*forests*

MDPI

*Article*

# IntegratingDeep Learning with Urban Greenery: Analyzing Visual Perception Through Street View Images in Tianjin, China

**Yu-Xiang Sun [1], Yuan-Yuan Sun [2,*], Qian Ji [2], Zi-Tong Zhao [2], Yan-Kui Yuan [1], Sheng-Bei Zhou [1] and Feng-Liang Tang [3]**

[1] International School of Engineering, Tianjin Chengjian University, Tianjin 300380, China; sunyuxiang7735@163.com (Y.-X.S.); yankui_yuan@163.com (Y.-K.Y.); shengbeizhou@163.com (S.-B.Z.)
[2] School of Architecture, Tianjin Chengjian University, Tianjin 300380, China; jiqian005@hotmail.com (Q.J.); zzttjcj99@163.com (Z.-T.Z.)
[3] Department of Urban Planning, School of Architecture, Tianjin University, 92 Weijin Road, Xuefu Street, Tianjin 300072, China; tang_fengliang@tju.edu.cn
* Correspondence: kaiyuan@tcu.edu.cn

**Abstract**

Rapid urbanization has intensified the demand for street designs that reconcile ecological quality with positive human experiences, particularly in high-density cities such as Tianjin, China. Streets function as key interfaces where ecological processes, social activities and human perception intersect. However, existing research tends to emphasize the amount of greenery while overlooking its structural characteristics, to treat perception as a psychological response decoupled from spatial context, and to make limited use of fine-grained functional data to examine how ecology and perception interact. This study develops an integrated analytical framework that combines the DeepLabV3+ model to extract the Urban Street Greenery Generalized Structure (USGGS) from Baidu Street View imagery with a vision transformer model trained on the Place Pulse 2.0 dataset to derive multidimensional perceptual metrics. Functional diversity is represented using point-of-interest (POI) data, and an enhanced Light Gradient Boosting Machine (LightGBM) model is employed to explore associations among greenery structure, perceived qualities and functional characteristics. Analyses of six urban districts in Tianjin indicate that ecological and perceived street qualities are closely related to the degree of coupling between vegetation structure and functional diversity. Streets characterized by multi-layered greenery and diverse, active functions tend to exhibit higher perceived aesthetics, safety and vitality, whereas streets with single-layer vegetation or functionally monotonous environments generally do not perform as well. Functional patterns appear to mediate relationships between greening and perception by shaping how ecological form is experienced through everyday social activities. Overall, the results suggest that closer coordination between ecological design and functional organization is important for fostering urban streets that combine environmental resilience with strong perceived appeal.

**Keywords:** urban street greening generalized structure (USGGS); street view imagery; human perception; urban planning; improved LightGBM model; sustainable ecology

Check for updates

Academic Editor: Nikolaos Proutsos

Received: 11 November 2025
Revised: 7 December 2025
Accepted: 15 December 2025
Published: 26 December 2025

**Copyright:** © 2025 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.

## 1. Introduction

Urbanization has changed the ecological and visual environment of cities [1,2]. The fast growth of built areas and the loss of vegetation have reduced many natural functions of cities [1,3]. Street greening, which includes trees, shrubs, and grass along roads, is

an important part of urban green infrastructure [4,5]. It helps to absorb carbon, lower temperature, improve air quality, and support biodiversity [5–8]. At the same time, street greenery enhances the visual quality of streetscapes and provides shade and thermal comfort for pedestrians. Given these functions, understanding the structure of street greenery is crucial for improving urban environmental quality and residents' quality of life [4]. As cities become denser, the spatial layout and plant composition of street greenery are becoming more complex. In many places, the form of street greening does not always match the needs of surrounding spaces or the daily experience of residents [9]. People's visual perception of streets is also closely related to how greenery is arranged and how urban functions are distributed [10]. Therefore, clarifying the relationship among street greening structures, human perception, and the spatial pattern of city functions is essential for creating healthier and more livable urban environments.

Accurately and consistently quantifying the structure of street greening remains challenging because street spaces contain complex forms and mixed vegetation layers [11]. Many studies have attempted to characterize urban street greenery [12,13], yet most face clear limitations. Traditional field surveys can record plant types and vertical layers in detail, but they are time- and labor-intensive and thus difficult to apply at large scales. Satellite and aerial imagery expand spatial coverage but cannot capture the fine structure of trees, shrubs and grass along narrow streets. To obtain a closer, human-scale view of urban greenery, researchers have increasingly used street-view imagery [14–16], which enables more precise identification of vegetation layers and visible greenery. For example, Xia et al. [17] developed a panoramic green index based on street-view images, improving the evaluation of greenery coverage and quality. Deep learning and computer-vision methods have further enhanced the accuracy of such analyses. In 2023, Zhang et al. [18] created the SSGS dataset by combining pixel-level labels of trees, shrubs and grass with street-view imagery, providing a major breakthrough for large-scale, fine-grained quantification of street-greening structures. Subsequent studies using this dataset have achieved better performance in mapping urban greenery and distinguishing vegetation types. Nevertheless, most existing work still concentrates on the physical distribution and classification of vegetation, without relating structural characteristics to broader spatial context or human perceptual dimensions.

Because most existing studies focus only on the physical form of street greenery, linking structural and perceptual perspectives has become increasingly important. Many scholars have explored how people perceive urban streets using different types of data and analytical methods [19]. The Place Pulse project collected a large number of street images and asked people to compare their sense of safety, beauty, and wealth, producing one of the first global datasets on urban perception [20,21]. These data were later used to train deep learning models that predict perception scores directly from visual information [22,23]. Other studies applied convolutional neural networks and semantic segmentation to identify urban elements such as trees, buildings, and roads and to examine how these elements influence visual comfort and attractiveness. Some researchers used social media images and online text to analyze emotional responses to urban environments, while others combined remote sensing and street view data to link greenery with subjective well-being [24,25]. These approaches have provided valuable insights into how visual conditions shape people's experiences in cities and have made perception a measurable attribute of the urban environment [26,27]. Yet, several important limitations remain. Most perceptual studies describe general feelings but rarely analyze the structural features of vegetation in detail, often emphasizing aesthetic impressions without explaining the spatial or ecological mechanisms behind them [28]. Many perception datasets are based on images from a limited number of cities, which constrains their ability to represent diverse urban

forms and cultural contexts. The models used are also sensitive to image quality and sampling bias [29]. In addition, few studies have integrated perceptual measures with fine-grained urban data that reflect land use, population density, or activity intensity [30,31]. Without this connection, it is difficult to explain how street greening interacts with the social and spatial structures of the city and how these relationships influence the visual and ecological quality of urban spaces.

To better understand the relationship between street greening and perception, it is important to include fine-grained urban data in the analysis. Among different forms of fine-grained data, points of interest (POI) are one of the most direct and flexible indicators for describing urban functions [32,33]. Each POI represents a specific activity or facility and abstracts multidimensional urban information into a single spatial point, reflecting both the spatial pattern and social attributes of the city. Because POI data can capture the diversity of urban functions and human activities, examining how street-greening structures and perceptual characteristics interact with them is essential for revealing the spatial organization of cities [34,35]. However, most existing studies still rely on traditional socioeconomic indicators such as population density, income level or housing price to analyze how urban structure affects green space or perception [36,37]. Although these indicators provide valuable insights, they are usually coarse in spatial resolution and cannot reflect the detailed functional composition of urban areas. In contrast, POI data provide fine-grained information about where different activities occur and how urban spaces are organized, offering a more accurate basis for examining the links among street greening, perception and urban structure [38]. At the same time, urban data often show nonlinear and complex relationships that traditional regression models cannot fully capture [39,40]. Machine learning methods are therefore well suited for modeling these interactions and identifying how multiple variables jointly influence street greening and perception. In this context, the SHAP (SHapley Additive exPlanations) framework can be used to quantify the contribution of each variable and to interpret model results in an explainable way [41–43].

Given the ecological and social importance of street greening and the limitations of previous research, there is a need for a more integrated analysis of how street-greening structures relate to visual perception and urban functions. Existing studies have largely focused on the overall quantity or coverage of greenery, treated human perception as an isolated outcome, and rarely incorporated fine-grained urban functional data. To address these gaps, this study takes Tianjin as a representative case and develops an urban street greening generalized structure (USGGS) derived from street-view imagery using deep learning. We link USGGS to perception scores generated by a deep learning model and to points-of-interest (POI) data that characterize urban functional spaces. By applying machine learning models and interpretable analyses, we identify the key factors shaping the interactions between physical greenery, human perception and urban structure. This integrated framework provides new empirical evidence on how street environments influence both ecological benefits and visual experiences, thereby supporting more sustainable and human-centered urban design.

## 2. Study Area and Data

### 2.1. Study Area

Tianjin is one of the four municipalities directly under the Central Government of China. It is located on the northern coast of the country and serves as an important economic and cultural center in the Bohai Rim region [44]. The city has a temperate monsoon climate with hot, humid summers and cold, dry winters. In recent decades, rapid urbanization has increased the density of buildings and reduced green space, creating challenges for maintaining urban ecological quality [45]. This study focuses on the six central districts

of Tianjin: Heping, Hexi, Hebei, Nankai, Hedong, and Hongqiao Figure 1. These districts form the core urban area of the city, with dense populations, diverse land uses, and complex street networks. They include the main residential, commercial, and cultural zones, which makes them suitable for studying how street greening structures relate to visual perception and urban functions. Focusing on this area helps reveal the interactions between ecological and social factors in a typical high-density urban environment and provides a basis for improving the quality of street spaces in Chinese metropolitan cities.

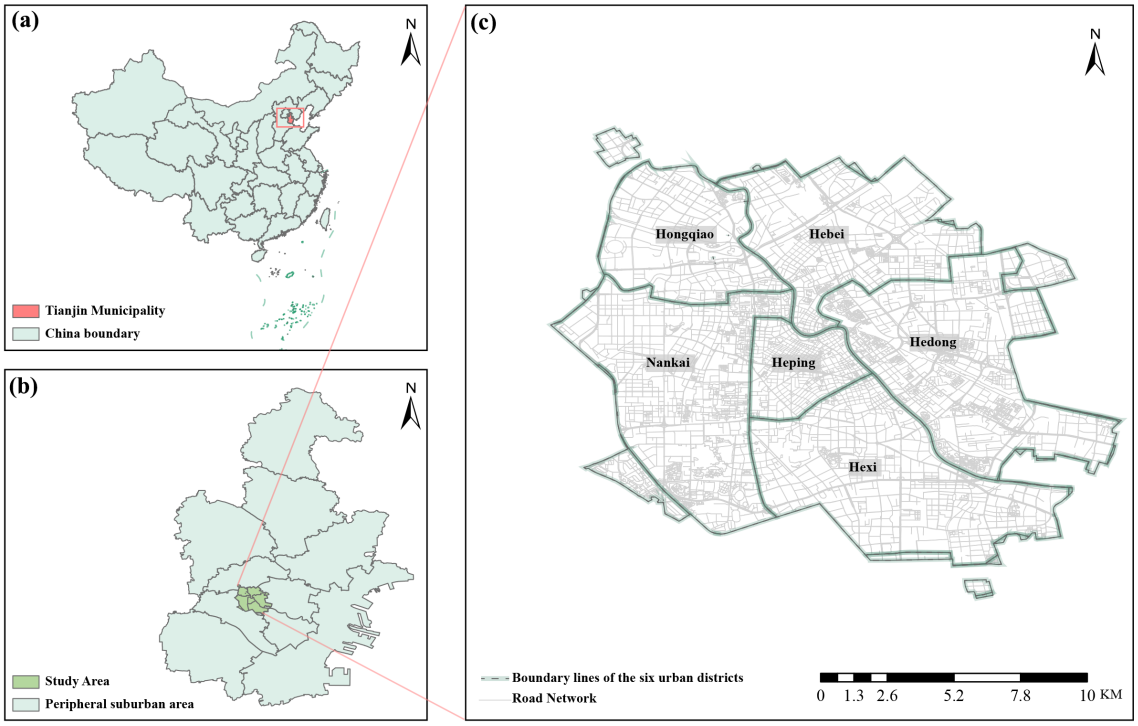



**Figure 1.** Study area location and extent. (**a**) Location of Tianjin within China and its administrative boundary; (**b**) location of the study area within Tianjin Municipality, highlighting the central urban districts versus peripheral suburbs; (**c**) boundaries and road network of the six central urban districts.

### *2.2. Data Sources*

The data used in this study include street view images, perceptual data, and fine-grained urban data. Street view images of Tianjin were obtained from the Baidu Street View (BSV) platform, which provides panoramic images with accurate geographic coordinates. Based on the road network data from Gaode Map (Amap),sampling points were generated at 50 m intervals within the six central districts of Tianjin. A spacing of 50 m was chosen as a compromise between spatial detail and computational efficiency, ensuring that each typical street block was represented by multiple viewpoints while keeping the total number of images manageable for deep-learning training and semantic segmentation. In total, about 120,000 sampling points were obtained, and for each point, four panoramic images were collected to cover different street directions. A total of about 120,000 sampling points were selected, and for each point, four panoramic images were collected to cover different street directions Figure 2. These images were used to extract the structural characteristics of street greenery through semantic segmentation. The Cityscapes and Street Greening Structure (SSGS) datasets were used to train and validate the DeepLabV3+ neural network model. Cityscapes provides pixel-level labeled street scenes for pretraining, while SSGS contains detailed annotations of trees, shrubs, and grass, improving segmentation accuracy for greenery extraction.

Perceptual metrics were derived from a deep learning model based on a ViT-Transformer architecture, fine-tuned on the Place Pulse 2.0 (PP2.0) dataset. PP2.0 provides large-scale pairwise comparisons of urban street images, in which volunteers judge which image appears safer, more affluent, more monotonous, more oppressive, more vibrant, or more aesthetically pleasing. Using these comparisons, we constructed continuous perceptual scores along six dimensions: sense of safety, sense of affluence, sense of monotony, sense of oppression, sense of vibrancy, and sense of aesthetics. For each street view image in Tianjin, the model predicts a numerical value for each of the six dimensions, reflecting the relative intensity of these perceptual attributes in the scene. The resulting scores are normalized and then averaged across the four directional images at each sampling point to obtain six street-level perceptual indicators. These indicators are subsequently georeferenced to build the spatial perception layer used in later analyses.Detailed urban functional data were obtained from the AutoNavi Open Platform, which provides POI information representing diverse urban functions (e.g., residential, commercial, educational, and recreational facilities). Each POI record includes a functional category and geographic coordinates, allowing precise characterization of the urban functional layout. By integrating the perceptual indicators with POI data, we establish a comprehensive analytical framework for examining the relationships between street greening structures, perceived environmental qualities, and the spatial configuration of Tianjin.

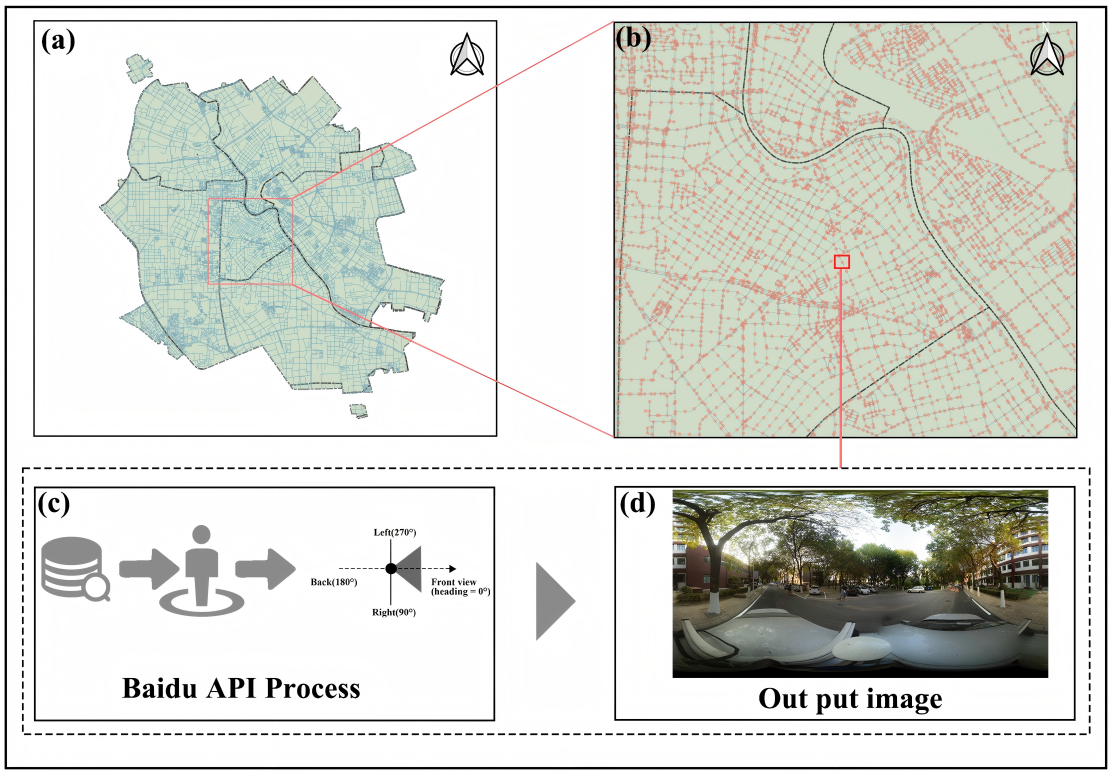



**Figure 2.** Street-view sampling and retrieval workflow: (**a**) road network with sampling grid; (**b**) zoomed-in sampling points; (**c**) Baidu Street View API processing; (**d**) example street-view image.

### *2.3. Data Processing*

The data processing consisted of two main parts: the acquisition of street view imagery and the collection of points of interest (POI) data. For the street view images, the road network data of Tianjin were obtained from the Amap open platform. Based on the road network, a total of 16,637 sampling points were generated at 50-m intervals within the six central districts using ArcGIS 10.8 and Python 3.8. Each point was assigned geographic

coordinates and a unique identifier, which were then used to request panoramic images from the Baidu Street View (BSV) API. Four directional images (north, south, east, and west) were captured at each point to ensure complete street coverage. After data cleaning and the removal of invalid or duplicate records, 106,240 valid street view images were obtained, covering most urban streets. In addition, 13,280 panoramic street view images were acquired to provide a broader perspective of the street environment. These images collectively formed a high-resolution dataset for quantifying the Urban Street Greening Generalized Structure (USGGS) through vegetation extraction and classification.

The POI data were obtained from the Amap open platform through the Amap Web Service API. A bounding box covering the six central districts of Tianjin was defined, and the region was divided into small tiles to ensure complete retrieval. For each tile, POI information was collected, including name, category, and geographic coordinates. After cleaning to remove duplicate and irrelevant records, a total of 116,417 valid POI points were retained. The data were converted into georeferenced point shapefiles using GeoPandas for spatial analysis. The dataset includes fourteen main categories such as catering, traffic facilities, healthcare, education, and leisure, which describe the fine-scale urban functions of Tianjin's core area. These variables, summarized in Table 1, were used to quantify functional diversity and to explore the relationships among street greening structure, perception, and urban functionality.

**Table 1.** Variables of POI Functional Types, Street Greening Structures, and Perceptual Indicators.

| Variable | Abbreviation | Description | Category |
|---|---|---|---|
| **Functional POI Variables (X)** | | | |
| X1 | C-A | Catering | Restaurant, café, dining facilities |
| X2 | C-E | Corporate Enterprises | Offices, companies, business centers |
| X3 | S-C | Shopping Consumption | Retail stores, malls, daily goods shops |
| X4 | T-F | Traffic Facilities | Transport stations, parking, infrastructure |
| X5 | F-I | Financial Institutions | Banks, ATMs, insurance services |
| X6 | H-A | Hotel Accommodation | Hotels, hostels, lodging places |
| X7 | S-E-C | Science, Education, and Culture | Schools, libraries, museums |
| X8 | T-A | Tourist Attractions | Scenic spots, tourist destinations |
| X9 | A-R | Automotive Related | Gas stations, repair shops, car services |
| X10 | C-R | Commercial Residence | Mixed commercial –residential buildings |
| X11 | L-S | Life Services | Daily services, salons, laundry, repair |
| X12 | L-E | Leisure and Entertainment | Cinemas, bars, KTVs, recreation venues |
| X13 | H-C | Healthcare | Hospitals, clinics, pharmacies |
| X14 | S-F | Sports and Fitness | Gyms, stadiums, fitness centers |

**Table 1.** *Cont.*

| Variable | Abbreviation | Description | Category |
|---|---|---|---|
| **Street Greening Structures and Perceptual Indicators (Y)** | | | |
| Y1 | T-B-G | Tree–Bush–Grass | Complex three-layer street greening structure |
| Y2 | T-B | Tree–Bush | Two-layer street greening structure |
| Y3 | T-G | Tree–Grass | Two-layer street greening structure |
| Y4 | B-G | Bush–Grass | Two-layer street greening structure |
| Y5 | S-T | Single Tree | One-layer street greening structure |
| Y6 | Safer | Safety Perception | Perceived safety level of streets |
| Y7 | Liveli | Liveliness Perception | Perceived activity and vitality |
| Y8 | Boring | Boring Perception | Perceived monotony or dullness |
| Y9 | Beauti | Beauty Perception | Perceived visual attractiveness |
| Y10 | Wealth | Wealth Perception | Perceived affluence and prosperity |
| Y11 | Depres | Depressing Perception | Perceived gloom or discomfort |

## 3. Methodology

### *3.1. DeepLabV3+ Model and Training*

Figure 3 This study outlines its research framework and analytical workflow. Previous research on street greening extraction has applied various deep learning models to semantic segmentation tasks, including fully convolutional networks (FCN), SegNet, and PSPNet. FCN introduces an end-to-end architecture for pixel-level prediction, effectively mapping image features to segmentation masks. However, its multiple pooling operations can lead to loss of spatial details, thereby reducing segmentation accuracy at object boundaries. SegNet enhances spatial detail through a symmetric encoder-decoder structure, where the decoder restores feature map resolution using the encoder's pooling indices. While SegNet demonstrates superior edge retention, it suffers from high computational complexity and struggles with multi-scale object detection. PSPNet introduces pyramid pooling to capture global contextual information, enhancing understanding of large-scale scenes. However, this model remains inefficient and exhibits limited adaptability for fine vegetation segmentation in dense urban environments. Overall, while these models make significant contributions, they still face limitations in accurately identifying mixed vegetation layers—such as trees, shrubs, and grass—within complex street spaces.

To overcome these limitations, this study adopted the DeepLabV3+ model [46], a state-of-the-art convolutional neural network designed for precise pixel-level segmentation of complex urban scenes Figure 4. DeepLabV3+ builds upon the DeepLab series and combines a powerful encoder–decoder architecture with atrous convolution and multi-scale contextual learning. The encoder captures high-level semantic information through a series of atrous convolutions, which enlarge the receptive field without decreasing spatial resolution. This allows the model to recognize vegetation objects of varying scales

while maintaining fine boundary details. The decoder gradually restores spatial resolution through upsampling and convolution operations, refining edges and improving segmentation accuracy along object boundaries such as between trees and buildings or between shrubs and sidewalks.

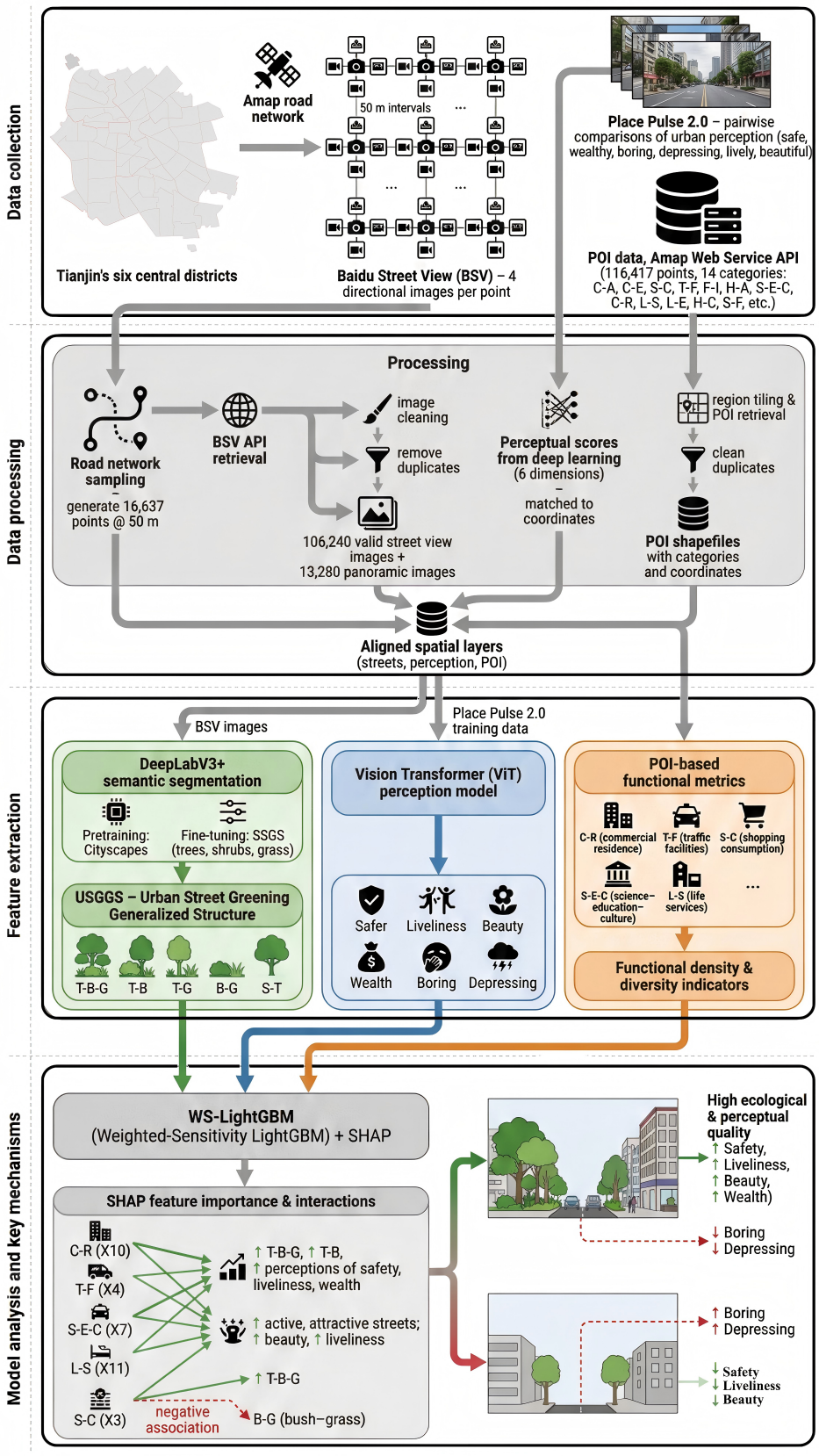



**Figure 3.** Research Framework.

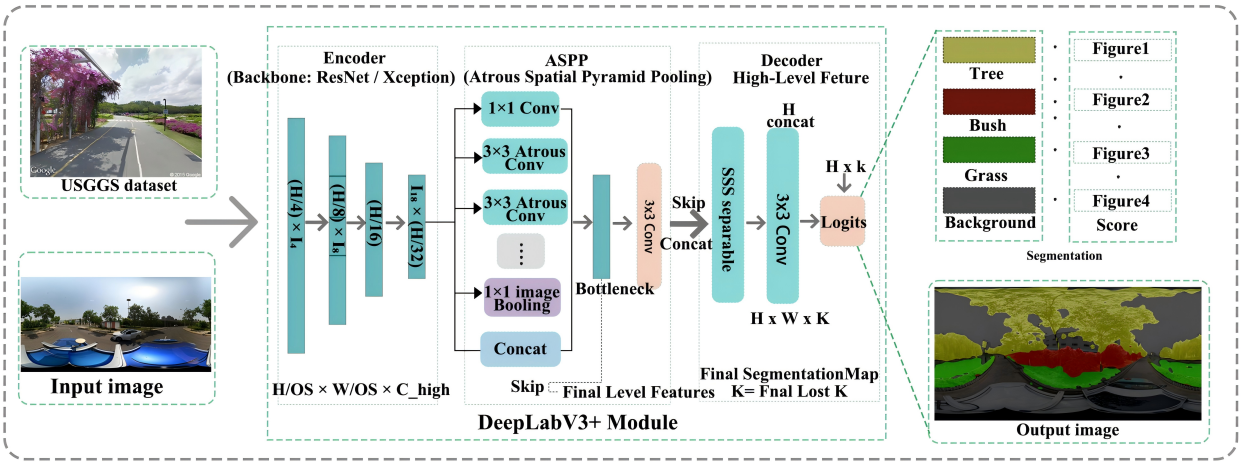



**Figure 4.** A Street View Vegetation Structure Semantic Segmentation Framework Based on DeepLabv3+ Model.

Within the encoder, DeepLabV3+ integrates multiple convolutional layers with batch normalization and activation functions to extract hierarchical features from low-level textures to high-level semantics. The inclusion of atrous convolution enables multi-scale feature extraction while keeping computational costs moderate. The operation of atrous convolution can be defined as:

$$y[i] = \sum_{k=1}^{K} x[i + r \cdot k]\, w[k], \tag{1}$$

where $x$ is the input feature map, $w$ is the convolution kernel, $K$ is the kernel size, $r$ is the dilation rate, and $y[i]$ is the output feature at position $i$. This operation enlarges the receptive field and enables the model to capture multi-scale contextual information that is crucial for vegetation segmentation in street scenes.

To further enhance multi-scale feature extraction, DeepLabV3+ incorporates the Atrous Spatial Pyramid Pooling (ASPP) module, which applies parallel atrous convolutions with different dilation rates to aggregate contextual information at multiple scales. The output feature of the ASPP module can be represented as:

$$F_{\text{ASPP}} = \text{Concat}(f_{r_1}(x), f_{r_2}(x), \ldots, f_{r_n}(x)), \tag{2}$$

where $f_{r_i}(x)$ represents the feature map obtained with dilation rate $r_i$, and $F_{\text{ASPP}}$ denotes the aggregated multi-scale feature representation. This process helps the model recognize both small shrubs and large trees across diverse urban street environments.

During training, DeepLabV3+ minimizes the pixel-wise cross-entropy loss between the predicted class probabilities and the ground truth labels to optimize segmentation accuracy. The loss function is expressed as:

$$L = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{C} y_{i,c} \log(p_{i,c}), \tag{3}$$

where $N$ is the number of pixels, $C$ is the number of classes, $y_{i,c}$ is the true label, and $p_{i,c}$ is the predicted probability. This loss function enhances the model's ability to produce accurate and consistent segmentation results for complex urban scenes.

DeepLabV3+ demonstrates several advantages that make it particularly suitable for this research. Its atrous convolution and ASPP modules allow for multi-scale contextual

learning, enabling accurate identification of vegetation structures across varying spatial resolutions. Compared with traditional models such as SegNet and PSPNet, DeepLabV3+ achieves higher segmentation accuracy, stronger boundary recognition, and faster convergence, while requiring fewer parameters. Its encoder–decoder architecture ensures both semantic richness and spatial precision, which are essential for distinguishing hierarchical vegetation layers in street images. As shown in Table 2, DeepLabV3+ outperforms the other two models in terms of mIoU, precision, and computational efficiency, confirming its suitability for vegetation segmentation in complex urban environments. The detailed mathematical derivations of the atrous convolution, ASPP module, and cross-entropy loss functions are provided in Appendix A.1, which formalizes how multi-scale contextual features are captured and optimized during model training.

**Table 2.** Performance comparison among different neural network models.

| Model | Epoch | mIoU (%) | mPA (%) | Recall (%) | Precision (%) | Per Second (s) |
|---|---|---|---|---|---|---|
| DeepLabV3+ | 1800 | 63.78 | 73.12 | 73.10 | 80.96 | 0.25 |
| PSPNet | 1800 | 56.02 | 70.68 | 70.55 | 66.21 | 0.47 |
| SegNet | 1800 | 54.11 | 68.93 | 68.74 | 64.85 | 0.55 |

Model training was performed using the SSGS dataset, which was specifically developed for street greenery research. The dataset contains pixel-level annotations of trees, shrubs, and grass across various urban scenes, including residential, commercial, and campus areas. These detailed labels allow the model to learn the hierarchical relationships among vegetation layers and to distinguish fine structural differences in urban green elements. Training with SSGS improves the robustness and precision of the segmentation model, enabling accurate quantification of the urban street greening general structure (USGGS) from Baidu Street View images in Tianjin. This approach provides a reliable foundation for analyzing how different vegetation configurations relate to visual perception and urban functions in subsequent stages of the study.

### *3.2. Vision Transformer Model and Perception Prediction*

In recent years, many studies have attempted to quantify urban perception from street view images by using convolutional neural network (CNN) architectures such as VGGNet, ResNet, and EfficientNet. These models have achieved remarkable success in image classification and feature extraction. VGGNet is characterized by its simple and deep architecture, which captures hierarchical visual features effectively but requires a large number of parameters and computational resources. ResNet introduces residual connections that alleviate the vanishing gradient problem and enable deeper networks to learn complex visual semantics; however, it still relies on local receptive fields and struggles to capture long-range spatial dependencies. EfficientNet improves parameter efficiency by balancing network depth, width, and resolution, but its local convolutional nature limits its ability to integrate global contextual information. Overall, CNN-based models perform well in object-level recognition but often overlook broader spatial relationships and visual coherence, which are crucial for human perception of urban streets.

To overcome these limitations, this study employed the Vision Transformer (ViT) model trained on the Place Pulse 2.0 (PP2.0) dataset Figure 5. The Vision Transformer is a deep learning architecture that adapts the Transformer framework, originally developed for natural language processing, to visual recognition tasks [47]. Unlike CNNs that process local receptive fields, ViT divides an image into a sequence of patches and models their global relationships through a self-attention mechanism. This design enables the model to capture both global and local contextual dependencies, which is particularly valuable for understanding the overall composition and perceptual qualities of urban streetscapes.

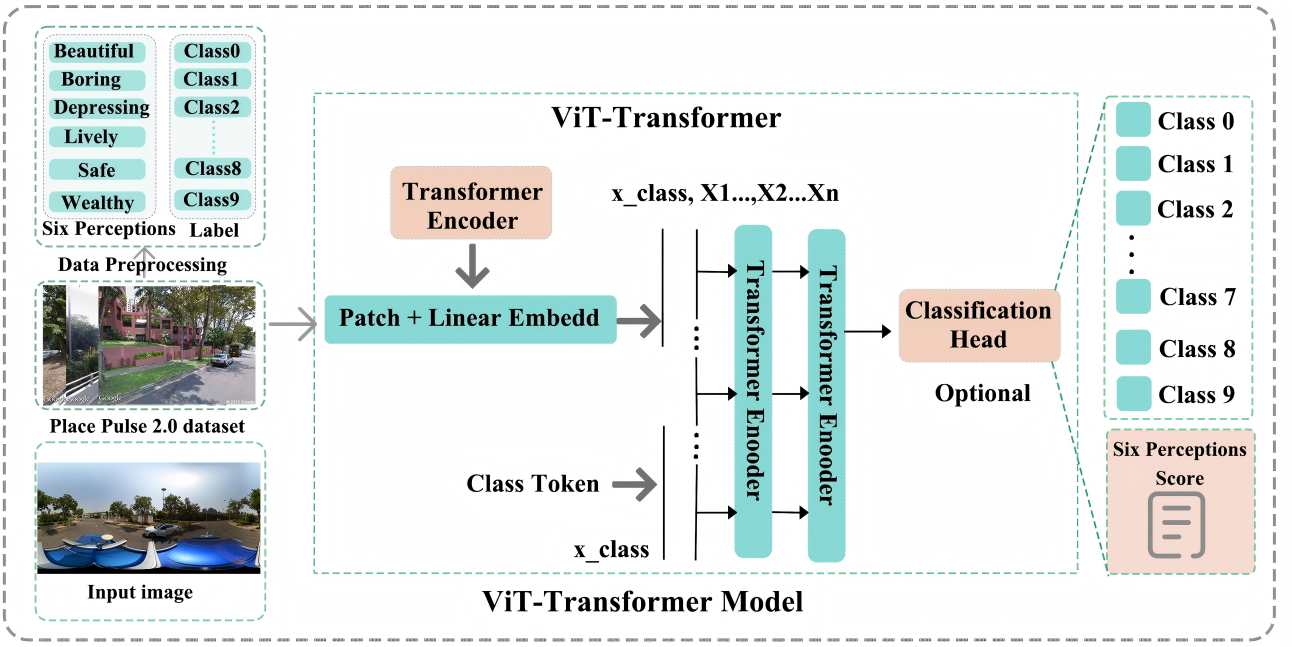



**Figure 5.** ViT-Transformer Perceptual Classification Model Framework.

The structure of ViT consists of three main parts: the patch embedding module, the transformer encoder, and the classification head. In the first stage, an input image $x \in \mathbb{R}^{H \times W \times C}$ is divided into $N$ non-overlapping patches of equal size $P \times P$. Each patch is flattened and linearly projected into a $D$-dimensional embedding vector. A learnable positional embedding is added to preserve spatial order, forming a sequential representation $z_0$ that serves as input to the transformer encoder. The patch embedding process can be expressed as:

$$z_0 = [x_p^1 E;\ x_p^2 E;\ \ldots;\ x_p^N E] + E_{pos}, \tag{4}$$

where $x_p^i$ denotes the $i$-th image patch, $E$ is the learnable linear projection matrix, and $E_{pos}$ is the positional encoding vector. This embedding operation transforms the spatial image into a one-dimensional token sequence, allowing the model to process images in a manner similar to language sequences.

The embedded sequence is then passed through a stack of $L$ identical transformer encoder layers. Each layer includes a multi-head self-attention (MHSA) module and a feed-forward network (FFN) with residual connections and normalization. The self-attention mechanism calculates the pairwise similarity between all image patches, allowing each patch to attend to all others. The MHSA operation is formulated as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{5}$$

where $Q$, $K$, and $V$ represent the query, key, and value matrices derived from the input embeddings, and $d_k$ is the dimension of the key vector. By computing weighted attention across all spatial tokens, the transformer learns global dependencies and context-aware feature representations. Multi-head attention extends this operation by applying multiple attention heads in parallel, which capture diverse feature relationships and improve representational richness.

In the final stage, a special classification token $z_{cls}$ is added to the sequence before entering the transformer encoder. This token interacts with all other patch tokens and

aggregates information across the entire image. The final representation of this token is used for classification or regression tasks. The output prediction can be defined as:

$$y = \text{softmax}(W_{cls}\, z_{cls}), \tag{6}$$

where $W_{cls}$ is the weight matrix of the classification head, and $y$ denotes the predicted probabilities for each perceptual dimension. The model was trained to predict six perceptual attributes from the PP2.0 dataset: safe, wealthy, boring, depressing, lively, and beautiful.

The Vision Transformer offers several advantages that make it particularly suitable for this study. Its self-attention mechanism enables global feature learning, allowing the model to capture long-range dependencies that CNNs cannot easily represent. This is especially beneficial for urban street images, where human perception depends on the overall spatial layout, openness, and vegetation composition rather than individual objects. The ViT model has strong generalization ability and can adapt to different urban contexts without extensive retraining, due to its parameter-sharing and attention-based representation. Moreover, it avoids the inductive bias inherent in convolution operations, enabling more flexible learning of complex spatial relationships between visual elements such as trees, roads, buildings, and sky. The model also achieves competitive accuracy with fewer parameters when pre-trained on large-scale datasets and fine-tuned on specific perception tasks. The detailed mathematical derivations of the patch embedding, multi-head self-attention, and output head are provided in Appendix A.2, which formalize how the Vision Transformer processes visual information and generates perception predictions.

In this study, the ViT model was fine-tuned using the PP2.0 dataset, which provides large-scale pairwise comparisons of urban street scenes with human-labeled perceptual scores. Each image includes annotations on six perceptual dimensions, enabling the model to learn robust mappings from visual features to perception attributes. The self-attention layers help the network capture global spatial semantics, while the transformer encoder integrates detailed textural and color information relevant to visual impressions. This structure aligns well with the goal of quantifying how street composition and greening patterns affect human perception. By using ViT, the model can effectively extract both holistic and detailed visual cues from street environments, producing reliable perceptual predictions that reflect multidimensional human experiences in Tianjin.

### *3.3. Integration of Greening Structure, Perception, and POI Data*

To analyze the relationships among street greening structure, visual perception, and urban functions, this study employed an improved Light Gradient Boosting Machine (LightGBM) model. Previous studies on urban environment evaluation have commonly used traditional statistical and machine learning approaches such as multiple linear regression (MLR), random forest (RF), and eXtreme Gradient Boosting (XGBoost). Linear regression models are simple and interpretable, but they assume linear relationships among variables and cannot capture complex nonlinear interactions. Random forest models improve predictive performance through bagging multiple decision trees but require high computational cost and offer limited interpretability. XGBoost enhances the efficiency and accuracy of gradient boosting by introducing second-order derivatives and regularization, yet it remains computationally expensive for large, high-dimensional urban datasets such as street view imagery and fine-grained POI data.

LightGBM, developed by Microsoft Research, is a gradient boosting framework based on decision tree learning that optimizes both computational efficiency and scalability. It employs a histogram-based algorithm and a leaf-wise tree growth strategy that selects the leaf with the highest information gain at each iteration. The objective function of gradient boosting can be defined as:

$$L(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{t=1}^{T} \Omega(f_t), \tag{7}$$

where $l(y_i, \hat{y}_i)$ is the loss function measuring the difference between the predicted value $\hat{y}_i$ and the true value $y_i$, and $\Omega(f_t)$ is the regularization term controlling model complexity across $T$ boosting iterations. Each new decision tree $f_t(x)$ is added to minimize the overall loss by fitting the negative gradient of the objective function.

The standard leaf-wise gain function in LightGBM is expressed as:

$$\text{Gain} = \frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} - \gamma, \tag{8}$$

where $G_L$ and $G_R$ denote the first-order gradients of the left and right child nodes, $H_L$ and $H_R$ are their corresponding second-order gradients, and $\lambda$ and $\gamma$ are regularization parameters to prevent overfitting.

To better adapt the model to heterogeneous urban datasets that combine physical (greening), perceptual, and functional (POI) indicators, we introduced a feature-sensitive weighting mechanism. The improved gain function becomes:

$$\text{Gain}_{\text{WS}} = \alpha_i \left[ \frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma, \tag{9}$$

where $\alpha_i$ is a dynamic sensitivity coefficient for feature $i$, defined as:

$$\alpha_i = 1 + \beta \cdot \frac{\sigma_i}{\bar{\sigma}}, \tag{10}$$

with $\sigma_i$ representing the standard deviation of feature $i$, $\bar{\sigma}$ the mean standard deviation across all features, and $\beta$ a hyperparameter controlling sensitivity adjustment. This weighting mechanism increases the influence of features with greater dispersion or interpretive relevance, enhancing the model's responsiveness to key factors such as perceptual scores or vegetation composition.

In addition to this structural improvement, Bayesian Optimization was used to fine-tune hyperparameters including learning rate, maximum depth, and number of leaves. The optimization process minimizes the validation loss:

$$\theta^* = \arg\min_{\theta \in \Theta} \mathbb{E}_{(x,y)\sim D}[L(y, f(x;\theta))], \tag{11}$$

where $\Theta$ represents the hyperparameter search space, and $f(x;\theta)$ is the model output parameterized by $\theta$. This ensures stable convergence and reduces overfitting.

Finally, SHapley Additive exPlanations (SHAP) were applied to interpret the marginal contribution of each feature to the model output. The SHAP value for feature $i$ is defined as:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} \left[ f(S \cup \{i\}) - f(S) \right], \tag{12}$$

where $F$ denotes the full set of features, and $f(S)$ the model output using subset $S$. SHAP enhances interpretability by quantifying the individual and interactive effects of greening structure, perception, and POI-related features.

Through these internal and external enhancements, the improved LightGBM model [48], referred to as the Weighted-Sensitivity LightGBM (WS-LightGBM), achieves both higher predictive accuracy and better interpretability. It effectively captures nonlinear interactions among ecological, perceptual, and urban functional dimensions, providing a

robust analytical framework for understanding the multi-dimensional mechanisms shaping Tianjin's urban streetscape. The detailed mathematical derivations of the second-order objective, weighted-sensitivity gain function, and Bayesian optimization process are presented in Appendix A.3, which formally illustrate how the WS-LightGBM model integrates feature sensitivity and interpretability in the analysis.

## 4. Results

### *4.1. Spatial Distribution and Pattern of Street Perception*

The distribution of street perception in Tianjin shows clear spatial variation across the six perceptual dimensions of safer, lively, boring, beautiful, wealthy, and depressing. As shown in Figure 6, these perceptual indicators form distinct clusters rather than uniform patterns. High perception values are mainly concentrated in the central and southern parts of the study area, where urban construction density and functional diversity are greater. Streets in these areas often have wider spaces, active commercial activities, and better maintenance. In contrast, the northern and western parts of Tianjin display lower perception scores. These districts contain older residential neighborhoods, narrow lanes, and limited greenery. The overall distribution suggests that the visual quality of urban streets is closely related to the level of development and functional intensity. The pattern also reflects the influence of urban form and socio-economic structure within the central city [49].

The six perceptual dimensions present different spatial characteristics. Safety and liveliness show similar patterns with high scores along major commercial corridors and near cultural and public spaces. These areas have strong street visibility, clear boundaries, and frequent pedestrian movement. Streets with more balanced spatial composition and moderate levels of greenery tend to be perceived as both safer and more vibrant. The boring and depressing dimensions display an opposite tendency [50]. High boring scores appear mainly in homogeneous residential areas where building forms are repetitive and vegetation diversity is low. Depressing values are higher in dense and aging neighborhoods where building maintenance is poor and public spaces are limited. The beauty perception shows a more balanced distribution across the study area but with clear spatial clustering along riverfront areas and historical streets. These streets often combine tree canopies, open views, and continuous architectural facades, creating a visually pleasing street environment. Wealthiness shows a similar spatial trend to beauty and safety. Higher values appear in modern residential and business districts, while older neighborhoods present relatively lower scores. The spatial overlap among beauty, safety, and wealthiness indicates that these dimensions may share common environmental characteristics related to order, maintenance, and greenery.

The overall pattern of street perception in Tianjin suggests that visual experience is strongly structured by spatial and functional differences within the city. Streets that are open, active, and well maintained usually produce positive perceptions such as safety, liveliness, and beauty. In contrast, streets that are enclosed, monotonous, or visually deteriorated are often linked to boredom and depression. These patterns show that perception can be regarded as a sensitive indicator of spatial quality and social conditions in urban areas. The uneven distribution of perceptual attributes also highlights the importance of incorporating visual assessment into urban planning and design. Understanding how perception varies across different urban contexts provides a foundation for further analysis of the relationships among street greening, visual experience, and urban function in the following sections.

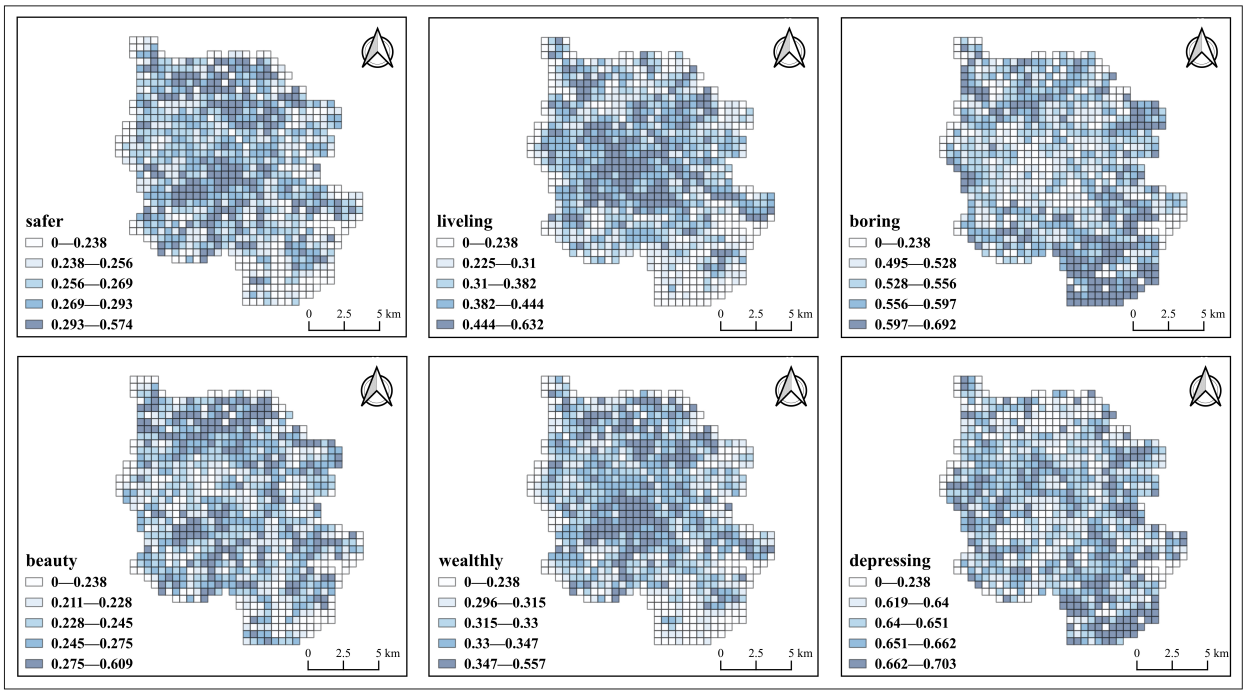



**Figure 6.** Spatial distribution of street perception in Tianjin across six perceptual dimensions: safer, lively, boring, beautiful, wealthy, and depressing.

### *4.2. Spatial Heterogeneity of Generalized Street Greening Structures*

The spatial distribution of street greening in Tianjin shows clear variation across different types of vegetation structures. As shown in Figure 7, five key indicators describe the overall pattern of urban street greenery: the Single Tree structure (S-T), the Tree–Bush–Grass composite structure (T-B-G), the Bush–Grass ratio (B-G), the Tree–Bush ratio (T-B), and the Tree–Grass ratio (T-G). Overall, the greening system demonstrates a mixed and uneven spatial distribution. The S-T and T-B-G patterns are more continuous and extensive than the others, reflecting the dominance of these two vegetation configurations in the city. Areas with higher greening values are mainly located in the southern and eastern parts of the study area, where street width and open spaces allow for more diverse planting arrangements. In contrast, the northern and western parts show relatively lower greening indices, which corresponds to older neighborhoods and compact street networks. The general pattern indicates that the structural composition of street greenery is influenced by urban development history and spatial form [18].

The S-T structure represents streets dominated by individual trees along sidewalks. This type of greening is distributed widely throughout the city but shows higher density in central residential zones and along arterial streets. These areas benefit from consistent street layouts that allow continuous tree planting. The T-B-G structure, which integrates trees, bushes, and grass layers, shows high values in planned communities and near major parks. This configuration provides a more complete ecological structure and a better shading effect. The B-G structure shows more scattered and localized distribution, with higher values in small-scale public spaces and along greenways where shrubs and grass coexist without tall tree canopies. The T-B and T-G ratios show complementary patterns. High T-B values occur in areas with dense roadside planting and limited ground vegetation, while T-G values are higher in wider streets and public squares where lawns dominate. The combination of these patterns suggests that different greening structures are related to specific street functions and physical conditions.

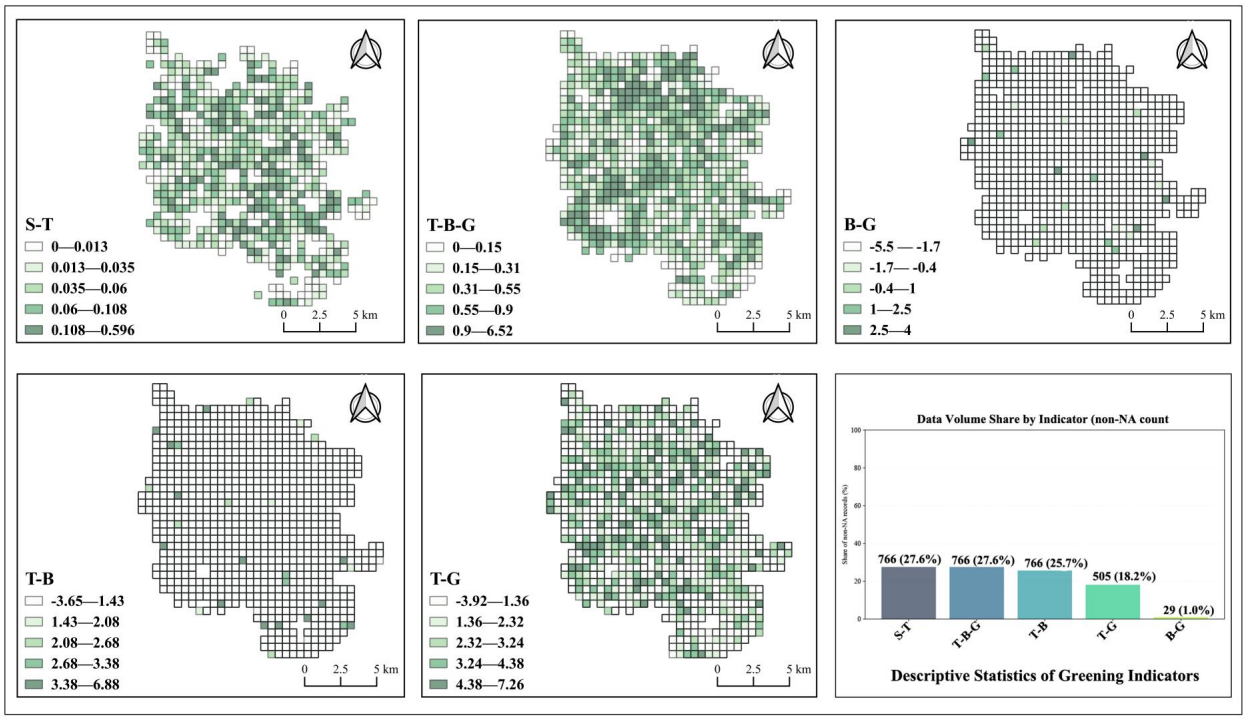



**Figure 7.** Spatial distribution of street greening structure in Tianjin across five vegetation configurations: S-T, T-B-G, B-G, T-B, and T-G.

In general, the greening structure of Tianjin's streets presents a complex but systematic spatial pattern. The southern and eastern parts of the city show more diverse and multilayered vegetation forms, while the northern and western parts are dominated by single-layer or simplified greening types. This pattern reflects the combined influence of planning era, construction density, and available public space on vegetation composition. Streets with composite greening systems such as T-B-G are often associated with high visual quality and better environmental comfort, while streets with single-layer systems such as S-T or sparse vegetation tend to appear less continuous and less ecologically effective. The spatial heterogeneity of greening structures provides an essential ecological background for understanding how street vegetation interacts with visual perception and urban functions, which will be further explored in the subsequent regression analysis.

### *4.3. Functional Coupling Between Street Greening Structures, Perceptions, and Fine-Scale POI Features*

The regression analysis examines how street greening structures and perceptual indicators are jointly influenced by urban functional features derived from POI data. In the models, both greening structures and perception indices were treated as dependent variables ($Y$), while the POI-based indicators served as independent variables ($X$). This design enables direct comparison of how physical and perceptual attributes of streets respond to variations in local functional diversity. The results show stable explanatory power, with a mean cross-validated $R^2$ of 0.58 across all targets. The model for Liveliness achieved the highest accuracy with an $R^2$ of 0.72, followed by Tree–Bush (T-B) with 0.68 and Wealth with 0.63. Safety and Depressing reached $R^2$ values of 0.55 and 0.52. These outcomes demonstrate that POI-based functional variables can effectively explain both vegetation configurations and perceptual dimensions. Activity-related perceptions and multi-layer greening structures respond more strongly, while emotional perceptions such as safety and depressing feelings show weaker but consistent responses. The results indicate that

both structural and perceptual qualities of streets are governed by common mechanisms associated with land-use intensity and functional complexity.

The overall pattern of coefficients and feature importance reveals strong consistency across all models. Five predictors—C-R (X10), T-F (X4), L-S (X11), S-E-C (X7), and S-C (X3)—emerge as the most influential factors. These variables represent commercial, transport, leisure, service, and cultural functions, which are central components of urban activity. Their recurrent presence among top predictors suggests that a small set of functional indicators governs both ecological layering and perceptual quality. Streets with higher concentrations of commercial and leisure activities tend to support more complex greening structures such as Tree–Bush–Grass (T-B-G) and generate stronger perceptions of beauty, liveliness, and safety. Streets with limited functions or low activity intensity are typically characterized by single-layer greening forms such as S-T and show weaker perceptual evaluations. This pattern suggests a strong coupling between ecological structure and socio-functional activity, where functional diversity supports both vegetation richness and positive perceptual outcomes.

The SHAP beeswarm analysis (Figures 7 and 8) further clarifies the mechanisms behind these relationships by quantifying the marginal contributions of individual POI-based features. Among all predictors, C-R (X10) and T-F (X4) have the largest overall SHAP values and display consistent positive effects across both greening and perception models. High levels of commercial and transport functions increase the probability of complex vegetation and enhance perceptions of liveliness, wealth, and safety. S-E-C (X7) also contributes positively, indicating that service and consumption activities maintain active street environments, which enhance ecological and perceptual conditions through frequent use and upkeep. L-S (X11) exerts a strong influence on perception-related models, particularly for Beauty and Liveliness, suggesting that leisure and social spaces play a key role in improving visual comfort. The S-C (X3) variable positively affects T-B-G but negatively relates to B-G, implying that cultural and institutional functions favor tree-dominant compositions rather than simple shrub–grass patterns. Together, these results highlight that urban functions related to mobility, commerce, and leisure exert the most consistent positive effects on both vegetation structure and perceptual quality.

The partial dependence analysis (Figure 9) provides further evidence by illustrating how variations in major predictors influence model outputs across their full value ranges. The dependence curve for C-R (X10) shows a steady positive relationship with both T-B-G and Liveliness, indicating that increases in commercial density correspond to richer vegetation layering and stronger positive perception. The effect of T-F (X4) follows a similar trajectory, with greening complexity and perceptual quality improving alongside transport intensity but stabilizing after reaching a threshold. L-S (X11) presents a nonlinear pattern: perception values rise rapidly when leisure facilities are moderately distributed but plateau in areas where they are overly concentrated. S-E-C (X7) maintains a consistent upward trend across both greening and perception targets, reflecting the importance of active service environments for sustaining high-quality street spaces. S-C (X3) shows a mixed response, strengthening T-B-G under moderate cultural intensity while providing diminishing benefits when cultural land uses dominate the surroundings. These partial dependence relationships suggest that balanced functional composition and moderate clustering are critical for achieving both ecological and perceptual benefits.

The regression results as a whole demonstrate that the diversity and intensity of urban functions directly shape both the physical and perceptual characteristics of street environments. Streets with diverse and balanced functional mixes tend to exhibit multi-layer vegetation and stronger positive perceptions such as beauty, liveliness, and safety. Streets dominated by single or low-intensity uses display simpler vegetation forms and lower

perceptual performance. These outcomes confirm that ecological complexity and social vitality are interdependent attributes of urban streets. Two main insights can be drawn from these results. First, functional diversity and greening complexity jointly determine the ecological and perceptual quality of urban streets, showing that improvements in one dimension reinforce the other. Second, a small and consistent group of urban functional variables exerts influence across both greening and perception models, which implies that enhancing local functional diversity can simultaneously promote ecological layering and positive visual experience. Together, these findings provide an integrated understanding of how functional composition, vegetation structure, and human perception interact to shape the character and quality of urban street spaces.

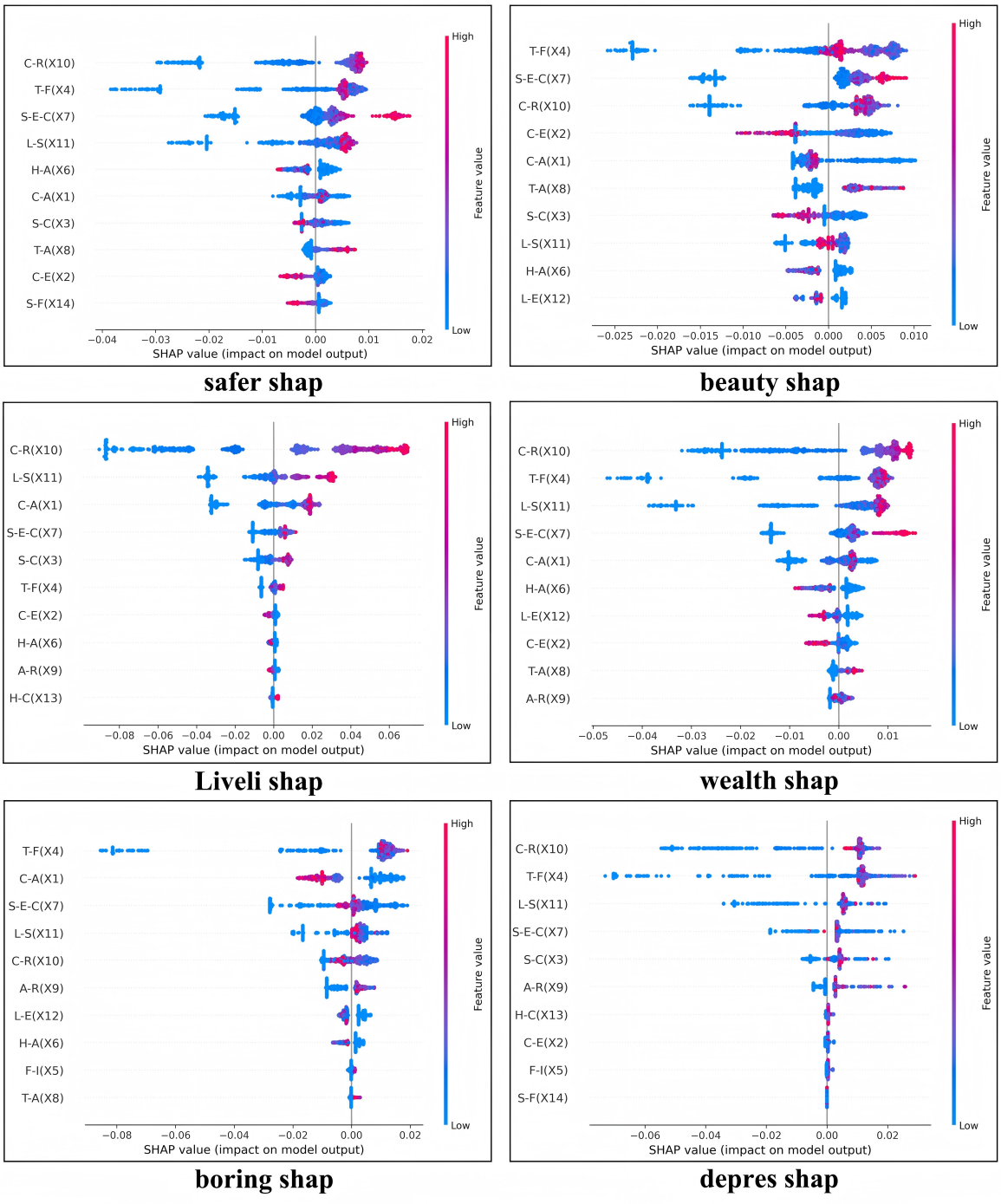



**Figure 8.** The SHAP swarm plot illustrates the contribution and direction of POI features within the perceptual model. Each subplot represents a primary model category.

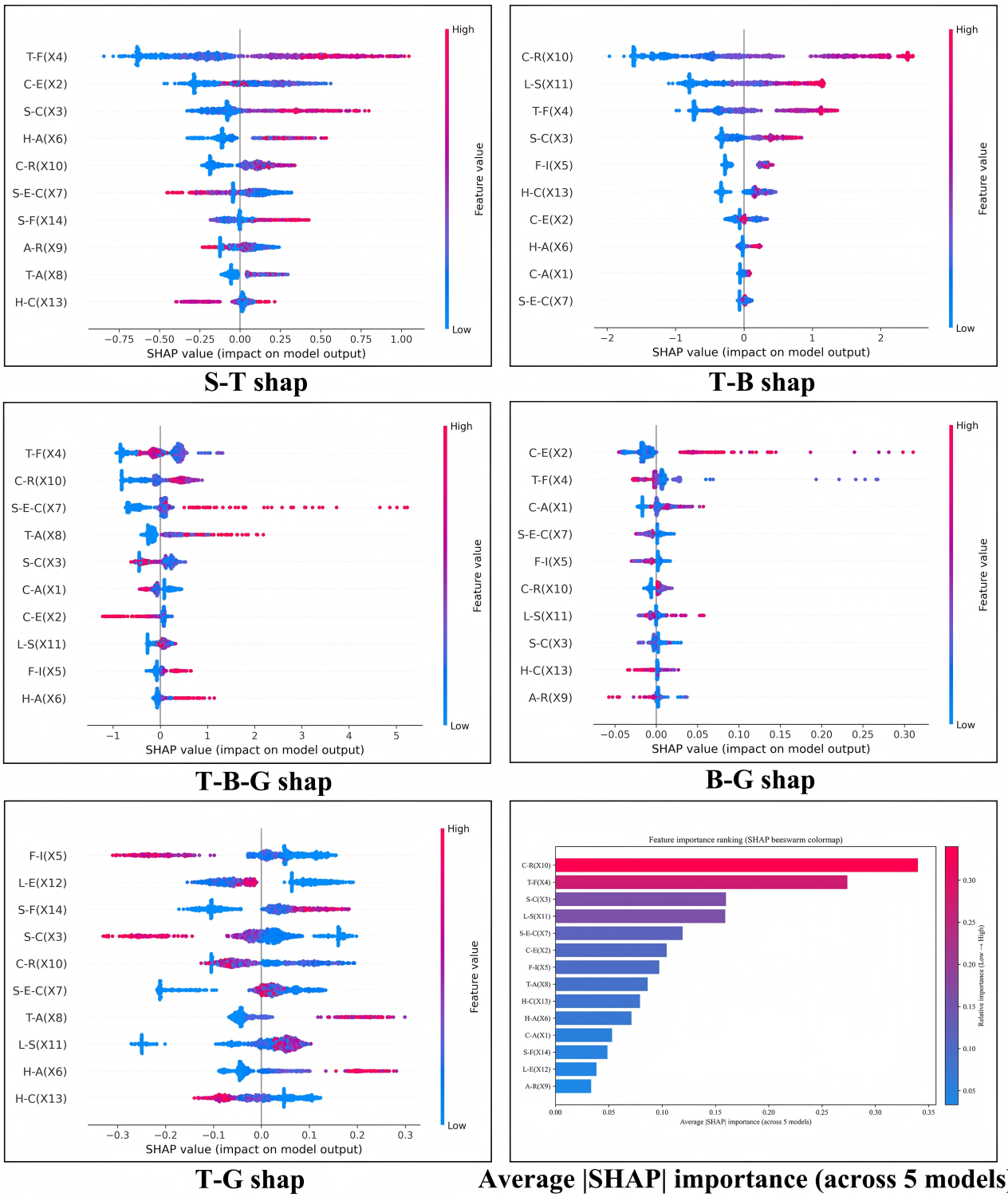



**Figure 9.** The SHAP swarm plot illustrates the contribution and direction of POI features in the greening model. Each subplot represents a primary model category.

## 5. Discussion

### *5.1. Interdependence of Functional Diversity and Greening Complexity in Shaping Street Quality*

Our empirical results show that functional diversity and greening complexity tend to co-occur and jointly shape perceived street quality. Streets dominated by multi-layer structures such as Tree–Bush–Grass (T–B–G) and Tree–Bush (T–B) are mainly concentrated

in the core districts with high densities of commercial, service and cultural POIs, and they exhibit higher levels of perceived beauty, safety and liveliness Figures 8–10. In contrast, streets characterized by Single Tree (S–T) or sparse vegetation are more common in areas with lower functional intensity and are associated with higher scores for boredom and depression. The WS-LightGBM results further indicate that POI variables such as commercial–residential, life services and traffic facilities are simultaneously important predictors of both multi-layer greening and positive perceptual outcomes. Together, these patterns suggest that ecological and social processes in urban streets evolve together rather than independently, and that street-greening structure is closely tied to the functional rhythms of everyday life.

The coexistence of diverse functions such as commerce, culture and services provides not only physical diversity but also the social foundation for ecological maintenance [51]. Streets with multiple uses attract continuous flows of people, which encourage investment in planting and care and help ensure that greenery remains visible and well-managed [52]. Where such interaction is absent, greenery often deteriorates into fragmented or ornamental patches with limited ecological or perceptual value [53]. Our findings that multi-layer plant systems are more prevalent in functionally mixed and perceptually positive corridors, whereas single-layer or sparse vegetation corresponds to areas of functional monotony, confirm that vegetation structure reflects the social rhythm of the city. Human activity produces environmental attention and care, while structured vegetation provides shade, comfort and identity that support continued use. The street, therefore, becomes a shared ecological and social space in which maintenance, perception and structure reinforce one another [54–56]. The influence of this interdependence extends beyond visual composition to affect how people feel and behave in urban environments [57–59]. Vegetation alters light, texture and sound, creating microclimates that shape sensory experience [60], while functional diversity amplifies these effects by giving them context and social meaning. This helps explain why, in our analysis, streets with multi-layer greening and high commercial–service intensity tend to score higher on liveliness and safety: people and greenery coexist within a coherent spatial system [61,62]. In contrast, disconnected systems weaken this link. A green corridor without activity can appear empty or insecure, while a busy commercial street without vegetation often feels harsh and uncomfortable. The balance between functional activity and vegetation layering thus determines how a street is used and remembered. Feelings of beauty or safety are not inherent qualities of greenery alone but expressions of how physical structure, social presence and spatial order work together [59,62,63]. Understanding perception in this way transforms it from a purely psychological response into a measurable reflection of environmental organisation.

These findings carry important implications for planning and design. First, they indicate that improving street quality cannot be achieved by treating greenery and function as separate concerns: planting additional trees without considering surrounding land use is unlikely to change perception or ecological efficiency, just as increasing commercial activity without greenery can diminish comfort and long-term vitality [64,65]. Second, they suggest that effective design depends on aligning ecological complexity with functional diversity so that social functions and ecological forms support one another. Planning should link vegetation layering to pedestrian movement, local services and public spaces, encouraging continuity between natural and social systems [66]. Streets designed through this lens can achieve both environmental stability and perceptual richness, creating urban environments that are sustainable not only in structure but also in human experience. In this sense, street quality emerges from the constant negotiation between what is natural and what is social, and the most resilient environments are those where ecological complexity and human diversity evolve together.

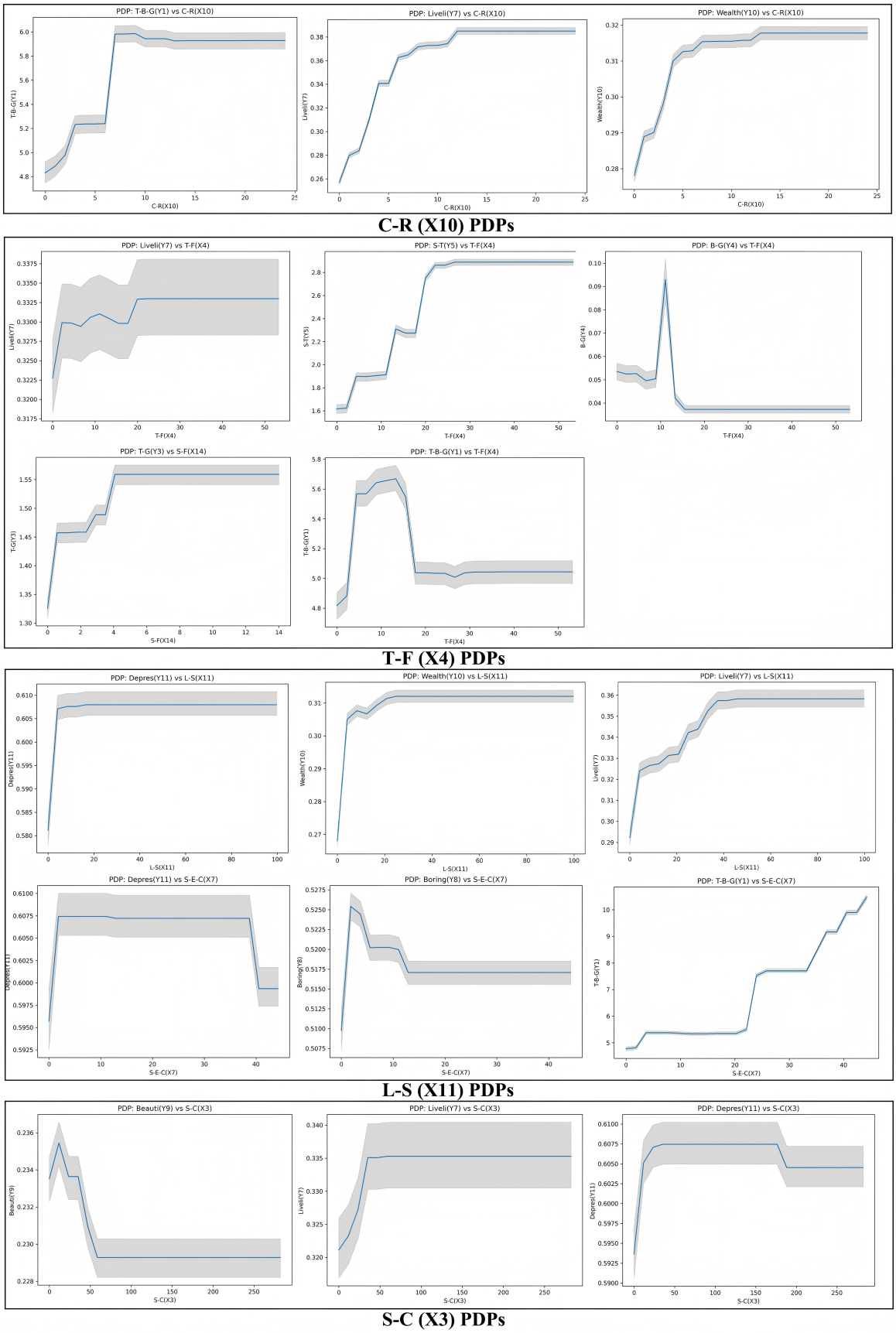



**Figure 10.** Partial dependence plots of key POI-based features showing marginal effects on greening and perception targets. Each subfigure corresponds to one major functional variable.The blue line indicates the estimated partial dependence (marginal effect), and the grey shaded band represents the 95% bootstrap confidence interval.

### 5.2. Functional Mediation of POI Features in Linking Greening and Perception

The connection between greening structure and human perception is rarely direct [67]. Our modeling shows that this relationship is strongly filtered through the social and functional organization of the city. In the WS-LightGBM models, a small set of POI-based indicators—particularly commercial–residential, life services, traffic facilities, science–education–culture and shopping consumption—repeatedly emerge among the top contributors for both greening structures (e.g., T–B–G, T–B) and perception indices such as liveliness, beauty, wealth and safety Figures 7–9. Liveliness shows the highest predictive performance, followed by T–B and Wealth, while Safety and Depressing have lower but still consistent $R^2$ values. Partial-dependence patterns indicate that moderate levels of these functions are associated with stronger ecological layering and more favorable perceptions, whereas overly concentrated single functions yield diminishing benefits. These results indicate that functional composition acts as a mediating layer that translates physical greening into perceived street quality.

Vegetation that appears identical in form can therefore produce very different perceptual responses depending on the urban context in which it is placed [68]. A street lined with trees beside active shops and community facilities conveys order and vitality, while the same vegetation along an underused corridor may seem isolated or even neglected. These differences arise because urban functions provide the framework that gives ecological form its meaning. The variety and intensity of local activities shape how greenery is seen, used and maintained. Where commercial, cultural and service functions are well distributed, people engage with their environment more frequently and interpret vegetation as part of the collective landscape [69,70]. In contrast, areas with limited functional support reduce both physical care and social attention, causing greenery to lose its symbolic and perceptual value. The interaction between ecological structure and functional composition therefore operates as a mediating process, transforming the physical presence of greenery into a social experience of comfort, beauty and belonging [71,72].

The mediating influence of functional composition works through multiple spatial and behavioral pathways. Functionally diverse streets generate continuous activity and movement, which improve environmental visibility and maintenance. People walking, shopping or resting near greenery provide passive surveillance and foster a sense of safety. Regular social use also encourages municipal and community investment in care, reducing decay and strengthening ecological health. Leisure and cultural functions play an equally important role. Parks, cafés and small cultural spaces invite people to pause and interact, extending the time spent in contact with vegetation. This prolonged engagement deepens the perceptual connection between ecological form and human activity, allowing greenery to become part of daily experience rather than a static visual background. Spatial proximity also matters: the closer greenery is to pedestrian routes and active façades, the more likely it contributes to perceptions of comfort and safety [73]. Functional elements therefore act as catalysts that reveal or suppress the perceptual potential of ecological features.Overall, two key insights emerge from this subsection. First, the perceptual impact of street greenery depends strongly on the functional context in which it is embedded; identical vegetation structures can generate very different perceptions under different activity patterns. Second, fine-grained functional data such as POI provide an effective proxy for these mediating processes, enabling researchers and planners to move beyond simple land-use categories toward a more nuanced understanding of how daily activities shape the greening–perception relationship.

### 5.3. Policy and Planning Implications

Improving the quality of urban streets requires recognizing that ecological form, human perception, and functional activity are parts of the same system. Policies that treat greenery as decoration or land use as an independent category cannot respond to the complexity revealed by this interaction. The interdependence between vegetation structure and functional diversity means that street environments perform best when ecological and social processes are designed to support one another. Planning should therefore begin with an understanding of how vegetation is maintained through human presence and how functional patterns shape the visibility and meaning of greenery. Streets with diverse land uses encourage continuous use and informal surveillance, which promote both safety and maintenance. Vegetation placed in these contexts acquires social relevance because it becomes part of collective daily life. In contrast, greenery installed in functionally inactive areas quickly deteriorates, losing ecological value and perceptual appeal. Recognizing this dependency is a first step toward policies that approach greening as an active system sustained by participation, rather than a passive amenity applied through top-down design.

Building on this understanding, the improvement of urban street environments should focus on creating balance between ecological structure and social function. Design interventions need to move beyond single-objective solutions such as expanding canopy coverage or increasing commercial density. The emphasis should instead be on spatial coordination, where vegetation, land use, and human movement reinforce each other. Greenery should follow the rhythm of activity: dense tree planting near pedestrian flows, mixed layers of shrubs and grass around seating zones, and open lawns near recreational or cultural nodes. This integration improves comfort and ensures that ecological functions, such as shading and biodiversity, overlap with areas of high public exposure. Functional diversity, in turn, enhances the legibility and safety of these spaces by keeping them active throughout the day. Policies should promote mixed-use zoning that supports ecological maintenance, encouraging small commercial units, cafés, and community facilities along green corridors to sustain social engagement. These measures link the physical management of greenery to the everyday social metabolism of the city. Maintenance schedules should also adapt to patterns of use, directing more frequent care to streets with high activity levels, where the condition of vegetation directly affects perception and use. The goal is not to maximize either vegetation density or land-use intensity but to maintain their equilibrium as a self-reinforcing system that evolves with social needs.

This study deepens the theoretical understanding of urban streetscapes and offers a basis for practical intervention. By proposing the Universal Structure of Green Streets (USGGS), it shifts attention from the quantity of greenery to the internal stratification and composition of street vegetation, thereby orienting green infrastructure research toward a more structural and configuration-focused perspective. Through the joint analysis of USGGS, point-of-interest (POI)–based functional metrics and perception indices, the study conceptualizes streets as socio–ecological–perceptual coupled systems in which ecological form, functional organization and human experience are interconnected and potentially mutually reinforcing. The integration of deep learning and interpretable machine learning (semantic segmentation, visual perception modeling, and WS-LightGBM combined with SHAP) further illustrates how data-driven methods can support theoretical development by revealing nonlinear interaction pathways among greening, functionality and perception.Building on these insights, fine-grained spatial data provide an analytical foundation for area-specific planning strategies. POI-based functional metrics help to identify mismatches between ecological structure and social activity, thereby informing targeted interventions—for example, activating streets with complex vegetation but limited social use, and enhancing ecological quality on streets with high functional intensity but weak

greening structure. This diagnostic capacity can support more efficient resource allocation by aligning investments in street greening with anticipated perceptual benefits and social returns. At the same time, integrating ecological stewardship with community-based maintenance and management may reinforce continuity between physical form and local identity. At the policy level, the framework can inform coordination across transport, landscape and community planning departments, facilitating the development of flexible zoning tools, adaptive planting guidelines and participatory maintenance schemes. Together, these measures have the potential to gradually reshape streets into adaptive public landscapes that sustain environmental quality and human well-being over the long term.

Despite these contributions, several limitations should be acknowledged. First, the present analysis relies on a relatively generic representation of urban street greening structure, which constrains the ecological resolution at which vegetation–perception–function linkages can be interpreted. Beyond the greening metrics employed here, future work could enrich the ecological dimensions of the USGGS by incorporating finer-scale tree and canopy attributes such as tree morphology and vertical stratification, crown size and canopy cover, species or functional types, vegetation density, and leaf area index (LAI). These descriptors would support more precise estimation of shading, evapotranspiration, thermal comfort, and biodiversity-related functions along streetscapes, and thus offer more ecologically grounded insights into how vegetation structure may relate to both perceived and functional diversity. In addition, the modeling framework is calibrated within a specific urban context and relies on POI-based indicators of functional diversity, which may not fully capture informal or temporally dynamic activities. Future research could therefore examine the transferability of the framework across cities with different urban forms and data conditions, and explore how alternative functional proxies together with richer ecological metrics might jointly improve the diagnosis and design of green, socially active streets.

## 6. Conclusions

Urban streets constitute a fundamental component of metropolitan life, connecting ecosystems with human activities and shaping how people experience the city. Yet, existing research often examines ecological structure, social function and human perception in isolation: some studies emphasize green coverage while neglecting internal structure, whereas others focus on perceptual experiences without linking them to quantifiable spatial–ecological elements. This study addresses these gaps by constructing an integrated analytical framework that systematically examines how the greening structure, perceptual attributes and fine-grained urban functions of streets in six central districts of Tianjin jointly relate to street quality.

Methodologically, the framework integrates deep learning-based vegetation segmentation, visual perception modeling and refined functional data. Using DeepLabV3+, it extracts the Urban Street Greenery Generalized Structure (USGGS) from Baidu Street View imagery, capturing the internal layering and combinatorial forms of street vegetation. A vision transformer model trained on large-scale image comparison data is used to predict six perceptual indicators, while block-level functional diversity is represented with point-of-interest (POI) data. Regression analysis and interpretable machine learning models are then employed to examine associations among ecological, perceptual and functional variables, enabling the joint consideration of physical, social and perceptual characteristics within a unified analytical framework.

The results indicate that a street's ecological and perceived quality is closely related not only to vegetation structure and functional diversity, but also to their degree of coordination. Streets that combine multi-layered greenery with diverse, active urban functions tend

to exhibit higher perceived aesthetics, safety and vitality, whereas streets with single-layer vegetation or functionally monotonous environments generally perform less well. Functional patterns appear to mediate the perception of greenery by translating ecological attributes into socio-emotional experiences: the coexistence of vegetation and visible activities is associated with perceptions of safety and vitality, while isolated vegetation may diminish the perceived impact of greening. These findings suggest that greening and functionality should not be treated as parallel, independent objectives, but rather as interdependent elements within a socio–ecological–perceptual system.

The applicability of this methodology is not limited to Tianjin. By leveraging widely accessible data sources—street-level imagery, POI records and open geospatial information—and employing transferable deep learning and gradient-boosting models, the framework can, in principle, be adapted to other high-density Chinese cities (e.g., Beijing, Shanghai, Guangzhou) and international metropolitan areas with comparable datasets. Applying the USGGS concept and integrated modeling approach across diverse urban contexts would enable comparative assessments of the interplay among ecological structure, functional organization and human perception, providing empirical support for designing streets that balance environmental resilience with perceptual appeal.

Despite these advances, several limitations should be acknowledged, which also point to directions for future research. First, the analysis is confined to a single metropolitan area at one point in time, and thus does not capture seasonal or long-term dynamics in vegetation and perception. Second, the characterization of street greenery structure remains relatively general; incorporating more detailed ecological attributes—such as species composition, canopy morphology, vegetation density or leaf area index—could more fully reveal connections to microclimate regulation, biodiversity and other ecosystem functions. Third, the functional dimension is approximated solely through POI data, which may under-represent informal or time-dependent activities. Perception metrics also rely exclusively on visual information, overlooking auditory, air quality or thermal environmental factors. Future work could extend this framework across multiple cities and time periods, integrate richer ecological and functional proxy indicators (e.g., mobile phone data, sensor data or participatory data), and combine image-based perception scores with field surveys. Such efforts would help to assess the robustness of the methodology in diverse urban settings and deepen understanding of how ecological and social processes jointly shape human experiences on urban streets.

**Author Contributions:** Conceptualization, Y.-X.S. and Y.-Y.S.; methodology, Y.-X.S. and Y.-Y.S.; software, Y.-X.S.; validation, Y.-X.S., Y.-Y.S. and Y.-K.Y.; formal analysis, Y.-X.S. and Y.-Y.S.; investigation, Y.-X.S.; resources, Y.-Y.S.; data curation, Y.-K.Y. and S.-B.Z.; writing—original draft preparation, Y.-X.S.; writing—review and editing, Y.-Y.S., Y.-K.Y., S.-B.Z., Q.J. and Z.-T.Z.; visualization, Y.-X.S.; supervision, Y.-Y.S. and F.-L.T.; project administration, Y.-Y.S.; funding acquisition, Y.-Y.S. and Z.-T.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Major Project of Social Sciences of the Tianjin Municipal Education Commission, "Art and Design Empowering the Transformation of Vernacular Culture and High-Quality Rural Development" (Grant No. 2023JWZD34).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author. The data are not publicly available due to legal and ethical restrictions.

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## Abbreviations

The following abbreviations are used in this manuscript:

| Symbol | Definition |
|---|---|
| $H, W, C$ | Height, width, and channel number of an input image. |
| $P$ | Patch size (pixels) in Vision Transformer embedding. |
| $N$ | Number of pixels (for segmentation) or number of patches (for ViT). |
| $D$ | Dimension of embedded token features. |
| $L$ | Number of encoder layers in the Vision Transformer. |
| $Q, K, V$ | Query, key, and value matrices in multi-head self-attention. |
| $d_k$ | Dimension of each attention head. |
| $E, E_{\text{pos}}$ | Linear projection matrix and positional embedding in ViT. |
| $z_{\text{cls}}$ | Classification token used to aggregate global information. |
| $W_{\text{cls}}, b_{\text{cls}}$ | Weight and bias of the ViT classification head. |
| $\phi(\cdot)$ | Non-linear activation function in feed-forward networks. |
| $r$ | Dilation (atrous) rate in atrous convolution. |
| $K$ | Kernel size of a convolution filter. |
| $w$ | Convolution kernel weight tensor. |
| $x, y$ | Input feature map and output feature map (segmentation context). |
| $F_{\text{ASPP}}$ | Output feature representation from ASPP module. |
| $F$ | Final fused feature map after ASPP. |
| $\sigma$ | Activation function (e.g., ReLU or Sigmoid). |
| $N, C$ | Number of pixels and number of segmentation classes (DeepLabV3+). |
| $y_{i,c}, p_{i,c}$ | True label and predicted probability of pixel $i$ for class $c$. |
| $L$ | Loss function (cross-entropy or mean-square error depending on context). |
| $g_i, h_i$ | First- and second-order derivatives of the loss with respect to predictions in LightGBM. |
| $f_t$ | Regression tree added at iteration $t$ in boosting framework. |
| $T$ | Number of leaf nodes in a tree. |
| $w_j$ | Leaf weight (output value) for leaf $j$. |
| $G_j, H_j$ | Summed first- and second-order gradients for leaf $j$. |
| $G_L, G_R$ | First-order gradients for left and right child nodes in a split. |
| $H_L, H_R$ | Second-order gradients for left and right child nodes. |
| $\lambda$ | Regularization coefficient for leaf weights. |
| $\gamma$ | Complexity penalty for creating an additional leaf. |
| $\Omega(f_t)$ | Regularization term combining $\lambda$ and $\gamma$ penalties. |
| Gain | Standard split gain used in LightGBM. |
| $\text{Gain}_{\text{WS}}$ | Weighted-sensitivity split gain in WS-LightGBM. |
| $\alpha_i$ | Dynamic sensitivity coefficient of feature $i$ in WS-LightGBM. |
| $\beta$ | Hyperparameter controlling the degree of feature-sensitivity adjustment. |
| $\sigma_i$ | Standard deviation of feature $i$. |
| $\bar{\sigma}$ | Mean standard deviation across all features. |
| $d$ | Total number of input features. |
| $\theta$ | Vector of model hyperparameters. |
| $\Theta$ | Search space of hyperparameters in Bayesian optimization. |
| $J(\theta)$ | Validation loss corresponding to hyperparameter configuration $\theta$. |
| $J^\star$ | Current best (minimum) observed validation loss. |
| $\text{EI}(\theta)$ | Expected Improvement acquisition function in Bayesian optimization. |
| $\phi_i$ | SHAP value representing marginal contribution of feature $i$. |
| $F$ | Complete set of features used in the SHAP decomposition. |
| $S$ | Subset of features excluding feature $i$ in SHAP definition. |

| | |
|---|---|
| $f(S)$ | Model output evaluated with subset $S$ of features. |
| $M$ | Monotonic constraint matrix applied to features $(-1, 0, +1)$. |
| $\mathcal{F}$ | Function space of regression trees in LightGBM. |
| $\mathcal{L}^{(t)}$ | Objective function of boosting iteration $t$. |
| $\tilde{\mathcal{L}}^{(t)}$ | Simplified form of objective after substituting optimal leaf weights. |
| $\gamma, \lambda$ | Regularization terms controlling model complexity and leaf weight penalty. |

## Appendix A. Technical Derivations

*Appendix A.1. DeepLabV3+: Atrous Convolution, ASPP, and Loss Derivation*

Appendix A.1.1. Atrous Convolution

For a one-dimensional sequence, the atrous (dilated) convolution is defined as

$$y[i] \;=\; \sum_{k=0}^{K-1} w[k]\; x[i + r \cdot k], \tag{A1}$$

where $r$ is the dilation rate. In the two-dimensional case, the convolution at pixel $p$ is written as

$$y(p) \;=\; \sum_{u=-a}^{a} \sum_{v=-b}^{b} w(u, v)\; x\big(p + (r \cdot u,\, r \cdot v)\big). \tag{A2}$$

The receptive field grows linearly with $r$ while the number of parameters remains constant, allowing larger contextual perception without loss of resolution.

Appendix A.1.2. Atrous Spatial Pyramid Pooling (ASPP)

Given an input feature map $x \in \mathbb{R}^{H \times W \times C}$, the ASPP module aggregates multiple parallel atrous convolutions:

$$F_{\text{ASPP}} \;=\; \text{Concat}\Big(f_{r_1}(x),\, f_{r_2}(x),\, \ldots,\, f_{r_m}(x),\, f_{\text{img}}(x)\Big), \tag{A3}$$

where $f_{r_j}$ denotes branches with different dilation rates and $f_{\text{img}}$ represents global average pooling followed by a $1 \times 1$ convolution and upsampling. The fused output is

$$F \;=\; \sigma\big(W\, F_{\text{ASPP}} + b\big), \tag{A4}$$

where $\sigma$ is the activation function. ASPP effectively captures multi-scale context for objects of different sizes such as trees, shrubs, and grass.

Appendix A.1.3. Cross-Entropy Loss and Gradient

The pixel-wise multi-class cross-entropy loss is given by

$$L \;=\; -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{C} y_{i,c}\, \log p_{i,c}, \quad p_{i,c} \;=\; \frac{\exp z_{i,c}}{\sum_{c'} \exp z_{i,c'}}, \tag{A5}$$

where $z_{i,c}$ is the logit for pixel $i$ and class $c$. The gradient with respect to $z_{i,c}$ is

$$\frac{\partial L}{\partial z_{i,c}} \;=\; \frac{1}{N}\,(p_{i,c} - y_{i,c}), \tag{A6}$$

which is used for back-propagation. DeepLabV3+ employs atrous convolution and ASPP in the encoder and a decoder for upsampling and boundary refinement, enabling precise segmentation of hierarchical vegetation layers.

*Appendix A.2. Vision Transformer: Patch Embedding, MHSA, and Output Head*

Appendix A.2.1. Patch Embedding

An input image $x \in \mathbb{R}^{H\times W\times C}$ is divided into $N = (H/P)(W/P)$ patches of size $P \times P$. Each patch is flattened and linearly projected to a $D$-dimensional vector:

$$z_0 \;=\; [x_p^1 E;\, x_p^2 E;\, \ldots;\, x_p^N E] + E_{\text{pos}}, \tag{A7}$$

where $E \in \mathbb{R}^{(P^2C)\times D}$ is the learnable projection matrix and $E_{\text{pos}}$ is the positional embedding. A classification token $z_{\text{cls}}$ may be prepended to the sequence.

Appendix A.2.2. Multi-Head Self-Attention (MHSA)

For the input $Z \in \mathbb{R}^{N\times D}$, each attention head computes

$$Q_h = ZW_h^Q, \quad K_h = ZW_h^K, \quad V_h = ZW_h^V, \tag{A8}$$

and

$$\text{Attn}_h(Z) \;=\; \text{softmax}\left(\frac{Q_h K_h^\top}{\sqrt{d_k}}\right) V_h. \tag{A9}$$

The outputs of all $H$ heads are concatenated and projected:

$$\text{MHSA}(Z) \;=\; \big[\,\text{Attn}_1(Z);\ldots;\text{Attn}_H(Z)\,\big]W^O. \tag{A10}$$

Each encoder layer applies layer normalization and a feed-forward network (FFN):

$$Z' \;=\; \text{LN}\big(Z + \text{MHSA}(Z)\big), \qquad Z^+ \;=\; \text{LN}\big(Z' + \text{FFN}(Z')\big), \tag{A11}$$

where $\text{FFN}(u) = \phi(uW_1 + b_1)W_2 + b_2$. The $O(N^2D)$ complexity allows ViT to learn global spatial dependencies critical for perceptual understanding of urban scenes.

Appendix A.2.3. Output Head and Loss

The final representation of the classification token yields the prediction:

$$\hat{y} \;=\; \text{softmax}\big(W_{\text{cls}}\, z_{\text{cls}} + b_{\text{cls}}\big). \tag{A12}$$

For multi-dimensional perception regression (safety, wealth, boring, depressing, lively, beauty), a multi-task mean-squared error loss can be used:

$$L \;=\; \frac{1}{M}\sum_{m=1}^{M}\frac{1}{N}\sum_{i=1}^{N}\big(\hat{y}_i^{(m)} - y_i^{(m)}\big)^2, \tag{A13}$$

or individual regression heads can be applied for each perceptual dimension. The combination of positional embedding and multi-head attention enables the model to jointly capture openness, continuity, and vegetation distribution that underlie street-level perception.

*Appendix A.3. WS-LightGBM: Second-Order Objective, Weighted Gain, and Bayesian Optimization*

Appendix A.3.1. Second-Order Approximation and Objective

Given training data $\{(x_i, y_i)\}_{i=1}^n$, the additive model at iteration $t$ is

$$\hat{y}_i^{(t)} \;=\; \hat{y}_i^{(t-1)} + f_t(x_i), \qquad f_t \in \mathcal{F}, \tag{A14}$$

where $\mathcal{F}$ denotes the space of regression trees. Using a second-order Taylor expansion of the loss $l(y, \hat{y})$, the objective becomes

$$\mathcal{L}^{(t)} \approx \sum_{i=1}^{n} \left[ l\left(y_i, \hat{y}_i^{(t-1)}\right) + g_i f_t(x_i) + \tfrac{1}{2} h_i f_t(x_i)^2 \right] + \Omega(f_t), \tag{A15}$$

where

$$g_i = \frac{\partial l}{\partial \hat{y}_i}, \qquad h_i = \frac{\partial^2 l}{\partial \hat{y}_i^2}, \qquad \Omega(f_t) = \gamma T + \tfrac{1}{2}\lambda \sum_{j=1}^{T} w_j^2. \tag{A16}$$

For a given tree structure,

$$w_j^* = -\frac{G_j}{H_j + \lambda}, \qquad \tilde{\mathcal{L}}^{(t)} = -\frac{1}{2} \sum_{j=1}^{T} \frac{G_j^2}{H_j + \lambda} + \gamma T, \tag{A17}$$

where $G_j = \sum_{i \in I_j} g_i$ and $H_j = \sum_{i \in I_j} h_i$.

Appendix A.3.2. Standard and Weighted-Sensitivity Gain

For a node split into left and right children, the standard gain is

$$\text{Gain} = \frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} - \gamma. \tag{A18}$$

To adapt the model to heterogeneous features from greening, perception, and POI data, a feature-sensitive weight $\alpha_i$ is introduced:

$$\text{Gain}_{\text{WS}} = \alpha_i \left[ \frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma, \tag{A19}$$

where

$$\alpha_i = 1 + \beta \cdot \frac{\sigma_i}{\bar{\sigma}}, \qquad \bar{\sigma} = \frac{1}{d} \sum_{k=1}^{d} \sigma_k. \tag{A20}$$

Here $\sigma_i$ is the standard deviation of feature $i$, $d$ is the total number of features, and $\beta$ controls sensitivity. When $\beta = 0$, the formulation reduces to the standard gain.

Appendix A.3.3. Bayesian Hyperparameter Optimization

Let $\theta$ denote the set of hyperparameters. The validation loss $J(\theta)$ is modeled by a Gaussian process surrogate. The next evaluation point is chosen by maximizing the expected improvement:

$$\theta_{t+1} = \arg\max_{\theta \in \Theta} \text{EI}(\theta) = \mathbb{E}\big[\max(0,\, J^\star - J(\theta))\big], \tag{A21}$$

where $J^\star$ is the current best observed loss. This procedure iteratively updates $\theta$ until convergence, ensuring stable and efficient optimization.

Appendix A.3.4. SHAP Interpretation

Given the feature set $F$, the SHAP value of feature $i$ is

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} [f(S \cup \{i\}) - f(S)], \tag{A22}$$

which quantifies the marginal contribution of feature $i$ to the model output. For tree models, SHAP values can be computed efficiently using path-dependent algorithms, preserving additivity and consistency.

Appendix A.3.5. Regularization and Monotonic Constraints

When specific POI features are expected to have monotonic relationships with perception or greening indicators, monotone constraints $M \in \{-1, 0, 1\}^d$ can be applied to restrict splitting directions. These constraints limit feasible thresholds but do not alter the second-order framework or gain formulation. The resulting WS-LightGBM maintains interpretability while adapting to the nonlinear and heterogeneous structure of urban data.

## References


1. Haaland , C.; van den Bosch, C.K. Challenges and strategies for urban green-space planning in cities undergoing densification: A review. *Urban For. Urban Green.* **2015**, *14*, 760–771. [CrossRef]
2. Li, H.; Zhao, Y.; Wang, C.; Ürge-Vorsatz, D.; Carmeliet, J.; Bardhan, R. Cooling efficacy of trees across cities is determined by background climate, urban morphology, and tree trait. *Commun. Earth Environ.* **2024**, *5*, 754. [CrossRef]
3. Nowak, D.J.; Hirabayashi, S.; Bodine, A.; Greenfield, E. Tree and forest effects on air quality and human health in the United States. *Environ. Pollut.* **2014**, *193*, 119–129. [CrossRef]
4. Lu, Y.; Ferranti, E.J.S.; Chapman, L.; Pfrang, C. Assessing urban greenery by harvesting street view data: A review. *Urban For. Urban Green.* **2023**, *83*, 127917. [CrossRef]
5. Liang, D.; Huang, G. Influence of Urban Tree Traits on Their Ecosystem Services: A Literature Review. *Land* **2023**, *12*, 1699. [CrossRef]
6. Bowler, D.E.; Buyung-Ali, L.; Knight, T.M.; Pullin, A.S. Urban greening to cool towns and cities: A systematic review of the empirical evidence. *Landsc. Urban Plan.* **2010**, *97*, 147–155. [CrossRef]
7. Nowak, D.J.; Crane, D.E.; Stevens, J.C. Air pollution removal by urban trees and shrubs in the United States. *Urban For. Urban Green.* **2006**, *4*, 115–123. [CrossRef]
8. Salmond, J.A.; Tadaki, M.; Vardoulakis, S.; Arbuthnott, K.; Coutts, A.; Demuzere, M.; Dirks, K.N.; Heaviside, C.; Lim, S.; Macintyre, H.; et al. Health and climate related ecosystem services provided by street trees in the urban environment. *Environ. Health* **2016**, *15*, S36. [CrossRef]
9. Jin, B.; Geng, J.; Ke, S.; Pan, H. Analysis of spatial variation of street landscape greening and influencing factors: An example from Fuzhou city, China. *Sci. Rep.* **2023**, *13*, 21767. [CrossRef]
10. Sharifi, F.; Nygaard, A.; Stone, W.M.; Levin, I. Accessing green space in Melbourne: Measuring inequity and household mobility. *Landsc. Urban Plan.* **2021**, *207*, 104004. [CrossRef]
11. Seiferling, I.; Naik, N.; Ratti, C.; Proulx, R. Green streets—Quantifying and mapping urban trees with street-level imagery and computer vision. *Landsc. Urban Plan.* **2017**, *165*, 93–101. [CrossRef]
12. Long, Y.; Liu, L. How green are the streets? An analysis for central areas of Chinese cities using Tencent Street View. *PLoS ONE* **2017**, *12*, e0171110. [CrossRef]
13. Dong, R.; Zhang, Y.; Zhao, J. How Green Are the Streets Within the Sixth Ring Road of Beijing? An Analysis Based on Tencent Street View Pictures and the Green View Index. *Int. J. Environ. Res. Public Health* **2018**, *15*, 1367. [CrossRef]
14. Li, X.; Ratti, C. Mapping the spatial distribution of shade provision of street trees in Boston using Google Street View panoramas. *Urban For. Urban Green.* **2018**, *31*, 109–119. [CrossRef]
15. Li, X.; Ratti, C.; Seiferling, I. Quantifying the shade provision of street trees in urban landscape: A case study in Boston, USA, using Google Street View. *Landsc. Urban Plan.* **2018**, *169*, 81–91. [CrossRef]
16. Bardhan, M.; Li, F.; Browning, M.H.E.M.; Dong, J.; Zhang, K.; Yuan, S.; İnan, H.E.; McAnirlin, O.; Dagan, D.T.; Maynard, A. ; et al. From space to street: A systematic review of the associations between visible greenery and bluespace in street view imagery and mental health. *Environ. Res.* **2024**, *263*, 120213. [CrossRef] [PubMed]
17. Xia, Y.; Yabuki, N.; Fukuda, T. Development of a system for assessing the quality of urban street-level greenery using street view images and deep learning. *Urban For. Urban Green.* **2021**, *59*, 126995. [CrossRef]
18. Zhang, L.; Wang, L.; Wu, J.; Li, P.; Dong, J.; Wang, T. Decoding urban green spaces: Deep learning and google street view measure greening structures. *Urban For. Urban Green.* **2023**, *87*, 128028. [CrossRef]
19. 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]
20. Zhang, F.; Zhou, B.; Liu, L.; Liu, Y.; Fung, H.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]

21. Wei, J.; Yue, W.; Li, M.; Gao, J. Mapping human perception of urban landscape from street-view images: A deep-learning approach. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *112*, 102886. [CrossRef]
22. Zhang, J.; Xue, D.; Wang, C.; Fang, D.; Cao, L.; Gong, C. Genetic engineering for biohydrogen production from microalgae. *iScience* **2023**, *26*, 107255. [CrossRef]
23. O'Regan, A.C.; Byrne, R.; Hellebust, S.; Nyhan, M.M. Associations between Google Street View-derived urban greenspace metrics and air pollution measured using a distributed sensor network. *Sustain. Cities Soc.* **2022**, *87*, 104221. [CrossRef]
24. Schaefer, A.; Scolaro, T.P.; Ghisi, E. Cluster analysis applied to obtaining reference models for building thermal performance studies . *J. Build. Eng.* **2024**, *89* , 109273. [CrossRef]
25. Meng, Y.; Shi, J.; Lyu, M.; Sun, D.; Fukuda, H. Research into the Influence Mechanisms of Visual-Comfort and Landscape Indicators of Urban Green Spaces. *Land* **2024**, *13*, 1688. [CrossRef]
26. Zhang, Y.; Kwan, M.-P.; Yang, J. A user-friendly assessment of six commonly used urban growth models *Comput. Environ. Urban Syst.* **2023**, *104*, 102004. [CrossRef]
27. Bai, Y.; Wang, R.; Yang, L.; Ling, Y.; Cao, M. The Impacts of Visible Green Spaces on the Mental well-being of University Students *Appl. Spat. Anal. Policy* **2024**, *17*, 1105–1127. [CrossRef]
28. Wang, R.; Browning, M.H.E.M.; Qin, X.; He, J.; Wu, W.; Yao, Y.; Liu, Y. Visible green space predicts emotion: Evidence from social media and street view data. *Appl. Geogr.* **2022**, *148*, 102803. [CrossRef]
29. Biljecki, F.; Ito, K. Street view imagery in urban analytics and GIS: A review. *Landsc. Urban Plan.* **2021**, *215*, 104217. [CrossRef]
30. Seresinhe, C.I.; Preis, T.; Moat, H.S. Using deep learning to quantify the beauty of outdoor places. *R. Soc. Open Sci.* **2017**, *4*, 170170. [CrossRef] [PubMed]
31. Fan, Z.; Zhang, F.; Loo, B.P.Y.; Ratti, C. Urban visual intelligence: Uncovering hidden city profiles with street view images. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, e2220417120. [CrossRef]
32. Hu, Y.; Han, Y. Identification of Urban Functional Areas Based on POI Data: A Case Study of the Guangzhou Economic and Technological Development Zone. *Sustainability* **2019**, *11*, 1385. [CrossRef]
33. Czekajlo, A.; Coops, N.C.; Wulder, M.A.; Hermosilla, T.; White, J.C.; Bosch, M.v.D. Mapping dynamic peri-urban land use transitions across Canada using Landsat time series: Spatial and temporal trends and associations with socio-demographic factors *Comput. Environ. Urban Syst.* **2021**, *88*, 101653. [CrossRef]
34. Wang, Y.; Gu, Y.; Dou, M.; Qiao, M. Using Spatial Semantics and Interactions to Identify Urban Functional Regions. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 130. [CrossRef]
35. Qin, Q.; Xu, S.; Du, M.; Li, S. Identifying urban functional zones by capturing multi-spatial distribution patterns of points of interest. *Int. J. Digit. Earth* **2022**, *15*, 2468–2494. [CrossRef]
36. Du, S.; Zhang, X.; Lei, Y.; Huang, X.; Tu, W.; Liu, B.; Meng, Q.; Du, S. Mapping urban functional zones with remote sensing and geospatial big data: A systematic review. *GIScience Remote Sens.* **2024**, *61*, e2404900. [CrossRef]
37. Larkin, A.; Gu, X.; Chen, L.; Hystad, P. Predicting perceptions of the built environment using GIS, satellite and street view image approaches. *Landsc. Urban Plan.* **2021**, *216*, 104257. [CrossRef] [PubMed]
38. Nesbitt, L.; Meitner, M.J.; Girling, C.; Sheppard, S.R.; Lu, Y. Who has access to urban vegetation? A spatial analysis of distributional green equity in 10 US cities. *Landsc. Urban Plan.* **2019**, *181*, 51–79. [CrossRef]
39. Gerrish, E.; Watkins, S.L. The relationship between urban forests and income: A meta-analysis. *Landsc. Urban Plan.* **2018**, *170*, 293–308. [CrossRef] [PubMed]
40. Wu, S.; Chen, B.; Webster, C.; Xu, B.; Gong, P. Improved human greenspace exposure equality during 21st century urbanization. *Nat. Commun.* **2023**, *14*, 6460. [CrossRef] [PubMed]
41. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [CrossRef] [PubMed]
42. Ouyang, J.; Fan, H.; Wang, L.; Zhu, D.; Yang, M. Revealing urban vibrancy stability based on human activity time-series. *Sustain. Cities Soc.* **2022**, *85*, 104053. [CrossRef]
43. Caicoya, A.T.; Poschenrieder, W.; Blattert, C.; Eyvindson, K.; Hartikainen, M.; Burgas, D.; Mönkkönen, M.; Uhl, E.; Vergarechea, M.; Pretzsch, H. Sectoral policies as drivers of forest management and ecosystems services: A case study in Bavaria, Germany *Land Use Policy* **2023**, *130*, 106673. [CrossRef]
44. Yang, F.; Yousefpour, R.; Zhang, Y.; Wang, H. The assessment of cooling capacity of blue-green spaces in rapidly developing cities: A case study of Tianjin's central urban area. *Sustain. Cities Soc.* **2023**, *99*, 104918. [CrossRef]
45. Sha, S.; Cheng, Q. Built or Social environment? Effects of perceptions of neighborhood green spaces on resilience of residents to heat waves. *Urban For. Urban Green.* **2024**, *94*, 128267. [CrossRef]
46. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 833–851. [CrossRef]

47. Khan, S.; Naseer, M.; Hayat, M.; Zamir, S.W.; Khan, F.S.; Shah, M. Transformers in Vision: A Survey. *ACM Comput. Surv.* **2022**, *54*, 1–41. [CrossRef]
48. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.M.; Liu, T.Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In Proceedings of the Advances in Neural Information Processing Systems 30 (NeurIPS 2017), Long Beach, CA, USA, 4–9 December 2017; pp. 3146–3154.
49. Zhou, S.; Ji, Q.; Zhang, L.; Wu, J.; Li, P.; Zhang, Y. Exploration of Differences in Housing Price Determinants Based on Street View Imagery and the Geographical-XGBoost Model: Improving Quality of Life for Residents and Through-Travelers. *ISPRS Int. J. Geo-Inf.* **2025**, *14*, 391. [CrossRef]
50. Tang, F.; Zeng, P.; Wang, L.; Zhang, L.; Xu, W. Urban Perception Evaluation and Street Refinement Governance Supported by Street View Visual Elements Analysis. *Remote Sens.* **2024**, *16*, 3661. [CrossRef]
51. Roman, L.A.; Battles, J.J.; McBride, J.R. Determinants of establishment survival for residential trees in Sacramento County, CA. *Landsc. Urban Plan.* **2014**, *129*, 22–31. [CrossRef]
52. Peters, K.; Elands, B.; Buijs, A. Social interactions in urban parks: Stimulating social cohesion? *Urban For. Urban Green.* **2010**, *9*, 93–100. [CrossRef]
53. McClintock, N.; Cooper, J.; Khandeshi, S. Assessing the potential contribution of vacant land to urban vegetable production and consumption in Oakland, California. *Landsc. Urban Plan.* **2013**, *111*, 46–58. [CrossRef]
54. Andersson, E.; Barthel, S.; Borgström, S.; Colding, J.; Elmqvist, T.; Folke, C.; Åsa Gren. Reconnecting cities to the biosphere: Stewardship of green infrastructure and urban ecosystem services. *Ambio* **2014**, *43*, 445–453. [CrossRef]
55. Rigolon, A. A complex landscape of inequity in access to urban parks: A literature review. *Landsc. Urban Plan.* **2016**, *153*, 160–169. [CrossRef]
56. Nordh, H.; Alalouch, C.; Hartig, T. Assessing restorative components of small urban parks using conjoint methodology. *Urban For. Urban Green.* **2011**, *10*, 95–103. [CrossRef]
57. Zhao, T.; Liang, X.; Tu, W.; Huang, Z.; Biljecki, F. Sensing urban soundscapes from street view imagery. *Comput. Environ. Urban Syst.* **2023**, *99*, 101915. [CrossRef]
58. Aslanoğlu, R.; Kazak, J.K.; Szewrański, S.; Świąder, M.; Arciniegas, G.; Chrobak, G.; Jakóbiak, A.; Turhan, E. Ten questions concerning the role of urban greenery in shaping the future of urban areas. *Build. Environ.* **2025**, *267*, 112154. [CrossRef]
59. Xie, Y.; Zhu, M.; Wang, F.; Zhao, J.; Sun, W.; Qin, Y.; Zhang, H. Unveiling changes in urban visual perception inequality in high-density cities: Insights from housing types and travel distances using time-series street view imagery. *Appl. Geogr.* **2025**, *185*, 103776. [CrossRef]
60. Kim, Y.; Brown, R.D. Climate-sensitive street design: Evaluating summer pedestrian activity and behavioral thermal adaptation on the High Line, NYC. *Build. Environ.* **2025**, *281*, 113203. [CrossRef]
61. Liu, Q.; Zou, G.; Luo, Y. How the neighborhood amenity mix shapes urban vitality: An exploratory analysis from a rhythm perspective. *Appl. Geogr.* **2025**, *185*, 103807. [CrossRef]
62. Wang, Z.; Ito, K.; Biljecki, F. Assessing the equity and evolution of urban visual perceptual quality with time series street view imagery. *Cities* **2024**, *145*, 104704. [CrossRef]
63. Nathvani, R.; Cavanaugh, A.; Suel, E.; Bixby, H.; Clark, S.N.; Metzler, A.B.; Nimo, J.; Moses, J.B.; Baah, S.; Arku, R.E.; et al. Measurement of urban vitality with time-lapsed street-view images and object-detection for scalable assessment of pedestrian-sidewalk dynamics. *ISPRS J. Photogramm. Remote Sens.* **2025**, *221*, 251–264. [CrossRef]
64. Navarrete-Hernandez, P.; Afarin, K. The impact of nature-based solutions on perceptions of safety in public space. *J. Environ. Psychol.* **2023**, *91*, 102132. [CrossRef]
65. Dong, X.; Yang, R.; Ye, Y.; Yi, S.; Haase, D.; Lausch, A. Planning for green infrastructure by integrating multi-driver: Ranking priority based on accessibility–equity. *Sustain. Cities Soc.* **2024**, *114*, 105767. [CrossRef]
66. Adams, C.; Frantzeskaki, N.; Moglia, M. Mainstreaming nature-based solutions in cities: A systematic literature review and a proposal for facilitating urban transitions. *Land Use Policy* **2023**, *130*, 106661. [CrossRef]
67. Ito, K.; Kang, Y.; Zhang, Y.; Zhang, F.; Biljecki, F. Understanding urban perception with visual data: A systematic review. *Cities* **2024**, *152*, 105169. [CrossRef]
68. Pröbstl-Haider, U.; Wanner, A.; Feilhammer, M.; Mostegl, N.; Dabrowska, K. The right fit: Acceptance of nature-based solutions across European cities. *Landsc. Urban Plan.* **2024**, *252*, 105189. [CrossRef]
69. Xu, L.; Tong, S.; He, W.; Zhu, W.; Mei, S.; Cao, K.; Yuan, C. Better understanding on impact of microclimate information on building energy modelling performance for urban resilience. *Sustain. Cities Soc.* **2022**, *80*, 103775. [CrossRef]
70. Biljecki, F.; Zhao, T.; Liang, X.; Hou, Y. Sensitivity of measuring the urban form and greenery using street-level imagery: A comparative study of approaches and visual perspectives. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *122*, 103385. [CrossRef]
71. Sang, Y.; Hu, Y.; Qin, X.; Yan, H.; Wu, R.; Qian, F.; Nan, X.; Shao, F.; Bao, Z. Impacts of street tree canopy coverage on pedestrians' dynamic thermal perception and walking willingness. *Sustain. Cities Soc.* **2025**, *121*, 106196. [CrossRef]

72. Fan, Z.; Feng, C.-C.; Biljecki, F. Coverage and bias of street view imagery in mapping the urban environment. *Comput. Environ. Urban Syst.* **2025**, *117*, 102253. [CrossRef]
73. Qiao, W.; Zheng, H. Predicting urban vitality and pedestrian road safety in urban areas based on machine learning. *Cities* **2025**, *166*, 106193. [CrossRef]

**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.