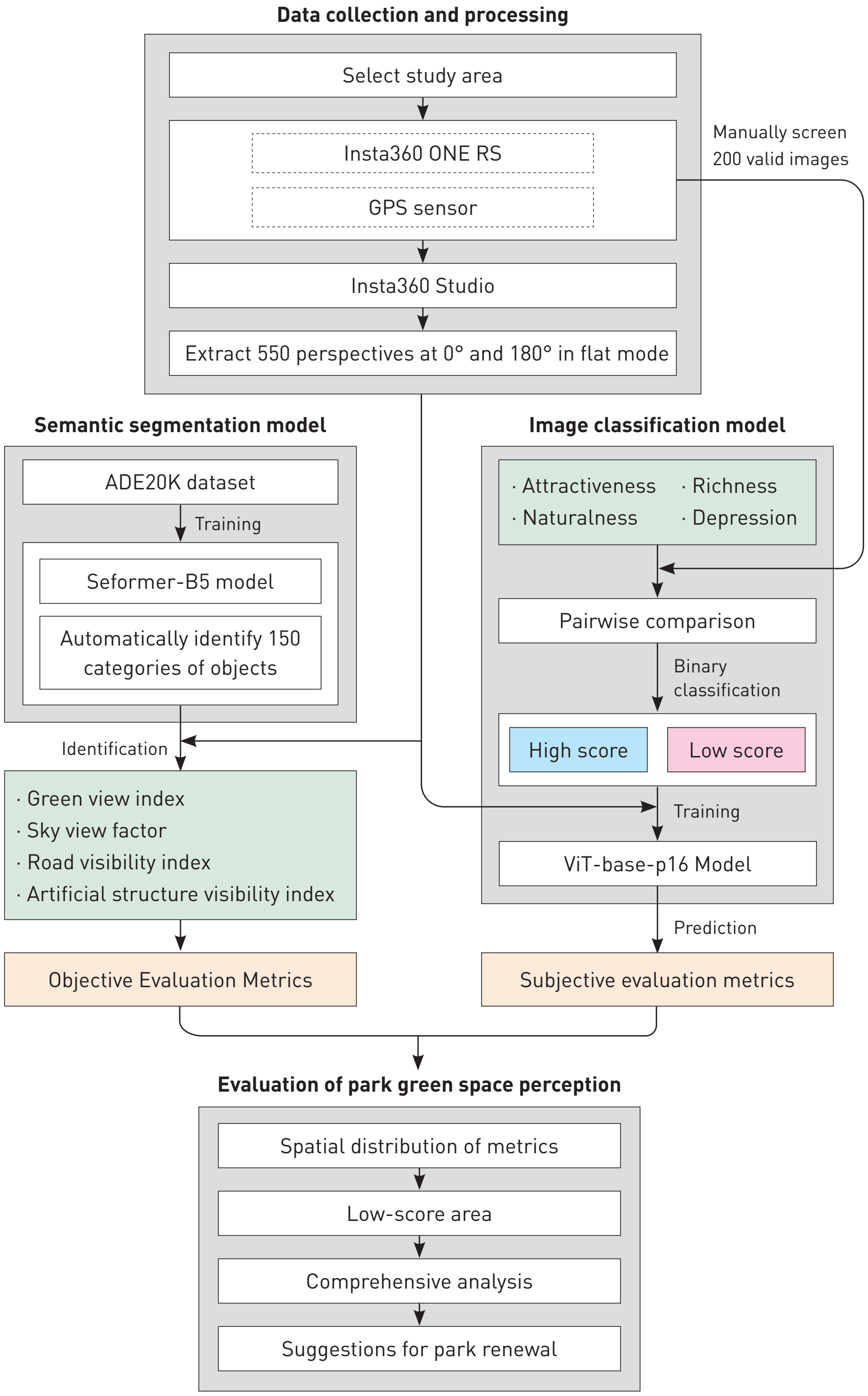
Physical elements in the environment (both natural and artificial) significantly influence the visual quality of landscapes and people's aesthetic perceptions. Semantic segmentation technology, a key technique for scene understanding, significantly improves the accuracy of identifying physical elements by pixel-level classification.

This study employed the SegFormer-B5 model[13], recognized for its high accuracy, to extract objective physical elements. The model consists of a hierarchical Transformer encoder and a lightweight All-MLP decoder. The Transformer encoder extracts image features using a self-attention mechanism to weigh important areas, enhancing segmentation performance. The All-MLP decoder fuses multi-level features and predicts semantic segmentation masks, outputting results through a fully connected layer. The model was trained using the ADE20K dataset[14], an open dataset for scene understanding released by MIT in 2016, which includes 150 element categories. Testing results show that the SegFormer-B5 model outperforms earlier models such as FCN, PSPNet, and DeepLabV3+[7], as well as advanced models like FPN and UPerNet on the ADE20K validation set①.

From the 150 element categories, this research extracted 13 common visual elements in parks②, and calculated GVI and SVF drawing from existing visual perception research[15]~[17]. Additionally, as Zhujiang Park has numerous roads and artificial structures (e.g., walls, benches, streetlights, and fences), this study introduced RVI and ASVI as metrics[11] (Table 1).

2.4.2 Subjective Perception Score Prediction With Image Classification Model

Traditional studies on subjective landscape perceptions often adopt methods like rating scales, pairwise comparisons, or categorization[18]. For instance, the Likert five-point scale requires respondents to rate images from 1 to 5. After obtaining scores, image classification models in deep learning can learn the relationship between scores and image features, simulating human perception process rating images from 1 to 5, enabling large-scale, rapid subjective perception evaluation. Existing studies mainly rely on large urban perception datasets. For example, the Place Pulse 2.0 dataset[19] includes over 110,000 images from 56 cities and over



1 © Xukai Zhao, Guangsi Lin

2 © Xukai Zhao, Guangsi Lin

1. Technical framework.
2. Image collection points.

① Model comparison data are available on the OpenMMlab GitHub page.
② The 13 common visual elements in parks include wall, building, sky, tree, shrub, ground cover, first-class road, second-class road, third-class road, fence, skyscraper, bench, and streetlight.

Table 1: Objective evaluation metrics

Dimension	Metric	Definition	Source
Natural	Green view index (GVI)	Proportion of pixels representing vegetation (tree, shrub, and ground cover)	Refs. [15][16]
	Sky view factor (SVF)	Proportion of pixels representing sky	Ref. [17]
Artificial	Road visibility index (RVI)	Proportion of pixels representing road (first-class, second-class, and third-class roads)	Ref. [11]
	Artificial structure visibility index (ASVI)	Proportion of pixels representing artifacts (wall, building, fence, skyscraper, bench, and streetlight)	Ref. [11]

80,000 online volunteers evaluated the images through pairwise comparisons across various dimensions to generate perception scores. This dataset has been used in subsequent research to train image classification models for subjective perception scoring prediction[6]. Relevant studies demonstrate that combining subjective visual surveys, image semantic segmentation, and image classification models can effectively and fairly collect and map street-level perceptions[6]. Although the dataset lacks data park-scene image data that can be directly applied into this study, its construction methods and subjective perception prediction approaches provide a foundation for the subjective evaluation of this study.

(1) Establishing subjective evaluation metrics

Drawing from traditional visual landscape assessment research[20]~[24], four subjective evaluation metrics were selected: attractiveness, richness, naturalness, and depression. Attractiveness refers to the degree to which a park scene appeals to individuals, encompassing factors like beauty and uniqueness[20]. Richness reflects the diversity and complexity of park elements, including species and design elements[20][21]. Naturalness represents the balance between human intervention and the natural state in the perceived park environment, informing park maintenance and management strategies[22]. Depression measures the extent to which a park induces feelings of melancholy or discomfort[23], often used to assess the impact of urban landscapes on physical and mental health[24]. Parks inducing high levels of depression may discomfort the visitors and negatively affect overall experience.

(2) Collecting pairwise comparison results

Compared with directly obtaining numerical ratings from participants, pairwise comparison is a more effective and accurate way to gather perception data[19]. First, to ensure the coverage of as many kinds of park scenes as possible, 550 photos were manually screened③, and those with excessive similarity were excluded, leaving 200 valid photos. Next, the research team developed an online rating system using JavaScript, which dynamically adjusted the displayed images based on user selections and the existing relationships between images to ensure each photo had enough comparison times and valid ratings. In each comparison, two images were randomly selected from the 200 photos (Fig. 3). Participants were asked to choose the image that better aligned with their preferences based on a question (e.g., “Which scene do you think exhibit more attractiveness/richness/naturalness/depression?”). Each participant performed four experiments, with each focusing on a single metric. To avoid fatigue, the number of comparisons in each experiment was limited to approximately 50 and kept less than 10 min. The experiment involved 35 master students, primarily majored in Landscape Architecture at South China University of Technology (12 males and 23 females), all of whom had no color blindness or color weakness. The experiment was conducted online over three days (March 3 ~ 5, 2024). On average, this research yielded 6,702 pairwise comparison results across four metrics, with an average of 1,675.5 results per metric.

③ Manual screening refers to image selection based on subjective perception and personal experience by the collector, without rigid quantitative criterion.

3. Subjective rating system based on pairwise comparison of images.

3 © Xukai Zhao, Guangsi Lin

(3) Calculation of subjective evaluation metrics

Drawing on existing research[25], this research used the "strength of schedule" method to statistically analyze the subjective ratings, obtaining high and low scores for each metric (Fig. 4).

For the subjective evaluation metric m, this research defined the frequency of a given image i when being selected ($W_{i,m}$) and not being selected ($L_{i,m}$) as follow:

$$W_{i,m} = \frac{w_{i,m}}{w_{i,m} + l_{i,m}}, \tag{1}$$

$$L_{i,m} = \frac{l_{i,m}}{w_{i,m} + l_{i,m}}, \tag{2}$$

where $w_{i,m}$, and $l_{i,m}$ represent the number of times the image being chosen, or not being chosen during comparisons.

The perception score ($Q_{i,m}$) of each image i for the evaluation metric m can be defined as follow:

$$Q_{i,m} = W_{i,m} + \frac{1}{n_i^w}\sum_{k_1=1}^{n_i^w} W_{k_1 m} - \frac{1}{n_i^l}\sum_{k_2=1}^{n_i^l} W_{k_2 m}, \tag{3}$$

where n_i^w and n_i^l represent the total number of times image i being selected and not being selected, respectively. To further categorize the image scores $Q_{i,m}$ into low and high value, the research team defined the following binary label $W_{i,m} \in \{0,1\}$, where 0 represents a low score and 1 represents a high score:

$$W_{i,m} = \begin{cases} 0 & if \;\; Q_{i,m} > \;\; \mu_m + \sigma_m \\ 1 & if \;\; Q_{i,m} < \;\; \mu_m - \sigma_m \end{cases}, \tag{4}$$

where μ_m and σ_m represent the mean and standard deviation of the perception scores across all data for the evaluation metric m, respectively.

(4) Image classification model training

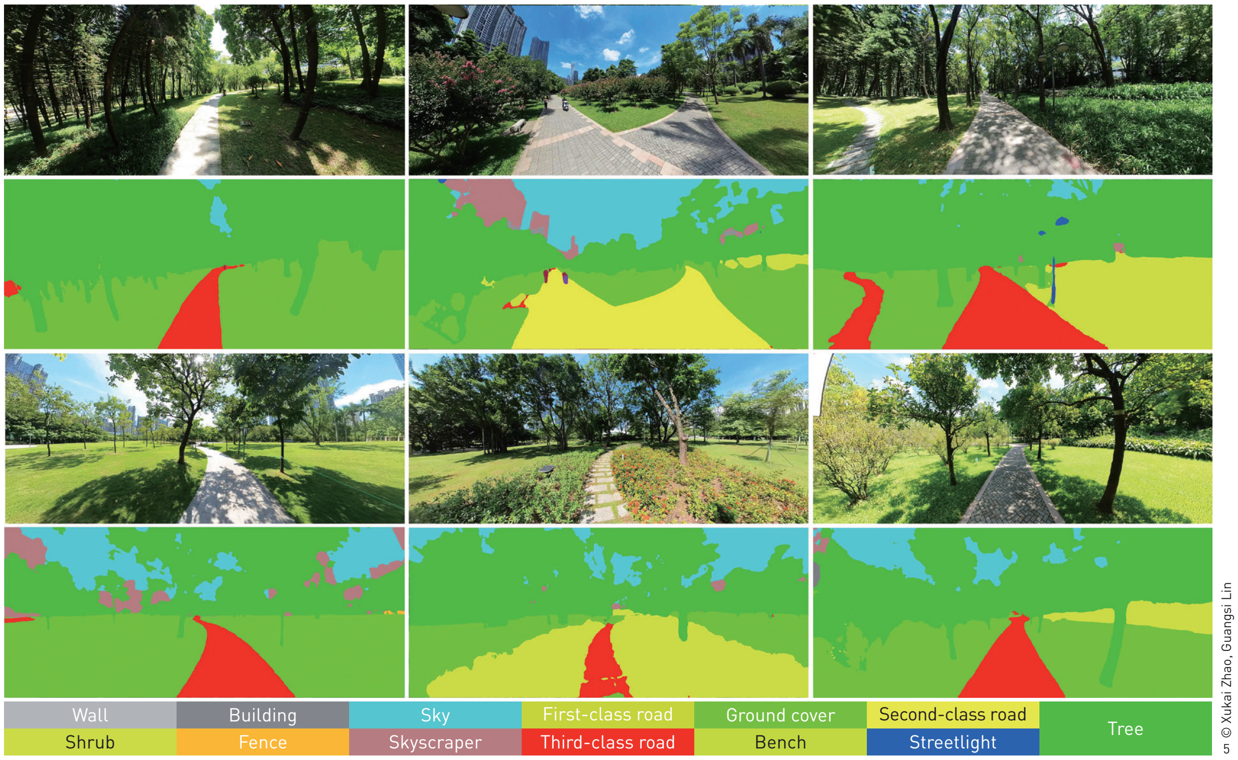
After the above calculations, each of the 200 images was assigned a value of "0" or "1" for all four metrics, the public perception dataset was formed. The image classification model used these values as labels and images as explanatory variables for training. This research employed the ViT-base-p16 model[26] for image classification. The ViT-base-p16 model divides input images into patches and treats each patch as a sequence element for input into a Transformer model. Using a self-attention mechanism, it weights important areas in the input images, effectively capturing important information. During the training, the ViT-base-p16 model was firstly pre-trained on the large-scale ImageNet-1k dataset to learn general representations of images. It was then fine-tuned on the public perception dataset for each of the metrics, resulting in four separate models for predicting attractiveness, richness, naturalness, and depression scores among all park images.

The performance of the model was evaluated using five-fold cross-validation. Specifically, the dataset of 200 images was randomly divided into five subsets. In each training iteration, four subsets were designated as the training set, and the remaining subset served as the validation set. The average accuracy across all five iterations was calculated to assess overall performance. The model with the highest accuracy was selected for scoring the subjective metrics. This approach ensured robustness of the training set while enhancing the model's generalizability to new data, enabling superior performance upon the limited sample size.

2.5 Integrated Evaluation of Subjective and Objective Metrics

The trained SegFormer-B5 and ViT-base-p16 models were employed to calculate both subjective and objective evaluation metrics for all 550 images. For each location, the average values of the two images were taken as the final scores. The score of these data points were visualized in ArcMap 10.6 to create spatial

4. Examples of image scoring across the four metrics.



5. Examples of semantic segmentation results.

distribution maps of both objective and subjective metrics, identifying low score areas. Since the data did not conform to a normal distribution, Spearman correlation analysis was thus applied to examine the relationships between subjective and objective metrics with the major elements in the images that take a larger proportion, including vegetation (trees, shrubs, and ground cover) and park paths (first-class, second-class, and third-class roads).

3 Results and Discussion

3.1 Results of Objective Evaluation Metrics

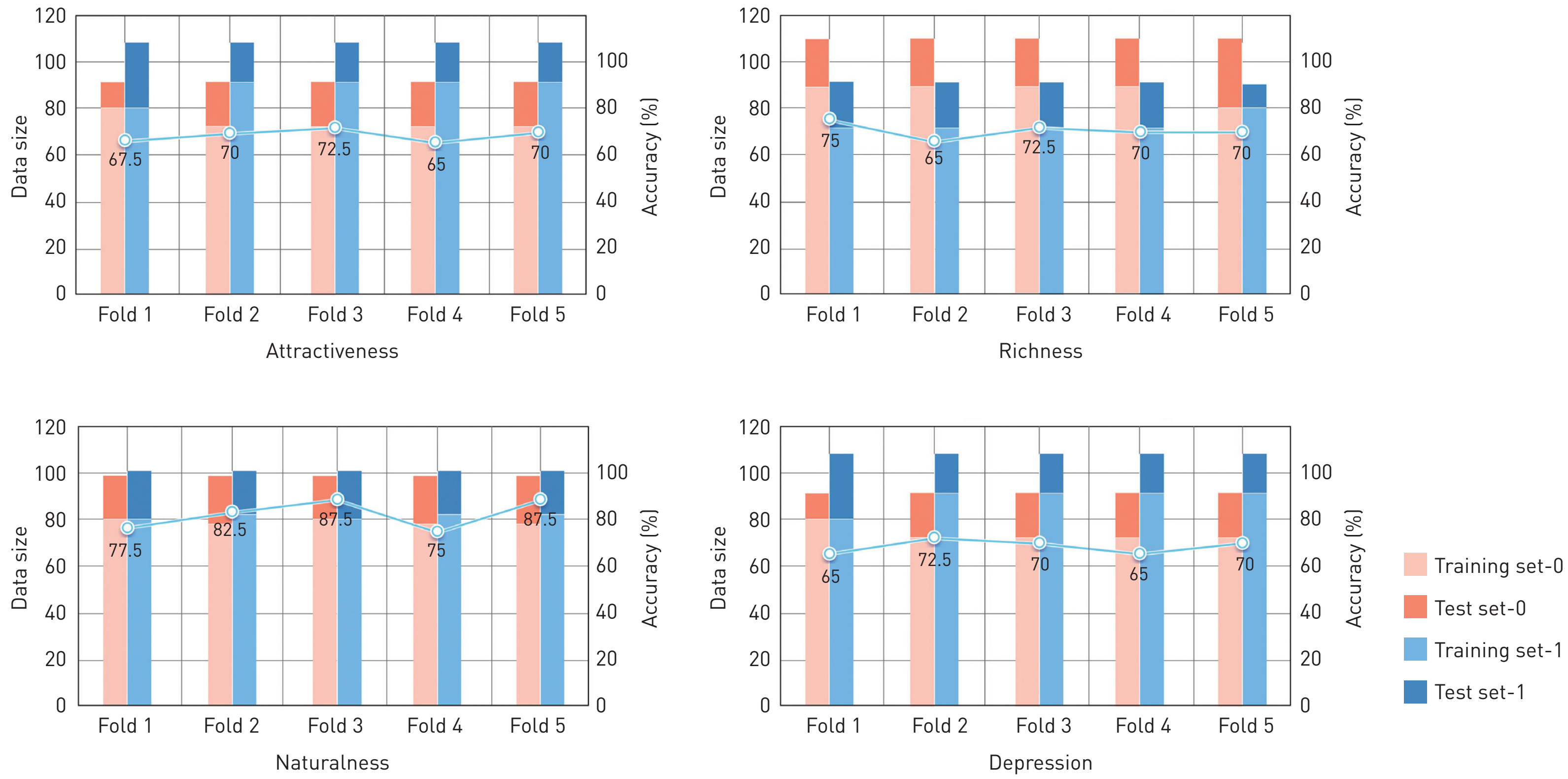
Figure 5 shows examples of different landscape elements identified through semantic segmentation using the SegFormer-B5 model. Table 2 summarizes the results of four objective evaluation metrics. Specifically, the average GVI was the highest (0.7115, with tree coverage at 0.3973, shrub coverage at 0.1691, and ground cover at 0.1450), indicating excellent vegetation conditions of the park, which is the main constituent of the park's landscape. The low average SVF (0.0737) corresponds to the high vegetation coverage, reflecting the dense tree canopy. The low averages of RVI and ASVI also reveal that the park is dominated by natural landscapes. The median values for these two metrics are close to their averages, indicating a relatively low coverage of roads and artificial structures. Moreover, the low *SD* suggests consistent internal park structures, contributing to a uniform visitor experience.

3.2 Training Results of Subjective Evaluation Metric Prediction Model

The distribution of five-fold cross-validation data and model prediction accuracy for the subjective evaluation metrics (Fig. 6) shows that, though the model's accuracy fluctuated across different metrics, the overall trend was stable. The average prediction accuracies for the test set were 69% (attractiveness), 70.5% (richness), 82% (naturalness), and 68.5% (depression), demonstrating a high reliability.

Table 2: Results of objective evaluation metrics

Objective metric	Minimum	Maximum	Average	Median	*SD*
GVI	0.0000	0.9731	0.7115	0.7351	0.1630
SVF	0.0000	0.2815	0.0737	0.0643	0.0557
RVI	0.0000	0.4237	0.1236	0.1011	0.0910
ASVI	0.0000	0.3787	0.0286	0.0127	0.0462



6. Results of five-fold cross-validation and model prediction accuracy.

6 © Xukai Zhao, Guangsi Lin

The statistical results for subjective metrics (Table 3) showed that naturalness had the highest mean value, indicating that the naturalness of Zhujiang Park was particularly prominent in human perception. This aligns with the semantic segmentation results. The range of naturalness was the largest (0.0443 ~ 0.8855), and the mean and median were close with the highest *SD*, reflecting significant spatial heterogeneity in vegetation distribution. Both attractiveness and depression also had relatively high mean values, suggesting that park scenes with high naturalness generally have strong appeal. However, excessively dense vegetation may increase feelings of depression. The *SD* for these two metrics were moderate and similar, indicating a relatively consistent variability across the sample. In contrast, the distribution of richness was more concentrated, with a lower *SD* and a narrower range (0.0732 ~ 0.5826), indicating relatively small differences in this metric. The lower mean value suggests that the diversity of visual elements in the park was relatively inadequate. This contrasts with the high variability in naturalness, indicating that while naturalness varies greatly across different scenes, the richness of visual elements is comparatively insufficient. This highlights a need to enhance landscape diversity in the park.

Table 3: Statistical results of subjective evaluation metrics

Subjective metric	Minimum	Maximum	Average	Median	*SD*
Attractiveness	0.0783	0.7540	0.4285	0.4485	0.1476
Richness	0.0732	0.5826	0.2841	0.2724	0.0942
Naturalness	0.0443	0.8855	0.4303	0.3776	0.2021
Depression	0.1619	0.8710	0.3821	0.4120	0.1468

3.3 Integrated Evaluation Results of Objective and Subjective Metrics

Overall, the spatial distribution patterns of objective and subjective metrics in Zhujiang Park showed similarities (Figs. 7, 8). The lawn area in front of the west entrance of the park (Zone C), characterized by open lawns and short trees with sparse shrubs, has wide paths and a higher SVF, but lower scores in GVI and naturalness, as well as relatively low attractiveness. The Kuailv Lake area in the central part of the park (Zone E), despite low GVI and high SVF, exhibited high attractiveness. This aligns with previous findings that people generally prefer water features[5]. The scenic forest area in the eastern part of the park (Zone F) had high GVI and naturalness, making it a valuable asset in the bustling city center of Guangzhou. Its winding, undulating paths, combined with a low proportion of road and artificial structures, contributed to the

7 © Xukai Zhao, Guangsi Lin

7. Examples of park scenes.

overall high attractiveness. Some areas in the southwest part of the park have dense vegetation and diverse spatial variations, leading to higher richness. However, the variability of scenes between different points results in varying levels of attractiveness. The service buildings on the eastern side of the park (Zone G) has monotonous facades and low attractiveness, requiring special attention in park management.

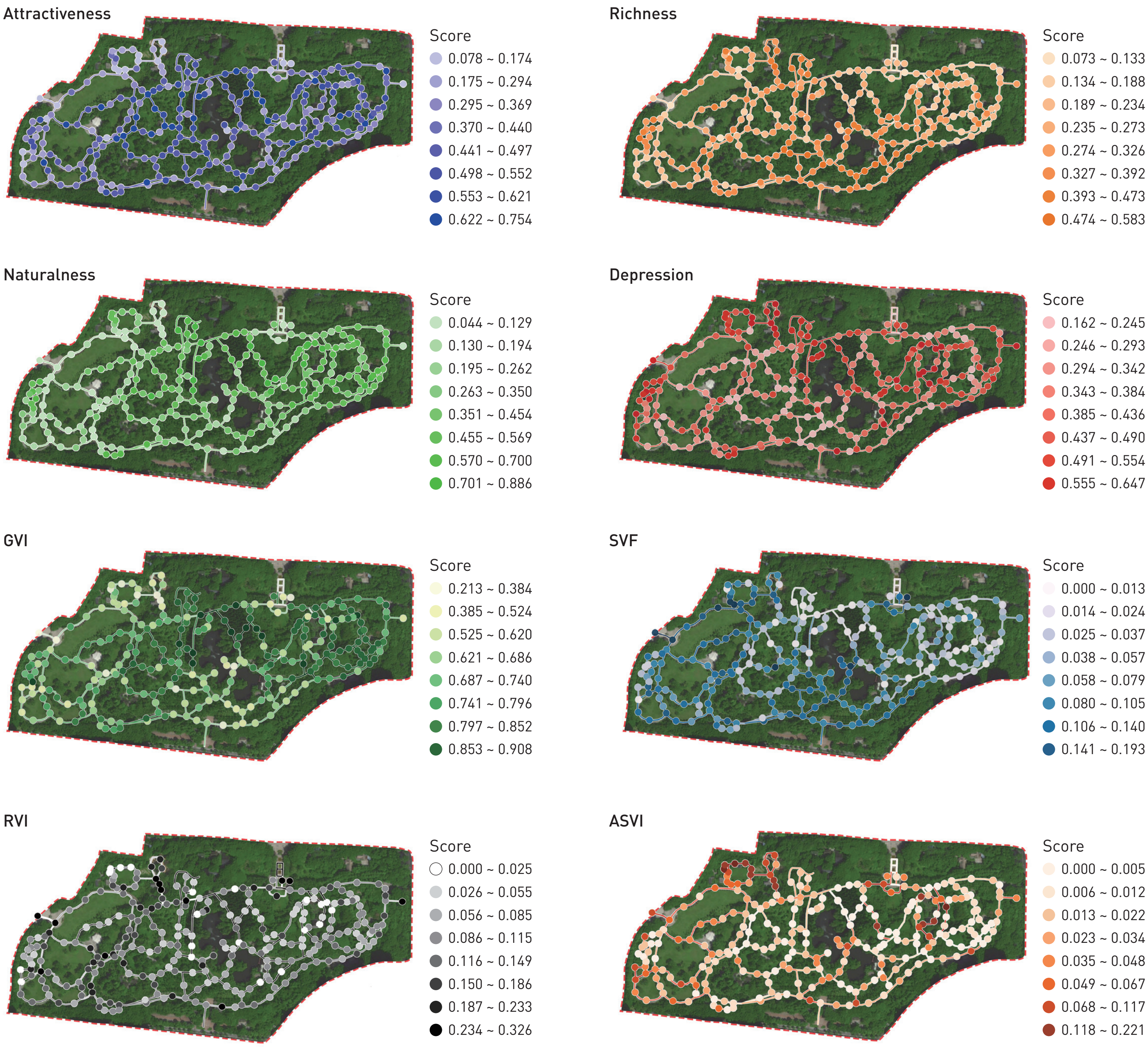
Spearman correlation analysis (Fig. 9) revealed a significant positive correlation between naturalness and attractiveness (r_s = 0.60), indicating that scenes with high naturalness are more favored by people. This finding aligns with previous research that visitors prefer environments with abundant vegetation. Such preferences may positively influence park usage frequency and visitor satisfaction[27]. The proportion of ground cover was significantly negatively correlated with richness (r_s = − 0.48), suggesting that an increase in ground cover may reduce the overall richness. In Zhujiang Park, areas with a high proportion of ground cover are primarily located in the western part of the park, characterized by open lawns, leading to lower spatial richness. Naturalness was significantly positively correlated with GVI (r_s = 0.71), proportion of tree (r_s = 0.47), and proportion of shrub(r_s = 0.46) . GVI and naturalness represent subjective and objective ecological environment, respectively, but the perception of naturalness is influenced not only by vegetation proportions but also by other factors, such as the overall composition of green elements and the presence of additional materials in the images (e.g., water, soil or permeable pavements). Depression shows a significant positive correlation with both naturalness (r_s = 0.64) and proportion of shrub (r_s = 0.65), indicating the dense vegetation may evoke feelings of depression.

Furthermore, SVF, RVI, and ASVI show positive correlations with each other and negative correlations with all four subjective perception metrics as well as GVI. This suggests that increases in the proportions of sky, roads, and artificial structures are associated with decreases in vegetation and naturalness. In Zhujiang Park, areas with higher proportions of sky, roads, buildings, walls, and benches, e.g., the children's play area in the northwest (Zone B) and the lawn area in front of the west entrance (Zone C), tend to have lower vegetation coverage, wide paths, and open spaces, and their attractiveness and naturalness are lower—compared with the scenic forest area in the eastern part that is densely vegetated—though the openness of these areas reduces feelings of depression.

4 Conclusions and Prospects

The European Landscape Convention emphasizes that landscapes are a vital public interest deserving recognition and protection[28]. Understanding how individuals observe and perceive landscapes and incorporating these insights into landscape planning and management is critical. This study adopts advanced image collection and AI technologies to develop a methodological framework for landscape research and practice centered on landscape perception. Overall, this study demonstrates three key contributions as follow.

1) This research implemented a convenient and efficient workflow by combining urban green space image collection by panoramic camera with advanced semantic segmentation and image classification models for unbiased assessments on park visual quality. This method overcomes limitations in traditional



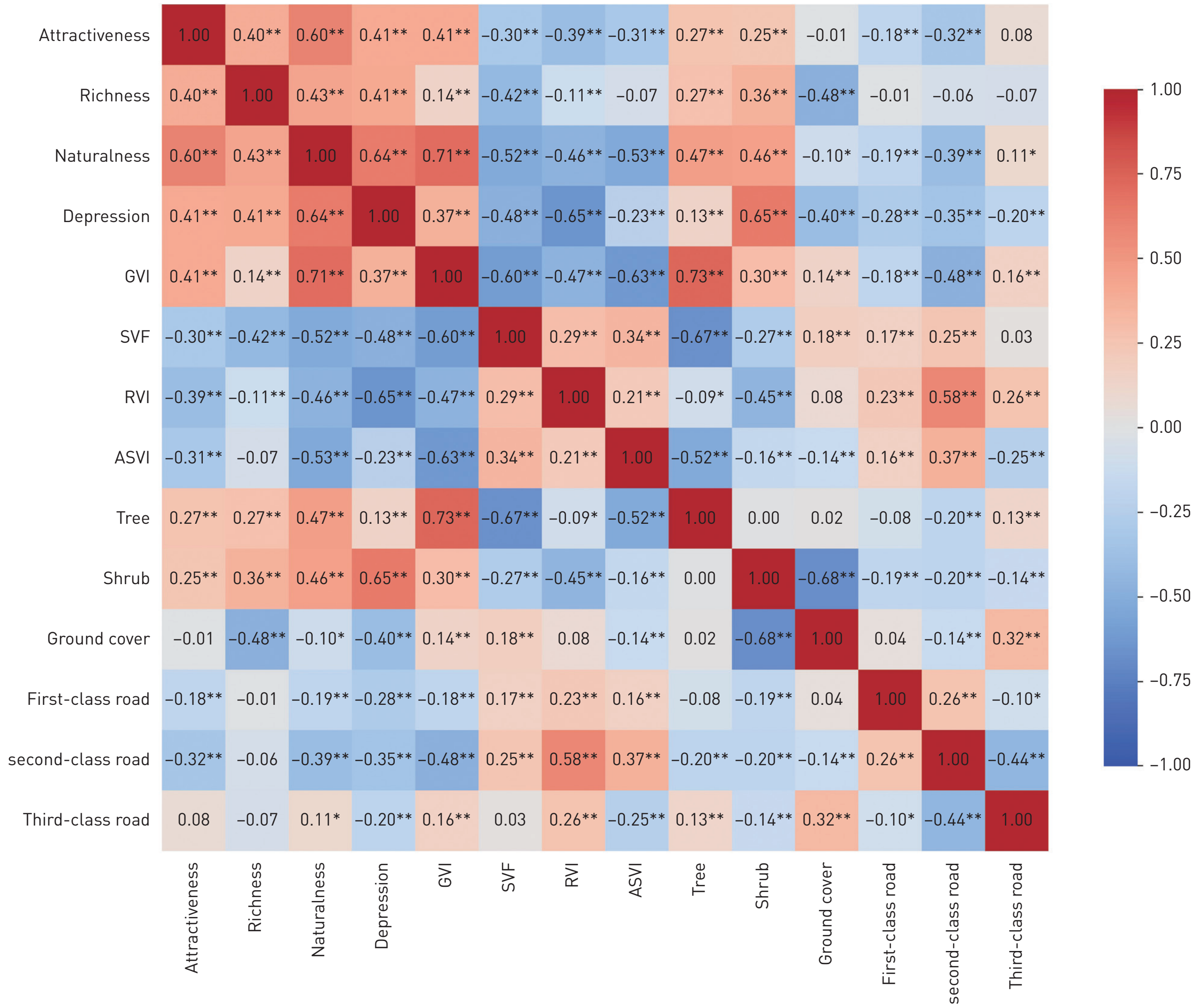
8. Spatial distribution results of objective and subjective metrics.

visual assessments such as inefficiencies in processing large volumes of images or fatigue in multi-scene evaluations, and validates the application of image big data and deep learning in landscape perception research.

2) Traditional visual assessment studies often rely on small-sized image datasets and lack accurate quantification of subjective and objective elements. This research precisely extracted and evaluated objective metrics and predicted scoring on subjective metrics, revealing that the presence of vegetation and water features enhances park attractiveness and stimulates positive perceptions. Conversely, higher proportions of sky, roads, and artificial structures are found to have negative effects.

3) Traditional research findings are often difficult to be applied directly to new scenarios' preference prediction. The intelligent method demonstrated in this paper can learn subjective scoring from a subset of scene images and predict scores for other new scenes, helping park managers efficiently identify low-scoring areas. This provides actionable guidance for urban green space renewal, demonstrating significant practical value.

Despite the contributions, this research has certain limitations.



9 © Xukai Zhao, Guangsi Lin

9. Spearman correlation analysis results (* indicates significant correlation at the 0.05 level, ** indicates significant correlation at the 0.01 level).

The image data and the number of participants were relatively limited, and the study focused exclusively on summer landscapes of Zhujiang Park, making it maybe difficult to generalize findings to parks of other types or in other seasons. Notably, the children's play area in the northwest had lower attractiveness, according to the research results, likely due to the preferences of the selected participants—university students—who may find areas characterized by low vegetation and richness less appealing. It also underlines a limitation of prior studies based on street view big data, which train models on generalized public preferences and fail to reflect the various needs of different user groups[19][25]. Additionally, panoramic camera images may have distortion, potentially affecting accuracy. Future studies should expand the dataset to include more diverse urban green spaces and seasonal landscapes; gather ratings from a broader range of users to improve green space perception datasets; and pay attention to the necessity of conducting preference surveys across diverse user groups.

Finally, the subjective and objective metrics extracted in this study could be integrated with other data, including park vitality, functional usage, and environmental quality, to further explore the relationships between factors including landscape attractiveness, user behavior patterns, and physical characteristics of a park. Such studies will support urban managers in systematic decision-making of developing more precise strategies, optimizing park functions, and improving urban landscape quality.

Competing interests | The authors declare that they have no competing interests.

REFERENCES

[1] Wolch, J. R., Byrne, J., & Newell, J. P. (2014). Urban green space, public health, and environmental justice: The challenge of making cities 'just green enough'. *Landscape and Urban Planning*, (125), 234–244.

[2] Daniel, T. C. (2001). Whither scenic beauty? Visual landscape quality assessment in the 21st century. *Landscape and Urban Planning*, 54(1–4), 267–281.

[3] Gobster, P. H., Ribe, R. G., & Palmer, J F. (2019). Themes and trends in visual assessment research: Introduction to the *Landscape and Urban Planning* special collection on the visual assessment of landscapes. *Landscape and Urban Planning*, (191), 103635.

[4] Daniel, T. C. (1976). *Measuring Landscape Esthetics: The Scenic Beauty Estimation Method*. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station.

[5] Cai, K., Huang, W., & Lin, G. (2022). Bridging landscape preference and landscape design: A study on the preference and optimal combination of landscape elements based on conjoint analysis. *Urban Forestry & Urban Greening*, (73), 127615.

[6] Zhao, X., Lu, Y., & Lin, G. (2024). An integrated deep learning approach for assessing the visual qualities of built environments utilizing street view images. *Engineering Applications of Artificial Intelligence*, (130), 107805.

[7] He, N., & Li, G. (2021). Urban neighbourhood environment assessment based on street view image processing: A review of research trends. *Environmental Challenges*, (4), 100090.

[8] Sanchez, T. W., Shumway, H., Gordner, T., & Lim, T. (2022). The prospects of artificial intelligence in urban planning. *International Journal of Urban Sciences*, 27(2), 179–194.

[9] Cheng, Y., & Fan, B. (2023). Digital landscape process. *Chinese Landscape Architecture*, 39(6), 6–12.

[10] Biljecki, F., & Ito, K. (2021). Street view imagery in urban analytics and GIS: A review. *Landscape and Urban Planning*, (215), 104217.

[11] Luo, J., Zhao, T., Cao, L., & Biljecki, F. (2022). Semantic Riverscapes: Perception and evaluation of linear landscapes from oblique imagery using computer vision. *Landscape and Urban Planning*, (228), 104569.

[12] Li, Y., & Long, Y. (2024). Inferring storefront vacancy using mobile sensing images and computer vision approaches. *Computers, Environment and Urban Systems*, (108), 102071.

[13] Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., & Luo, P. (2021). SegFormer: Simple and efficient design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems*, (34), 12077–12090.

[14] Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., & Torralba, A. (2017). Scene parsing through ADE20K dataset. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 633–641). Computer Vision Foundation.

[15] Qiu, W., Li, W., Liu, X., Zhang, Z., Li, X., & Huang, X. (2023). Subjective and objective measures of streetscape perceptions: Relationships with property value in Shanghai. *Cities*, (132), 104037.

[16] Song, Q., Li, W., Li, M., & Qiu, W. (2022). Social inequalities in neighborhood-level streetscape perceptions in Shanghai: The coherence and divergence between the objective and subjective measurements. *Social Science Research Network*.

[17] Xia, Y., Yabuki, N., & Fukuda, T. (2021). Sky view factor estimation from street view images based on semantic segmentation. *Urban Climate*, (40), 100999.

[18] Lange, E., & Legwaila, I. (2012). Visual landscape research—Overview and outlook. *Chinese Landscape Architecture*, 28(3), 5–14.

[19] Dubey, A., Naik, N., Parikh, D., Raskar, R., & Hidalgo, C. A. (2016). Deep learning the city: Quantifying urban perception at a global scale. *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I* (pp. 196–212). Springer.

[20] Sun, D., Li, Q., Gao, W., Huang, G., Tang, N., Lyu, M., & Yu, Y. (2021). On the relation between visual quality and landscape characteristics: A case study application to the waterfront linear parks in Shenyang, China. *Environmental Research Communications*, 3(11), 115013.

[21] Zhang, G., Yang, J., & Jin, J. (2021). Assessing relations among landscape preference, informational variables, and visual attributes. *Journal of Environmental Engineering and Landscape Management*, 29(3), 294–304.

[22] Wartmann, F. M., Stride, C., Kienast, F., & Hunziker, M. (2021). Relating landscape ecological metrics with public survey data on perceived landscape quality and place attachment. *Landscape Ecology*, (36), 2367–2393.

[23] "Depressing." *Oxford English Dictionary*. Oxford University Press.

[24] Gong, Y., Palmer, S., Gallacher, J., Marsden, T., & Fone, D. (2016). A systematic review of the relationship between objective measurements of the urban environment and psychological distress. *Environment International*, (96), 48–57.

[25] Zhang, F., Zhou, B., Liu, L., Fung, H. H., Lin, H., & Ratti, C. (2018). Measuring human perceptions of a large-scale urban region using machine learning. *Landscape and Urban Planning*, (180), 148–160.

[26] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. *International Conference on Learning Representations*.

[27] Talal, M. L., Santelmann, M. V., & Tilt, J. H. (2021). Urban park visitor preferences for vegetation—An on-site qualitative research study. *Plants, People, Planet*, 3(4), 375–388.

[28] Council of Europe. (2000). *Explanatory Report to the European Landscape Convention*.

https://doi.org/10.15302/J-LAF-0-020024 收稿日期 | 2023-11-02 录用日期 | 2024-04-05 中图分类号 | TU986 文献标识码 | A

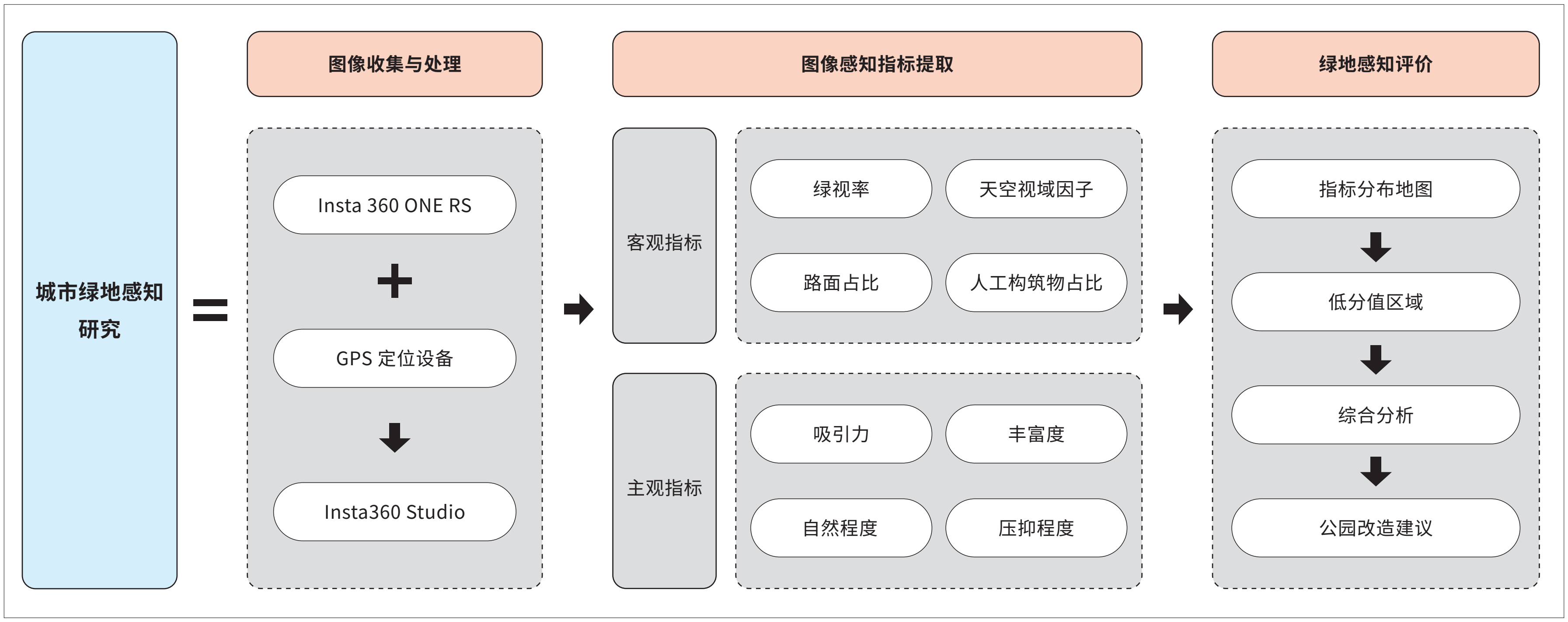
基于全景影像采集与深度学习技术的城市绿地感知评价研究——以广州市珠江公园为例

赵旭凯[1,2]，林广思[1,2,3,*]

1 华南理工大学亚热带建筑与城市科学全国重点实验室，广州 510641
2 华南理工大学建筑学院风景园林系，广州 510641
3 华南理工大学广州市景观建筑重点实验室，广州 510641

*通讯作者
地址：广东省广州市天河区五山路381号
邮编：510641
邮箱：asilin@126.com

图文摘要



摘要

城市绿地的视觉质量评估是景观设计学领域的重要话题，但传统研究方法在实际操作中存在一定局限。人工智能技术与街景大数据的发展为城市绿地感知评估带来了契机。然而，由于中国城市的绿地尚未被街景服务全面覆盖，相关研究的开展受到了限制。本文立足于景观的公众感知评价，以中国广州市珠江公园为例，采用便捷的全景相机图像采集与处理流程，利用Segformer-B5语义分割模型和ViT-base-p16图像分类模型分别对公园图像计算客观评价指标（绿视率、天空视域因子、路面占比、人工构筑物占比）与主观评价指标（吸引力、丰富度、自然程度、压抑程度），从而进行公园绿地视觉质量评估。基于各项评价指标空间分布图，进行综合分析并识别低分值区域。结果发现，植被与水体有助于提升公园的吸引力与游客的积极感知，而过多的天空与构筑物则可能会产生相反效果；消极的人工景观和压抑的建筑也会降低景观质量。本研究所提出的图像采集与视觉感知评估方法可为城市绿地更新与管理提供科学依据。

关键词

景观感知评价；视觉景观评估；全景相机；人工智能；城市绿地；语义分割；图像分类

文章亮点

- 探索实践了一种便捷的基于全景相机的城市绿地实景图像的采集流程
- 建构了基于深度学习的公园景观视觉质量评估方法，实现无差别分析
- 运用定量计算与统计方法，综合主观与客观评价指标快速识别待优化区域

基金项目

- 国家自然科学基金面上项目“应对主被动排斥的城市绿色空间游憩场所包容性设计研究”（编号：52378054）
- 中央高校基本科研业务费专项资金项目“基于公众感知的绿地供给评价方法研究”（编号：CGPY202410）
- 华南理工大学百步梯攀登计划项目“深度学习驱动的公园感知评价方法研究及其应用”（编号：j2tw202402095）

编辑 高雨婷，马锡栋，田乐

1 引言

城市绿地——城市中自然或半自然的土地利用状态——作为景观的重要组成部分，为城市居民提供了广泛的生态系统服务，以及亲近自然和游憩交往的机会[1]。视觉是公众感知景观环境的重要方式之一[2]，而评估城市景观的视觉质量，即视觉景观评估，是景观研究领域的重要话题[3]，对其进行分析有助于为研究人员和政府提供了解城市景观质量的重要途径。传统的视觉景观评估方法，如美景度评价法[4]和问卷调查法[5]，尽管可有效收集人们对特定景观的偏好，但仍存在高度依赖于专家或受访者对图像的主观评判、人力物力成本巨大、操作复杂、数据来源相对有限等诸多弊端[6][7]。此外，实际景观的复杂性使得相关研究结果难以直接被应用于新场景的偏好预测。

近年来，人工智能（AI）技术的发展为解决上述问题带来了契机。AI技术已在智能建成环境研究中展现出巨大潜力，并被广泛认为在可持续智能城市与景观规划领域极具发展前景[8][9]。其中，街景图像作为新兴的众包数据，可展现出真实城市环境并利于人们实际感知的反馈，是衡量城市建成环境视觉质量的优质数据源[6][10]。相关研究内容广泛，涵盖了景观设计及环境评估、热环境、社区形态、邻里环境感知、社会经济因素分析五大方向[7]。

此外，随着卫星数量的增加与街景服务覆盖范围的扩展，卫星与街景图像已成为理解大尺度城市景观的重要数据源，但这些数据源也存在一定局限性。卫星影像难以反映人眼的真实视觉感知；当前，中国多数城市中的绿地、社区、高校等区域未向地图服务商开放图像采集，某些道路也尚未被街景服务覆盖。此外，更新不及时也是公开数据的一大不足[11]。因此，有学者尝试使用可穿戴相机、无人机等设备人工采集图像，以补充或替代街景图像。例如，李彦等人驾车使用GoPro采集西宁城区的街道图像，并利用目标检测技术开发估算模型来推断西宁市的店面空置率[12]；罗俊杰等人使用无人机建立京杭大运河天津区段河流景观的无人机斜向数据集，并开展视觉评价[11]。然而，目前尚未有针对公园这一绿地类型进行人工采集图像来补充或替代街景图像的研究。同时，由于城市绿地内部地形复杂，部分路线（如阶梯、汀步等）并不适合骑行或驾车使用GoPro采集图像，且常常存在清晰度较差、采集视角偏高等问题，而无人机也无法还原人类视角的景象，因此需要进一步探索图像采集设备与设计在视觉景观评估中的应用。

与此同时，随着生活质量的不断提升，人们对高品质绿地的需求越来越高。这要求城市规划者能够准确识别并改善绿地中的低质量区域。然而，现有研究难以在实践中应用，使得相关评估主要依赖于设计师的个人经验和主观判断，往往忽视了公众的实际需求和偏好。AI技术能有效模拟公众感知并对环境图像进行视觉评估[6][7][10]，具有解决此项难题的潜力，但针对公园这一特定环境类型的基于AI算法的公众感知评价方法尚未得到开发。

本研究旨在建立一种基于城市公园图像采集与深度学习技术的绿地智能感知方法，以实现对公园视觉质量的快速、精准和全面评估，指明低质量区域并为空间的更新改造提供指引。本研究将聚焦以下问题：如何更便捷地采集公园绿地的实景图像？如何设计能够准确反映公众对公园环境的感知和偏好的AI算法系统，从而识别出视觉质量较低的空间？该系统的感知评价结果能够在哪些方面支持视觉景观评估相关理论？这些探索可促进城市景观感知的定量与循证研究的发展，并为城市绿地的更新提供有效的决策建议。

2 研究材料与方法

2.1 研究区域

本文研究区域为位于中国广东省广州市天河区的珠江公园。珠江公园是集生态、游憩、文化于一体的城市公园，具有较高的空间多样性和

丰富的活动类型，占地面积约28hm^2，人气旺盛，为中国亚热带地区十分具有代表性的绿地。

2.2 技术路线

首先，本研究采用了一种便捷的、使用全景相机的公园图像采集方法，并通过人工实地操作方式验证其可行性；接着，利用在ADE20K数据集上训练的Seformer-B5模型自动识别图像中的150类物体，计算绿视率、天空视域因子、路面占比、人工构筑物占比4个客观评价指标；此外，选取吸引力、丰富度、自然程度、压抑程度作为主观评价指标，并通过人工成对比较图像的方式——即每张图像在4个主观指标中被划分为高分值或低分值两类——建立公众感知数据集；基于所构建的公众感知数据集训练ViT-base-p16模型，以实现对主观指标的有效预测；然后，通过可视化呈现主、客观评价指标的空间分布图，识别低分值图像的空间分布区域；最后，通过相关性分析主、客观评价指标间的相关性，为公园改造提供建议（图1）。

2.3 数据收集与处理

图像采集时间为2023年7月6日9:00~13:00，当天天气晴朗，气温约30℃。一位图像采集人员使用Insta360 ONE RS全景相机步行沿珠江公园所有道路进行拍摄，拍摄高度约为1.7m。同时使用智能手持GPS传感器（佳明eTrex 221x）记录拍摄点位的位置信息。依据实践经验，为完整反映视觉信息并提高采集效率，研究团队在道路交叉点、拐点、中点（两个拐点之间道路的中间点）及重要标志物（建筑、亭台、雕塑等）等处进行图像采集，且相邻图像采集点之间的间隔不超过50步（约40m）。本研究共采集275个点位的全景图像，均位于道路中线（图2）。

本研究使用Insta360 Studio进行图像处理，所有图像均清晰可用。随后，研究团队于平铺模式下截取0° 和180° 两个视角的照片共550张，以展示点位四周的场景环境，并在ArcMap 10.6中将照片与GPS空间信息进行数据匹配。

2.4 基于深度学习技术的图像评价方法

2.4.1 使用图像语义分割模型提取客观评价指标

环境中的客观物理要素（包括自然和人工要素）对景观的视觉质量和人们的审美认知有很大的影响。近年来，语义分割技术作为场景理解的关键技术之一，通过像素级别的图像分类显著提升了图像要素识别的精度。

本研究采用目前识别准确率较高的SegFormer-B5模型[13]进行客观物理要素的提取。模型主要包含分层的Transformer编码器和轻量级的All-MLP解码器两个模块。其中，Transformer编码器用于提取图像特征，采用自注意力机制来对输入图像的重要区域进行加权，使模型可以有效捕捉其重要信息，提高图像分割的性能。All-MLP解码器可以直接融合多级特征并预测语义分割掩码，并通过全连接层输出结果。研究基于ADE20K数据集[14]对模型进行训练，该数据集是2016年由MIT开放的场景理解的数据集，包括150个要素类别。测试结果表明，SegFormer-B5模型在ADE20K数据集验证集上的预测表现优于早期提出的FCN、PSPNet、DeepLabV3+[7]等常用模型及先进的FPN、UPerNet模型①。

本研究从150种要素中提取出13种公园场景中常见的视觉要素②，并借鉴现有视觉感知研究计算绿视率和天空视域因子[15]-[17]。前者反映了公园的生态和自然程度，后者则可衡量空间的开放程度[17]。此外，珠江公园中道路与人工构筑物（墙体、座椅、路灯、栏杆等）也较多，故本研究还引入了路面占比和人工构筑物占比两个指标[11]（表1）。

2.4.2 使用图像分类模型预测主观感知分数

在传统的图像主观感知研究中，常用评级测度法、配对比较法、分类法等方法获取受访者的景观感知评价[18]。以李克特五点式量表为例，

① 模型比较数据可通过 OpenMMlab 的 github 网页获取。

② 13 种公园场景常见视觉要素包括墙体、建筑、天空、乔木、灌木、地被、一级道路、二级道路、三级道路、围栏、摩天大楼、座椅、路灯。

表 1：客观评价指标

维度	评价指标	定义	来源
自然	绿视率	植被（乔木、灌木、草地）的像素占比之和	参考文献 [15][16]
	天空视域因子	天空像素占比	参考文献 [17]
人工	路面占比	硬质路面（公园一、二、三级道路）的像素占比之和	参考文献 [11]
	人工构筑物占比	人工结构（墙体、建筑、围栏、摩天大楼、座椅与路灯）的像素占比之和	参考文献 [11]

受访者需从1~5的等级对图像进行评分。在获取评分后，深度学习中的图像分类模型可学习评分与图像特征之间的关系，从而模拟人类感知过程，将图像划分为1~5的等级，实现大规模、快速的主观感知评分。现有的此类研究主要基于大规模城市感知数据集，例如包含来自56个城市11万余张图像的Place Pulse 2.0数据集[19]，通过8万余名在线志愿者对其从不同维度的两两对比获得图像评分，后续研究据此训练图像分类模型以预测主观感知分数[6]。相关研究证明，将主观视觉调查、图像语义分割和图像分类模型相结合可以有效、公正地收集和绘制街道感知情况[6]。虽然该数据集缺乏公园场景图像信息，因而无法直接应用于公园场景中，但其数据集的构建方法及基于此数据集的主观感知预测方法为本研究的主观评分方法奠定了基础。

（1）主观评价指标建立

借鉴传统视觉景观评估研究[20]-[24]，选取吸引力、丰富度、自然程度和压抑程度作为主观评价指标。其中，吸引力指公园场景对人们的吸引程度，包括景观的美观性、独特性等特征[20]。丰富度指公园环境组成元素（包括物种与各类设计元素）的多样性和复杂性[20][21]。自然程度指游客对公园环境在人为干预和自然状态之间平衡程度的感知，测定并理解公园的自然程度感知有助于制定公园维护管理策略[22]。压抑程度指的是令人抑郁、沮丧、消沉的程度[23]，常被用于衡量城市景观对人身心健康的影响[24]，压抑程度高的公园可能会让人感到不适，影响园内体验。

（2）两两对比结果收集

与直接获取被试者的评分数值相比，两两对比是一种更有效、准确的感知获得方式[19]。首先，在尽可能涵盖所有公园场景的前提下，人工判读③550张照片并剔除相似度过高的照片，最终获得200张有效照片。随后，研究团队利用Java Script建立在线评分系统，该系统会根据用户的选择和图像已有的对比关系，动态调整所展示的图片，以保证每张图片均获得充分对比与有效评分。每次对比随机从200张照片中抽取2张（图3），被试者需根据问题（“哪个场景让您感到更有吸引力／丰富／自然／压抑？”）选出更符合个人偏好的图片。每位被试者进行4次实验，每次实验只针对一项指标进行评分；同时为避免疲劳，单次实验的对比次数被要求控制在50次左右，时长不超过10分钟。实验共招募35名华南理工大学风景园林专业为主的在读硕士生（男女比例12:23），均无色盲或色弱，完成为期三天（2024年3月3～5日）的在线实验。最终，四项指标共获得对比结果6 702项，平均每个指标获得结果1 675.5项。

（3）主观评价指标计算

参考现有研究[25]，本文使用“赛程强度”（strength of schedule）方法来统计主观评分，以此获取每个指标的高低得分（图4）。

对于主观评价指标m，研究定义图像i被选择和未被选择频率分别为$W_{i,m}$和$L_{i,m}$：

$$W_{i,m} = \frac{w_{i,m}}{w_{i,m}+l_{i,m}}, \tag{1}$$

$$L_{i,m} = \frac{l_{i,m}}{w_{i,m}+l_{i,m}}, \tag{2}$$

式中，$w_{i,m}$、$l_{i,m}$分别表示在比较中被选择与未被选择的次数。

每个图像i的主观评价指标m的评分分数$Q_{i,m}$可以定义为:

$$Q_{i,m} = W_{i,m} + \frac{1}{n_i^w}\sum_{k_1=1}^{n_i^w} W_{k_1 m} - \frac{1}{n_i^l}\sum_{k_2=1}^{n_i^l} W_{k_2 m}, \tag{3}$$

其中，n_i^w和n_i^l分别表示图像i被选择与未被选择的总次数。为了进一步以低、高划分图像得分$Q_{i,m}$，研究定义了下述二进制标签，$W_{i,m}\in\{0,1\}$，0代表低分值，1代表高分值：

$$W_{i,m} = \begin{cases} 0 & if\ \ Q_{i,m} > \ \mu_m + \sigma_m \\ 1 & if\ \ Q_{i,m} < \ \mu_m - \sigma_m \end{cases}, \tag{4}$$

其中，μ_m和σ_m分别表示感知指标m所有数据的平均值和标准差。

（4）图像分类模型训练

上述200张图片在4项指标上均被赋值为“0”或“1”，从而构成公众感知数据集。图像分类模型可以将这些数值作为标签，以图像作为解释变量进行训练。本研究采用ViT-base-p16模型[26]进行图像分类。ViT-base-p16模型通过将输入图像分解成一系列的图像块，并将每个块作为序列元素输入Transformer模型，利用自注意力机制对输入图像中的关键区域进行加权，从而有效捕捉图像中的重要信息。在训练阶段，ViT-base-p16模型首先使用大规模的ImageNet-1k数据集进行预训练，以学习图像的通用表示，接着在公众感知数据集中分别对各项指标进行微调，最终得到4个模型分别用于预测公园所有图像的吸引力、丰富度、自然程度和压抑程度。

本研究采用5折交叉验证评估模型性能。具体而言，包含200张照片的主观评价数据集被随机划分为5个子集，在每轮训练中，4个子集被作为训练集，余下1个子集作为验证集，计算5轮准确率的平均值来获得总体的性能评估，最终选取准确率最高的模型进行主观指标评分。此方法不仅能够保证模型在训练集上的良好表现，还能提升模型对新数据的泛化能力，从而在有限数据上获得更优的模型表现性能。

2.5 主观与客观指标综合评价

本研究利用训练好的Seformer-B5与ViT-base-p16模型，分别对550

③ 人工判读指图像采集者依据主观感知和个人经验筛选图像，无量化指标。

张图片进行主观与客观评价指标的计算，并取每个点位的两张图片指标的平均值作为该点位的最终指标值。研究首先在ArcMap 10.6中对这些点位的数值进行可视化，生成主、客观评价指标的空间分布图，以识别低分值图像的分布区域。由于数据不符合正态分布，随后采用Spearman相关性分析来检验主、客评价指标与在图片中占比较多的植物（乔木、灌木、地被）以及道路（一、二、三级道路）之间的相关性。

3 研究结果与讨论

3.1 客观评价指标提取结果

图5列举了经Segformer-B5模型语义分割后图像中不同景观元素的构成情况，表2为4个客观评价指标的统计信息。4项指标中，各点位绿视率平均值最高（0.7115，其中乔木占比0.3973、灌木占比0.1691、地被占比0.1450），表明珠江公园中的植被条件十分优秀，构成了公园景观的主要骨架。相应地，高植被覆盖度也导致天空视域因子较低（平均值为0.0737），由于树冠、植被的浓密程度，公园的天空可视范围较低。路面占比和人工构筑物占比的低平均值则揭示了公园以自然景观为主导的特征。这两项结果的中位数均较接近平均值，表明大部分区域中路面和建筑物的分布较少；同时，标准差数值较低也表明了公园内部结构的相对均匀性，游客体验较为一致。

3.2 主观评价指标预测模型训练结果

主观评价指标的5折交叉验证数据分布情况与模型预测准确率结果显示（图6），模型的预测准确率在不同指标上有所波动，但整体表现较为稳定。测试集的平均准确率依次为69%（吸引力）、70.5%（丰富度）、82%（自然程度）、68.5%（压抑程度），预测结果可信度较高。

主观指标评分统计结果（表3）显示，自然程度的均值最高，表明珠江公园的自然程度在人类感知中较为突出，与语义分割结果可相互佐证：该指标的数值范围最大（0.0443 ~ 0.8855），平均值和中位数接近，且标准差最高，反映出公园的植被分布存在较大的空间异质性。吸引力和压抑程度的均值也较高，表明自然程度较高的公园场景整体上具有较强的吸引力，但过于茂盛的植被可能增加压抑感；二者的标准差相似且适中，表明这两个指标在样本中的分布离散性较为适中。相比之下，丰富度的分布更为集中，标准差较低，数值范围较窄（0.0732 ~ 0.5826），说明该指标在样本中的差异性相对较小；其均值也最低，表明公园视觉要素的多样性较低，与自然程度的较高波动性形成对比，表明虽然不同场景中自然程度变化较大，但视觉元素的丰富性相对不足，公园在提升景观多样性方面仍有提升空间。

3.3主观与客观指标综合评价结果

总体而言，珠江公园主、客观指标的得分的空间分布模式较为相似（图7，8）。公园西门入口前的草坪区（C区）主要由开阔的草坪与低矮的乔木组成，辅以少量灌木，整体空间较为开阔、道路较为宽敞，天空占比较高，绿视率与自然程度较低，同时吸引力偏低。公园中部的快绿湖区域（E区），虽然绿视率较低、天空占比较高，但吸引力较高，符合人们对水体有普遍偏好的既有研究发现[5]。公园东部为植被茂盛的风景林区（F区），绿视率与自然程度较高，在广州繁华的市中心十分难得，该区域道路曲折蜿蜒、起伏多变，路面与人工构筑物占比较低，总体吸引力也较高。此外，公园西南侧部分区域植被茂盛且空间变化多样，丰富度较高，不同点位场景差异较大，吸引力也有所不同。而位于公园东侧的服务型建筑立面较为单一（G区），吸引力偏低，需要公园管理者重点关注。

Spearman相关性分析结果（图9）显示，首先，自然程度与吸引力之间存在显著的正相关（r_s=0.60），表明自然程度高的场景更受人们的喜爱，与前人研究结果相符，即游客偏好自然植被丰富的环境，这可能会积极影响游客的公园使用频率和满意度[27]。其次，丰富度与地被占比呈负相关关系（r_s= - 0.48），意味着地被的增加可能导致整体丰富度的降低。例如，珠江公园中以地被为主的区域（画面中地被占比较高），集

表 2：客观评价指标结果

客观评价指标	最小值	最大值	平均值	中位数	标准差
绿视率	0.0000	0.9731	0.7115	0.7351	0.1630
天空视域因子	0.0000	0.2815	0.0737	0.0643	0.0557
路面占比	0.0000	0.4237	0.1236	0.1011	0.0910
人工构筑物占比	0.0000	0.3787	0.0286	0.0127	0.0462

表 3：主观评价指标结果统计

主观评价指标	最小值	最大值	平均值	中位数	标准差
吸引力	0.0783	0.7540	0.4285	0.4485	0.1476
丰富度	0.0732	0.5826	0.2841	0.2724	0.0942
自然程度	0.0443	0.8855	0.4303	0.3776	0.2021
压抑程度	0.1619	0.8710	0.3821	0.4120	0.1468

中在公园西侧，且以开阔草坪为主，空间丰富度较低。自然程度与绿视率（r_s=0.71）、乔木（r_s=0.47）、灌木（r_s=0.46）占比存在显著正相关趋势。其中，自然程度和绿视率分别反映了图像中主观和客观的生态环境状况，而主观自然程度的感知不仅受到植被占比的影响，还涉及其他影响因素，如绿色元素在图像中的整体构成比例、画面中呈现的其他材料（如水体、泥土、透水路面）等。对于压抑程度，自然程度（r_s=0.64）和灌木（r_s=0.65）占比与之呈显著正相关，表明植被郁闭度高的地方可能会增加人们的压抑感。

此外，天空视域因子、道路占比和人工构筑物占比之间存在一定的正相关性，且三者与4个主观感知指标以及绿视率均存在负相关关系。这意味着天空、道路和人工构筑物占比的提高可能与植被数量的减少及自然程度的下降有关。在珠江公园这一以自然景观为主的公园中，天空、道路、建筑、墙体、座椅等占比较多的区域，如西北侧的儿童游乐区（B区）和西门入口前的草坪区（C区）植被较少、道路宽敞、视野开阔，这些区域的吸引力与自然程度可能不如植被茂盛的东侧风景林区，但由于环境开阔，压抑感也有所降低。

4 结语与展望

《欧洲景观公约》指出，景观是一项重要的、应予认可和保护的公共权益[28]，深入了解人们对景观的观察和感知，并将其纳入景观规划和管理至关重要。本研究采用视觉评估领域前沿的图像采集及AI技术，尝试构建以景观感知为核心的景观学研究与实践方法论。总体而言，本研究在以下三个方面展现出积极意义：

1）本研究采用了一套便捷且高效的工作流程，通过全景相机采集城市绿地的实景图像，进一步结合先进的语义分割和图像分类模型对公园视觉质量进行无差别评估。这种方法克服了传统视觉景观评估中难以高效评估大批量图像与多场景评估时易于出现的视觉疲劳问题，验证了基于图像大数据与深度学习技术的城市感知研究在绿地中应用的有效性。

2）传统视觉评估研究中所采用的图像数据量通常较小，且缺乏对主、客观要素的准确量化。本研究通过精准提取并评估客观指标与预测主观评分，发现植物与水体的存在有助于提升公园场景的吸引力与使用者的积极感知，而天空、道路与人工构筑物占比过高则会产生负面效果。

3）传统的研究结果难以直接应用于新场景的偏好预测。本研究采用的智能方法，通过部分场景照片主观评分的学习，能够预测新场景的主观评分，帮助公园管理者高效识别出低分值区域，为城市绿地的更新设计提供指引，具有较强的实践应用价值。

然而，本研究经由人工采集与筛选的图像数据与被试者的样本量相对有限，且仅针对珠江公园的夏季场景，难以预测与珠江公园差异较大或其他季节的公园场景。同时值得注意的是，本研究结果显示，公园西北侧的儿童游乐区的吸引力偏低，这可能与受访对象的偏好有关：此区域植被较少且丰富度较低，对于本次实验的受访对象——大学生来说吸引力较差。这也揭示了以往基于街景大数据、通过全体居民的普遍偏好进行模型训练的研究存在局限性[19][25]，难以有效反映不同群体的多样化需求。此外，全景相机采集的图像可能存在一定畸变，会导致一定误差。未来研究中应补充更多绿地与不同季节的景观图像，并纳入更多使用者的评分，以完善绿地感知数据集，并有必要针对不同群体的偏好进行细化研究。

最后，本研究所提取出的主、客观感知评价指标，可以与公园活力、功能使用、环境质量等多元数据结合，进一步挖掘公众感知与公园的景观吸引力、使用者行为模式和公园客观物理特征等方面的关系，系统化地为城市管理者提供科学的决策支持，帮助其制定更精准的策略和方法，优化公园功能配置，提升城市景观质量。