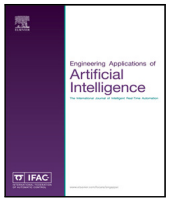
Engineering Applications of Artificial Intelligence 136 (2024) 108905

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai



Research paper

An estimation method for multidimensional urban street walkability based on panoramic semantic segmentation and domain adaptation

Jiaxuan Li [a], Xuan Zhang [a,b,*], Linyu Li [c,d], Xu Wang [e], Jing Cheng [f], Chen Gao [g], Jun Ling [a]

[a] School of Software, Yunnan University, Yunnan 650091, China
[b] Yunnan Key Laboratory of Software Engineering, Yunnan 650091, China
[c] School of Computer Science, Peking University, Beijing 100871, China
[d] Key Lab. of High-Confidence Software Technologies (PKU), Ministry of Education, Beijing 100871, China
[e] School of Economics, Yunnan University, Yunnan 650091, China
[f] Smart City Business Group, Yunnan Nantian Electronics Information Corp., Ltd., Yunnan 650041, China
[g] School of Information Science and Engineering, Yunnan University, Yunnan 650091, China

ARTICLE INFO

Keywords:
Panoramic image semantic segmentation
Domain adaptation
Human–machine adversarial
Unsupervised learning
Urban walkability analysis

ABSTRACT

Urban walkability is a critical aspect of urban planning. Since the traditional measurement methods constrained by cost, time, and scalability researchers have turned to computer-assisted audits based on Panoramic Street View Images (PSVIs). However overlook image distortion and data annotation issues, impacting predictive accuracy. Current evaluations often lack a holistic approach, making comparison challenging. In response, a multidimensional evaluation approach is proposed through three indices: street red quality, physical walkability, and perceived walkability. To enhance accuracy, a Transformer-based Doubly Deformable Panoramic semantic segmentation Network (TDDPassNet) is introduced to calculate key metrics informing ecological quality and spatial layout evaluations. An unsupervised domain adaptation method is proposed for insufficient labeled data. Furthermore, a Geographic Information System (GIS) analysis was conducted to assess the physical walkability index. Human–machine adversarial technology and a random forest model evaluate the perceived walkability index. A comprehensive evaluation framework is presented using the Analytic Hierarchy Process to assign weights to assessment indices across the three dimensions. A case study was conducted in Lijiang City, China, to demonstrate the practical application of the methodology. Extensive experiments are conducted, TDDPassNet exhibits an average increase of 5.3% in mIoU across diverse datasets compared to the prevailing models. This study evaluated 47,758 sampling sites, providing insights into urban planning and development in similar contexts

1. Introduction

Accelerating urbanization has led to increased car ownership, resulting in smaller walkable city spaces and a worse ecological environment. This, in turn, restricts urban walkability, generating traffic congestion and aggravating environmental pollution. As a result, cities fall into a vicious circle. Scientific and reasonable adjustment of urban planning to enhance walkability has become crucial in urban planning and construction (Scorza et al., 2021). Walking is an eco-friendly mode of transportation that reduces traffic congestion and environmental pollution while promoting physical and mental health. Walkable urban streets and districts can provide a space for residents to socialize, exercise, and engage in daily activities. This can contribute to the economic and social development of the city (Kim and Woo, 2022). Therefore, urban planning and design should measure walkability to develop appropriate urban design and retrofit recommendations. This is important for promoting urban walkability and urban habitability (Gong et al., 2018).

Early researchers focused on describing residents' community life through their social relationships. Studies on street walkability were conducted using face-to-face interviews, questionnaires, and ratings from expert panels. For instance, Ewing et al. (2006) and Ewing and Handy (2009) measured the street environment and tested for significant associations with walking habitability. Peiravian et al. (2014) developed and applied indices such as land use diversity and street green visibility to calculate the Pedestrian Environment Index (PEI). The traditional survey-based approaches can capture residents' opinions regarding street quality, but only for limited areas of interest

* Corresponding author.
E-mail address: zhxuan@ynu.edu.cn (X. Zhang).

https://doi.org/10.1016/j.engappai.2024.108905
Received 8 April 2024; Received in revised form 9 June 2024; Accepted 27 June 2024
Available online 10 July 2024

due to their time-consuming nature, small sample sizes, high cost, and inefficiency. In recent years, computer vision techniques have enabled large-scale quantitative measurement of Street View Images (SVIs). This trend has emerged due to the continuous rise of digital mapping services and the constant updating of crowd-sourced street view data (Arellana et al., 2020; Wang et al., 2021; Ogawa et al., 2023).

The rapid development of computer technology has facilitated the integration of deep learning into urban and rural planning and landscape architecture analysis. An increasing number of researchers are utilizing deep learning-based methods for urban planning research. This includes using techniques such as semantic segmentation and image classification to analyze street indices and using street images to understand the relationship between various types of street attributes (Mohanty et al., 2020; Sun et al., 2021b). In street view big data, panoramic street view images (PSVIs) differ from remote sensing images. They provide a three-dimensional and rich view of 360° urban scenes from the pedestrians' point of view to conduct analyses of visual elements and urban planning (Kim et al., 2022). As a result, a large number of urban planning studies are currently advancing on this basis. For different analysis indices based on PSVIs data, which are processed using computer vision, e.g., Tang et al. (2022) and Zhao et al. (2023). However, panoramic images often have significant image distortions and object deformations due to the inherent equirectangular projection (Sun et al., 2021a). As a result, traditional convolutional neural networks and learning methods may not be the best solution for processing panoramic images. In addition to the deformation of panoramic images, the lack of labeled data is another significant challenge that hinders the progress of visual processing of panoramic images (Yang et al., 2021b). Most researchers in this field currently use traditional pinhole image-oriented models, such as SegNet (Badrinarayanan et al., 2017), UNet (Ronneberger et al., 2015), PspNet (Zhao et al., 2017), VPLR (Zhu et al., 2019) and DeepLabV3+ (Chen et al., 2018), which are challenging to handle severe image distortions and are not trained with labeled panoramic image datasets. The models' weak feature extraction capability and the lack of relevant datasets for feature training may reduce their ability to predict the semantic segmentation of images. This, in turn, affects the reliability of subsequent urban studies' results.

A quantitative analysis of only some features of streetscape images is insufficient to fully represent urban walkability. A comprehensive assessment from multiple dimensions is necessary. Zhou et al. (2019) did not consider people's walking needs at the physical and perceptual levels. They measured the quality of the street visual environment solely in terms of greenery, visual congestion, outdoor fencing, and visual pavement width. Horak et al. (2022) assessed the walkability of urban physical spaces using an approach based on accessibility by calculating neighborhood environment indices, excluding subjective evaluations. Meanwhile, Koohsari et al. (2021) investigated the link between walking perception scores and built environment attributes but did not conduct actual neighborhood physical walking index measurements. The methods mentioned above have one-sided metrics and a single assessment model. However, methods for measuring urban walkability require more comprehensive assessment criteria and systematic assessment processes.

To solve the above problems, this paper presents a framework for assessing urban walkability from a comprehensive urban planning perspective. The framework has three dimensions: street ecological quality, physical walkability, and perceived walkability. The ecological quality of the street, such as green coverage and sky view factor, is assessed to reflect the ecological condition of the city street. To address the pivotal distortion issue in panoramic image data, a Transformer-based (Vaswani et al., 2017) Doubly Deformable Panoramic semantic segmentation Network (TDDPassNet) is proposed. To address the issue of limited sample data, the panoramic data is annotated using an unsupervised multi-stage prototypical domain adaptation method. Subsequently, variables such as walking paths, point of interest (POI) facilities, and essential urban land use classification (EULUC) data are analyzed and computed using walkability analysis to reflect the actual physical walkability of city streets. Then, Human–machine adversarial technology is employed to collect perceptual data on the urban walking environment. The four-dimensional perceptual scores of Safety, Convenience, Comfort, and Attractiveness of different streets in the city are analyzed to assess the perceived walkability of the city and reflect pedestrians' subjective feelings towards the urban walking environment. Finally, the city's Comprehensive Walkability Index (CWI) is calculated. This paper's primary contributions can be summarized as follows:

1. From urban planning perspective, a three-dimensional urban walkability framework is proposed to measure the Comprehensive Walkability Index (CWI) for cities. The framework integrates multiple data sources, including PSVIs, POI, and EULUC, and uses techniques such as semantic segmentation, GIS, and human–computer adversarial for analysis. By combining the street ecological quality index, physical walkability index, and perceived walkability index into a single metric, the CWI provides a more comprehensive evaluation of walkability, addressing the limitation of partial evaluation from a single perspective prevalent in previous studies.
2. In the field of urban walkability research, the consideration of image distortion and object deformation in panoramic image data is addressed for the first time. A transformer-based Dual Deformable Panoramic semantic segmentation Network (TDDPassNet) is proposed. TDDPassNet incorporates a broad perspective, handles panoramic-specific semantic distributions in its design, and perceives image distortion through deformable MLP modules, multi-scale strategies, and dual attention mechanisms.
3. A novel unsupervised domain-adaptive method for self-annotation of panoramic images is developed to address the paucity of labeled data in the urban planning domain. The method is applied to the construction and opening of a total of 47,758 panoramic image datasets.[1]
4. The spatial distribution analysis of different indicators is performed, and the walkability of the ancient city area is visually analyzed through spatial distribution maps. A series of comparison and ablation experiments for the models and methods, as well as a more comprehensive qualitative analysis, are carried out. The experimental results demonstrate that TDDPassNet has the best performance compared to all other models, with an average increase of 5.3% in mIoU across different datasets.

The abbreviations of the most frequently utilized definitions in the article are presented in Table 1, accompanied by their complete nomenclature. The remainder of the study is organized as follows: Section 2 discusses related work in the field of urban walkability. The detailed methodology is presented in Section 3. Section 4 describes the experimental setup and analyses the results. A discussion of the study's outlook is presented in Section 5. Finally, Section 6 summarizes the work presented in the article.

2. Related work

2.1. Definition and approach to urban walkability

Walkability refers to the extent to which a street's environment is pedestrian-friendly (Guzman et al., 2022). It is a complex concept influenced by multiple variables and is challenging to quantify simply and specifically. Although there is no consensus on the measure of walkability, several studies have attempted to assess walkability from different aspects quantitatively. Earlier studies usually judged the walkability of urban streets from a single perspective only. For example,

[1] LjPass Dataset:https://github.com/Qingkongmu/LjPass.

Table 1
List of Abbreviations.

Abbreviation	Full name
PSVIs	Panoramic Street View Images
TDDPassNet	Transformer-based Doubly Deformable Panoramic Semantic Segmentation Network
PoI	Point of Interest
EUCUL	Urban Essential Land Use Categories
GIS	Geographic Information System
CWI	Comprehensive Walkability Index
mIoU	Mean Intersection over Union
RF	Random Forest

Cerin et al. (2007) focused on factors such as the reasonable time and cost required to reach various destinations, primarily considering the physical accessibility of walkable spaces while overlooking the ecological and perceptual dimensions. Ewing and Handy (2009) sought to gauge the subjective quality of urban street environments through ratings provided by a panel of experts. However, this relies too heavily on individual perceptual variables, and there is a one-sided evaluation of the walking environment.

Researchers have started exploring multiple dimensions to measure walkability to avoid the limitations of studies from a single perspective. Gebel et al. (2009) used a community questionnaire to consider elements of pedestrians' perceptions of walkable streets, considering various variables in the walking environment. However, their small and costly study area still has significant limitations. Another approach involves using user voting on browser platforms to select visually appealing, environmentally friendly, and physically and mentally pleasing streetscape data (Quercia et al., 2014). While this method has advantages regarding the ecological environment and sensory experience, it requires significant manual labeling, increasing time and labor costs. Tsiompras and Photis (2017) assessed the built environment features that influenced people's travel behavior and geo-visualization through a comprehensive walkability index approach based on weighted GIS. Although this approach has better human and time costs, the focus is limited to the environmental quality of the street and the actual walking index. It does not consider the pedestrian's needs for the walking environment at the visual perception level. The application of artificial intelligence to urban research has become increasingly common due to the recent emergence of deep learning techniques and the constant updating of crowdsourced map data.

In the domain of deep learning techniques for auditing urban walkability, Ki and Lee (2021) explored the correlation between urban greenery and pedestrian activities, while Hua et al. (2022) established a connection between street-level greenery in densely populated urban areas and the surrounding urban morphological conditions. In addition to these methodologies, Li et al. (2022b) integrated components of an image multi-classification module and virtual reality (VR) into the semantic segmentation model based on SVIs, facilitating evaluations closely mirroring real-world perceptions. Meanwhile, Li et al. (2023a) attempted to construct a comprehensive research methodology for assessing urban street walkability by amalgamating GIS, environmental sensors, and image semantic segmentation, along with incorporating various walking-related variables. He and He (2023) recognized the challenge of accurately gauging safety perceptions in urban walkability. They employed natural language processing technology to conduct public opinion analyses of street safety incidents, thereby refining the assessment model at the perceptual level. Based on the findings of these studies, the Comprehensive Walkability Index (CWI) is proposed as a tool for evaluating the walkability of urban streets from various perspectives in urban planning. The CWI incorporates assessments of the ecological quality of streets, as well as observations of pedestrian behavior and perceptions of willingness to walk.

2.2. Panoramic Street Images and image semantic segmentation

The global availability of Panoramic Street View Images (PSVI) has established it as a primary data source for urban studies, particularly in relation to the impact of the street-level built environment on human walkability (Biljecki and Ito, 2021). Major mapping services, including Google Maps, Tencent Maps, and Baidu Maps, provide researchers with high-resolution and realistic street-level panoramic image data through their APIs. PSVI offers significant advantages in observing and characterizing the built environment virtually. Compared to field surveys, it provides a cost-effective and time-efficient way of conducting large-scale assessments (Farahani et al., 2023). Additionally, PSVI-based audits demonstrate a high degree of consistency with physical audits, particularly in the areas of pedestrian infrastructure, traffic safety, and streetscape aesthetics. The audits exhibit over 80% consistency (Jamei et al., 2021). Research indicates that virtual audits conducted through PSVI can significantly reduce labor and time costs compared to field surveys. According to Bartzokas-Tsiompras et al. (2023), such audits can save up to 97% of labor costs and 90% of time costs.

Image semantic segmentation techniques are most commonly used by related researchers in PSVI-based quantitative analysis of cities (Suel et al., 2021). The process aims to classify image content semantically. Scene parsing based on semantic segmentation is a core topic in computer vision, with the goal of assigning category labels to each pixel in an image. This can be seen as an extension of target detection and a process of categorizing image pixels. The aim is to determine the category of a pixel from a set of discrete categories (Long et al., 2015).

Urban research scholars typically adopt existing semantic segmentation models for PSVI-based studies related to urban walkability. This approach is understandable as developing a more robust algorithm is technically challenging. However, these studies have paid little attention to the limitations of applying semantic segmentation. Recent studies on walkability that use SVI-based methods are summarized in Table 2. The year of the article, study area, information about the semantic segmentation model, the training set, the types of predicted image data used, and the accuracy of the model are shown in Table 2. Accurate extraction of street elements is a valid prerequisite for measuring walkability in walkability studies. However, the adoption of existing semantic segmentation models vastly overestimates the performance of the model. In particular, some of the models' performance has not been validated against a test set. Therefore, models cannot provide accurate results, and models lacking accuracy validation have limited contribution to subsequent studies (He and He, 2023). Additionally, these studies did not compare the performance of models that effectively identify target street features before model selection. It is necessary to compare models because they may have varying accuracies for specific street features, and even the same model may perform differently with different training datasets.

In particular, when selecting PSVI data for research purposes, the models chosen by the researchers were originally proposed for planar pinhole image data. As a result, these models may not be suitable for dealing with panoramic image data with obvious image distortion and object deformation problems (Zhao et al., 2023; Yang et al., 2021b). Furthermore, these models are not trained using panoramic images with labels as a training set, greatly reducing their performance in PSVI semantic segmentation prediction. This reduction in performance will seriously impact the accuracy of subsequent studies (Zhang et al., 2022). Therefore, developing a common application paradigm in PSVI-based research related to walkability is crucial, and it should include model testing, comparison, and selection. Future studies following this paradigm will have greater contribution and reference value.

Table 2
Summary of deep learning methods in recent SVI-based research related to walkability.

Article (Year)	Study area	Model used for semantic segmentation	Training set	Split image type	Accuracy rate
Tang and Long (2019) 2019	Binjiang, China	SegNet	ADE20K	Pinhole Image	82.54%
Nagata et al. (2020) 2020	Tokyo, Japan	DeepLab v3+	Cityscapes	Pinhole Image	82%
Suel et al. (2021) 2021	London, UK	U-Net	Not mentioned	Panoramic Image	Not mentioned
Jiang et al. (2021) 2021	Hong Kong, China	PSPNet	Cityscapes	Pinhole Image	Not mentioned
Ki and Lee (2021) 2021	Seoul, South Korea	FCN-8 s	Cityscapes	Panoramic Image	84.56%
Li et al. (2022b) 2022	Osaka, Japan	DeepLab v3+	Cityscapes	Panoramic Image	Not mentioned
Kim and Woo (2022) 2022	Seoul, South Korea	HRNetV2-W48	ADE20K	Panoramic Image	Not mentioned
Li et al. (2023a) 2023	Osaka, Japan	DeepLab v3+	Cityscapes	Panoramic Image	81.20%
Kang et al. (2023) 2023	Jeonju, South Korea	DeepLab v3	ADE20K	Panoramic Image	87.50%
He and He (2023) 2023	Shenzhen, China	VPLR, DeepLabV3, ResNet	Cityscapes, ADE20K	Panoramic Image	92%

2.3. Unsupervised domain adaptive approach

Unsupervised learning (UL) has proven valuable in comprehending the intricacy of cities. Unlike supervised learning methods, UL uncovers patterns from inherent data structures without the need for manual labeling, which is believed to be crucial in producing truly AI decisions (Wang and Biljecki, 2022). Supervised learning has proven to be useful for various applications and datasets. However, it may not be suitable for addressing all research questions due to challenges such as obtaining training data. Real-world urban data often lacks labeling information, as noted by Yang et al. (2021a).

Cities are complex creations with patterns believed to be hidden in their physical form and daily operations (Anthony, 2023). With the increasing amount of urban data and the application of machine learning techniques, it has become possible to identify patterns from large-scale data automatically. This has gained momentum in providing researchers with assistance in unraveling the complexity of cities to inform urban interventions and facilitate data-driven planning (Li et al., 2023b). Unsupervised Learning (UL) infers patterns from unlabeled data, unlocking the potential to further understand dynamic and large-scale data in urban research. In the prevailing trend of interdisciplinary GeoAI research (Liu and Biljecki, 2022), UL is crucial for learning the rich semantics of spatial representations and spatial data infrastructures.

Labeling panoramic images could be a time-consuming and expensive task due to the ultra-wide field of view and distorted elements. To address this issue, existing Domain Adaptation (DA) methods can be classified into two main types: Semi-supervised Domain Adaptation (SDA) and Unsupervised Domain Adaptation (UDA). SDA assumes that labeled data from the target domain can be used in addition to labeled data from the source domain to train/adapt the model. In contrast, the latter does not require labels from the target domain but explores the similarity between the two in the data distribution. This scheme adapts the model from the source domain to the target domain, improving model generalization to unseen domains (Oza et al., 2023). In this context, UDA is selected as the primary solution for the task.

Unsupervised domain adaptation mainly includes self-training and adversarial learning. Self-training utilizes unlabeled data in the target domain to generate pseudo-labels for training by the source domain model. After iteration, the model then adapts to the target feature distribution to improve its generalization ability (Zhang et al., 2019; Li et al., 2022a; Huo et al., 2022). Adversarial learning is a technique that utilizes Generative Adversarial Networks (GANs) (Goodfellow et al., 2014). GANs consist of a generator network and a discriminator network. The generator network produces target samples, while the discriminator network distinguishes between the generated and real samples. Through an iterative process, adversarial relationships are formed to reduce the difference in the distribution of the source and target domains, ultimately enhancing the model's performance (Gallego et al., 2020; Chang et al., 2019). Both types of methods reshape the feature space to enable the generalization of classifiers trained on transformed source data to target data (Kouw and Loog, 2019).

To tackle the issue of insufficient labeled data, it is crucial to explore Unsupervised Domain Adaptation (UDA) by utilizing sub-optimal but label-rich resources to train panoramic models. This involves adapting a pinhole image to a target panoramic image (Farahani et al., 2023). A multi-stage prototype adaptation method is proposed for panoramic image data, simultaneously considering both approaches mentioned above. The gap between different domains is bridged by self-learning to deeply mine different feature spaces under multiple datasets, source labels in the output space, and pseudo-labels in the target domain. Meanwhile, adversarial learning is employed to warm up the optimal model at each stage, facilitating the synthetic transfer of features between different fields of view.

3. Methodology

This section presents the specific details of the methods employed. In this article, different symbols and abbreviations are employed to represent different concepts. Table 3 provides a comprehensive overview of the symbols and abbreviations, thereby facilitating a clear understanding of the symbols used.

3.1. Research framework

This paper presents a framework for measuring cities' Comprehensive Walkability Index (CWI) from an urban planning perspective. The framework focuses on three-dimensional urban walkability research. Walking behavior is influenced by various variables, including the quality of the street, the distribution of public facilities, and the willingness of the pedestrian to walk. To meet various travel requirements and simplify path selection, developing a comprehensive framework to evaluate the feasibility of walking is an effective solution. The research framework integrates the effects of mesoscale variables, such as street environment characteristics, and microscale variables, such as the condition of surrounding facilities Additionally, pedestrians' perceived willingness to walk is considered, and these variables are used to obtain a Comprehensive Walkability Index (CWI), which provides a more scientific and reliable evaluation method. The evaluation architecture diagram for this study is shown in Fig. 1.

First, city-related data are collected, including road network data from Open Street Map (OSM), Panoramic Street Image (PSVI) data, and Point of Interest (PoI) data from the Application Programming Interface (API) provided by Baidu Maps. Additionally, land use type data of the study area is collected using the Urban Essential Land Use Categories (EULUC) data provided by EULUC-China (Gong et al., 2020).

To evaluate the ecological quality of roads, it is necessary to classify and predict each element type in the panoramic image data. A panoramic semantic segmentation model with domain adaptive processing is proposed to address image distortion and annotation scarcity in panoramic data. The TDDPassNet model utilizes multiple datasets for source and target domain adaptation and enhances the model's adaptability to the target domain through adversarial training methods. Through unsupervised domain adaptation, labeled data of panoramic images of the study area were obtained. The environmental quality assessment indices of the target streets were also acquired based on the percentage of different visual elements in the TDDPassNet output.

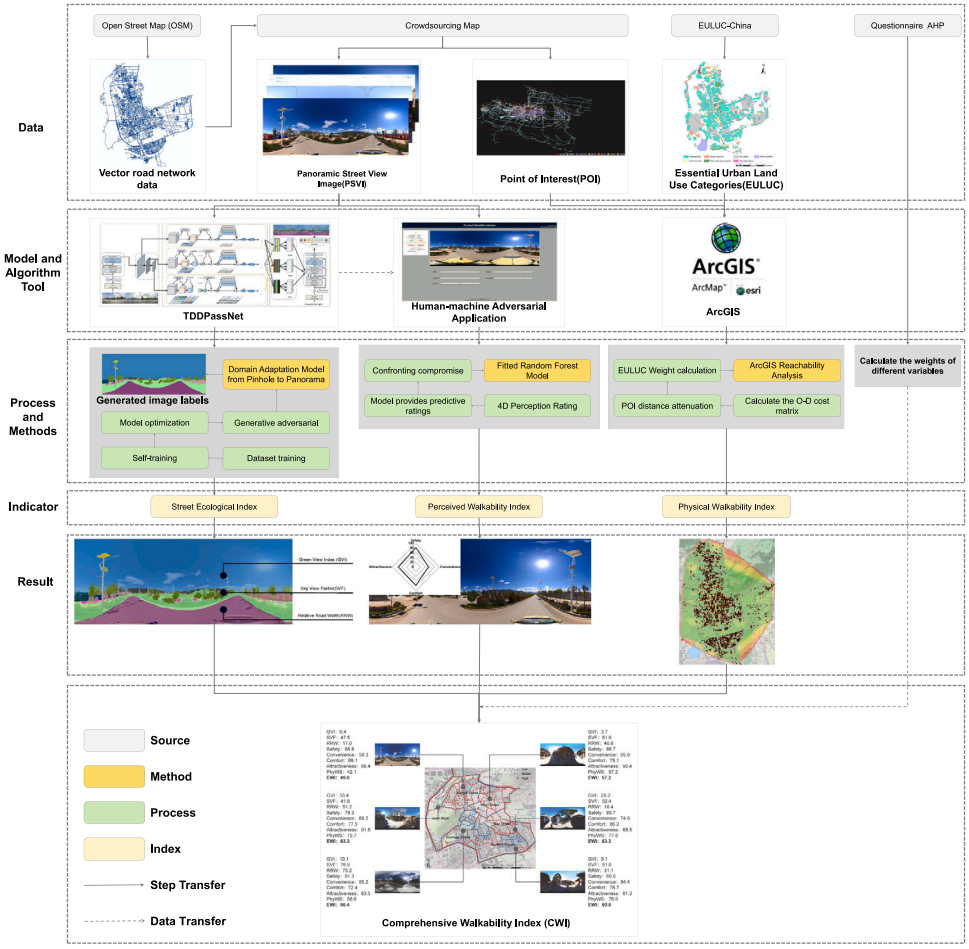


Fig. 1. Evaluation Architecture.

Table 3
Symbols and Explanations.

Symbol	Explanation
θ	Latitude
ϕ	Longitude
f_i	Feature mapping
S_i	Downsampling step size
C_i	The Channel size
$I^{s/t}$	Source or target domain datasets, s and t represent the source domain and target domain, respectively
${x_{(i,j)}}^{s/t}$	Image x, i, j represents pixel information, s and t represent the source and target domains
$y^s_{(i,j,n)}$	Source Domain $x^s_{(i,j)}$, The label information for (i, j) represents category n
$\hat{y}^t_{(i,j,n)}$	Pseudo label category, target domain $x^t_{(i,j)}$ is predicted as information for category n
${p_{(i,j,n)}}^{s/t}$	The probability representation of pixel ${x^{s/t}_{(i,j)}}$ being predicted as the nth class
$\rho_{x,y}$	Pearson correlation coefficient, X is a certain perceptual dimension, Y is a certain visual element
$\mathrm{SEQ}_{\mathrm{index}}$	Street Ecological Quality Index
PhyWI_i	The physical walkability index of the sample point
PerWI_i	The Physical walkability index of the sample point

To evaluate the walkability of an area, a human–machine adversarial web application was created using a group decision-making process. Volunteers completed an online perceptual walkability assessment and input the data into a random forest model for training. The model used the proportion of each visual element as a reference variable to determine the influence of each element on perceptual walkability. Additionally, the POI and EULUC data were input into ArcGIS to analyze the spatial accessibility of pedestrians and obtain an objective physical walkability index.

Finally, the weights of each index are determined using the Analytic Hierarchy Process (AHP) (Saaty, 1988), and then the street ecological quality index, the perceived walkability index, and the physical walkability index were combined into a Comprehensive Walkability Index (CWI).

3.2. Data collection and processing

This paper utilizes four types of data: road network data, PSVIs, POIs, and EULUC.

Open Street Map (OSM) provides road network data, which requires processing such as rasterization, vectorization, and geo-alignment to convert latitude and longitude coordinates to the local coordinate system. This enables subsequent sampling of street points and image

Table 4
Classification and Number of Facilities at Points of Interest.

Facility classification	Place types of Baidu Map API	Weight	Number
Entertainment	KTV, ……., Playground, Cinema	1	517
Traffic	Aircraft, ……., Parking Lot, Coach	3	1,323
Estate	Villa, ……., dormitory, Industrial Park	1	527
Hospital	First-Aid Center, ……., Specialized Hospital	1	720
Attractions	Scenic Spot, ……., Memorial Hall, Aquarium	2	404
Auto	Charging Station, ……., Auto Repair, Auto Sale	3	1,032
Life services	Lotto, ……., Information, Post Office, Agency, Public Toilet	1	2,842
Education	Museum, ……., Middle School, Vocational School	1	760
Shop	Emporium, Minimart, Supermarket, Mall, Grocery	1	6,783
Sport	Stadium, Spa, Natatorium	2	275
Finance	ATM, Insure, Investment, Bank	2	281
Hotel	Budget Hotel, ……., Starred Hotel	1	5,268
Restaurant	Teahouse, ……., Chinese Food	2	5,510

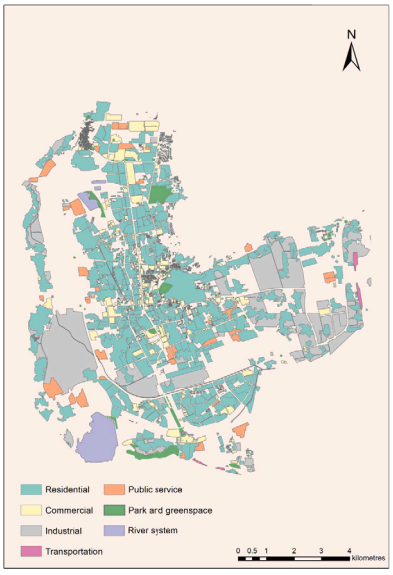


Fig. 2. The essential classification of urban land use in the ancient city.

processing. In our method, the vectorized road network map is sampled at every 30 m street point, with corresponding parameters such as pitch and field of view width being set to ensure image acquisition quality and accuracy. Subsequently, a program is developed to access the Baidu Street View map application programming interface (API) and collect the PSVIs with corresponding point of interest (POI) data. The categorized acquisition of POI data is presented in Table 4. The essential urban land use classification (EULUC) data for the study area is obtained from EULUC-China. The land use type of the ancient city of Lijiang City area from China is depicted in Fig. 2. EULUC-China utilizes 10-megapixel satellite imagery, open street maps, nighttime lighting, PoIs, and Tencent social big data as input features to provide a new national urban land use map.

Finally, to ensure the reliability of subsequent urban planning analysis and deep learning model training, the collected data, including PSVIs, POIs, and EULUC, are integrated and classified. The data is also cleaned to remove any invalid data that may have been accessed.

3.3. Street ecological quality evaluation based on panoramic image semantic segmentation models

After the data processing is completed, the PSVIs of the study area are next fed into our proposed TDDPassNet model (which will be introduced in the following sections) for semantic segmentation prediction of the panoramic images. TDDPassNet extracts valuable information about the 19 visual elements present in each panoramic image, including their respective categories and proportions. In order to perform the street ecological quality assessment, we first need to calculate the specifics of variables such as Green View Index, Sky View Factor, and Relative Road Width. The calculation formula is shown in (1):

$$SEQ_{\text{index}} = \frac{\text{Area}_{p\_\lambda}}{\text{Area}_{t\_\lambda}} \times 100\% \tag{1}$$

where index represents the calculated index considering three variables crucial for assessing the ecological quality of streets under panoramic view, namely the Green View Index (GVI), Sky View Factor (SVF), and Relative Road Width (RW). Area_{p_λ} denotes the total pixel count of the indexed element captured in panoramic image λ at each sampling point, while Area_{t_λ} signifies the total pixel count in a sampling point containing 19 elements in the PSVIs. Area_{p_λ} and Area_{t_λ} are introduced and employed to characterize the image at the sampling point. The resulting values for these three visual element variables require subsequent weighting to derive the street's ecological quality index.

3.3.1. Transformer-based doubly deformable panoramic semantic segmentation network (TDDPassNet)

The Transformer model is a deep learning model based on a self-attention mechanism. It has achieved great success in the field of natural language processing. Its unique architecture and advantages have also made it a popular technology in computer vision and urban planning research. This study presents TDDPassNet, an architecture based on Transformer (as shown in Fig. 3), which mitigates target distortion and warping effects in panoramic semantic segmentation. A feature pyramid structure is added between the encoder and decoder for multi-scale spatial feature extraction, with three stages of layers 2,4,8, each sampling the image at a different resolution. During the process of image segmentation, TDDPassNet first patterns the input image with the shape H×W×D. The image is passed through three layers of maxpool of different sizes, which are used to construct spatial feature maps to map features $f_i \in (2, 4, 8)$ for different resolution scales. Downsampling is then performed for processing steps $S_i \in (8, 16, 32)$ corresponding to different channel sizes $C_i \in (128, 256, 512)$. In each scale space, three nested convolutional layers with the same filter size are used to explore each scale's depth space feature maps. The multi-scale feature maps f_i are first transformed into a uniform shape of $\frac{H}{f_i} \times \frac{W}{f_i} \times D$ and then inputted into the decoder. The number of embedding channels D is set to 128. The prediction layers output the final semantic segmentation results that match the size of the input image and are based on the number of semantic categories for their respective tasks.

When predicting panoramic image data, Kim and Woo (2022) considered the issue of distortion and deformation commonly found in such images. To mitigate this problem, Kim performed image cropping, retaining only the main undistorted body part in the center of all cropped images. This method could effectively extract semantic information from the most information-intensive parts of an image, but it

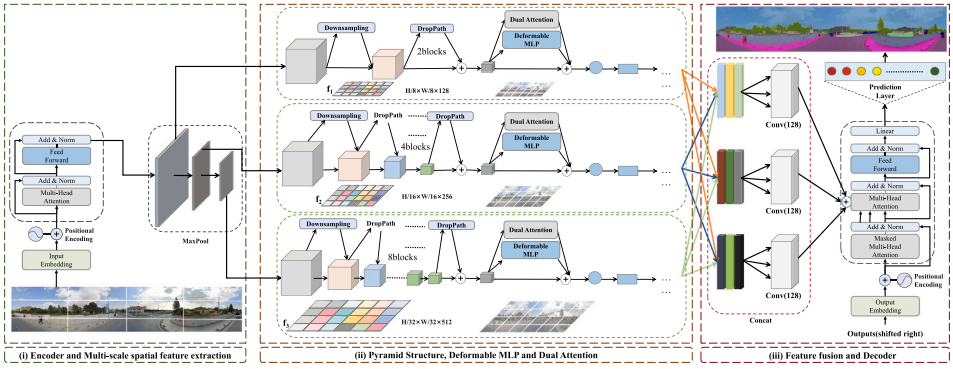


Fig. 3. Schematic of the architecture of Transformer-based doubly deformable panoramic semantic segmentation network (TDDPassNet).

may result in a lack of information from the edge parts. To address this, we take a holistic approach by first transforming the 360° image data from a spherical coordinate system to ensure accurate acquisition of global information and mitigate errors caused by image distortion. The panoramic image is originally in the coordinate system of latitude $\theta \in [0, 2\pi]$ and longitude $\phi \in \left[-\frac{1}{2\pi}, \frac{1}{2\pi}\right]$. Eq. (2) converts it to the Cartesian coordinate system of x and y. The center latitude and longitude are $(\theta_0, \phi_0) = (0, 0)$. However, there is also considerable distortion behind this equirectangular projection. As per the distortion transformation formula (3) proposed by Lai et al. (2017) for panoramic videos, the input panoramic image is represented by a set of coordinates (θ, ϕ), and the model divides the image into k slices, $(\theta^k, \phi^k), \forall k \in [1, N]$, where N represents the number of pixels in the region. Eq. (4) is derived from the equirectangular projection transformation. It is evident that there is a strong correlation between the image distortion and $cos(\phi)$. This means that any pixel position of the image from $\phi = 0$ could be transformed with the corresponding distortion.

$$\begin{cases} x = (\theta - \theta_0)\, cos(\phi_0), \\ y = (\phi - \phi_0), \end{cases} \tag{2}$$

$$(\theta, \phi) = \left(\frac{\sum_{k=1}^{N} \theta^k}{N}, \frac{\sum_{k=1}^{N} \phi^k}{N} \right) = \frac{\left| \frac{\partial x}{\partial \theta} \cdot \frac{\partial x}{\partial \phi} \right|}{\left| \frac{\partial y}{\partial \theta} \cdot \frac{\partial y}{\partial \phi} \right|} \tag{3}$$

$$(x, y) = \frac{cos(\phi) \left| \frac{d\theta}{d\phi} \right|}{\left| \frac{dy}{dx} \right|} = \frac{cos(\phi)}{(\theta, \phi)} \tag{4}$$

where x and y are the Cartesian coordinates. θ and ϕ are the latitude and longitude coordinates of the panoramic image.k is the slice of the image. N is the total number of pixels in the image. $\frac{d\theta}{d\phi}$ represents the change in latitude with respect to longitude. $\frac{dy}{dx}$ is the derivative of y with respect to x

Using the formulation of the three equations mentioned above, we incorporate a deformable MLP module (Zhang et al., 2022) onto the multi-scale feature map $U \in R^{\frac{H}{f_i} \times \frac{W}{f_i} \times D}$ extraction. This helps to preserve the object's shape and spatial layout, resulting in a more accurate and consistent 2D reference index of the feature map and panoramic representation. The effect after reprojection is shown in Fig. 4. Unlike (a) standard Transformer patch extraction, our TDDPassNet (b) could embed and extract patches along the distortion degree after deformation and, simultaneously, consider the distortion to perform good panoramic image segmentation.

The feature map $U \in R^{\frac{H}{f_i} \times \frac{W}{f_i} \times D}$ consists of three dimensions: width, height, and channel. Among them, width and height can build spatial features, while channels represent regional features in space. Obviously, the spatial position information focusing on key features could help enhance images' feature representation in the spatial domain. At the same time, highlighting channels with relatively rich information

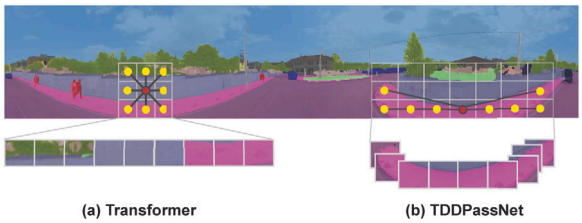


Fig. 4. Schematic Diagram of Standard Transformer and TDDPassNet Projection Transformation and Patch Embedding.

could enhance the meaningful features of each spatial area. Therefore, the Dual Attention (Woo et al., 2018) mechanism has been introduced to combine channel attention and spatial attention. This helps to learn key local features in panoramic images and improve the semantic segmentation performance of panoramic image scenes. This method is applied to different scales of spatial feature maps U, aiming to enhance the multi-scale and multi-sensor performance of advanced features in local semantics.

The fundamental methodology underlying the model could be discerned in Algorithm 1. The proposed TDDPassNet algorithm exhibits a computational complexity of $O(H \times W \times D \log(D))$ in time and $O(H \times W \times D)$ in space. This complexity is attributed to the transformation, feature extraction, and attention mechanisms employed for panoramic semantic segmentation, which are represented by H, W, and D, respectively, and account for the height, width, and depth of the input image.

3.3.2. Multi-stage prototyping method for urban scenes with unsupervised domain adaptation

A multi-stage prototype method for unsupervised domain adaptation of urban scenes is proposed to address the problem of missing annotations in PSVIs data (Fig. 5). Our approach involves training TDDPassNet on different datasets as source domains in multiple stages. By preserving the optimal prototype of the first stage and mapping it to the unlabeled Lijiang PSVIs dataset, intermediate domains are added to improve model performance.

Previous studies commonly use a single-stage, single-prototype method for the domain adaptation process in image data. However, those approaches struggle to learn the potential relationship between the source and target domains and to accurately fit the label requirements of the target domain.

To address the lack of complete and sufficient sample size panoramic image datasets in the field of panoramic image data, the utilization of SynPass synthetic panoramic image datasets as an intermediate domain prototype is proposed. This improves the model's ability to accurately identify patterns between pinhole images and

Algorithm 1 TDDPassNet: Transformer-based Panoramic Semantic Segmentation

Require: Image: input image with shape $H \times W \times D$
Ensure: Semantic segmentation results

function TDDPassNet($Image$)
  Perform transformation from spherical coordinates to Cartesian coordinates:
  if $Image$ is panoramic **then**
    $x = (\theta - \theta_0)\cos(\phi_0)$
    $y = \phi - \phi_0$
  end if
  Compute coordinates' mean:
  $N \leftarrow$ total number of pixels
  $\theta \leftarrow \frac{\sum_{k=1}^{N} \theta^k}{N}$
  $\phi \leftarrow \frac{\sum_{k=1}^{N} \phi^k}{N}$
  if $Image$ is panoramic **then**
    Compute distortion correction factor:
    $x, y \leftarrow \frac{\cos(\phi)|\frac{d\theta}{d\phi}|}{|\frac{dy}{dx}|}$
  end if
  Incorporate deformable MLP module onto multi-scale feature map extraction:
  $U \in R^{\frac{H}{f_i} \times \frac{W}{f_i} \times D}$
  for $i \leftarrow 1$ to D **do**
    Apply deformable MLP to $U[:, :, i]$
  end for
  Apply Dual Attention mechanism to enhance semantic segmentation:
  if Panoramic image **then**
    Combine channel attention and spatial attention for U
  else
    Apply regular attention mechanism to U
  end if
  return Semantic segmentation results
end function

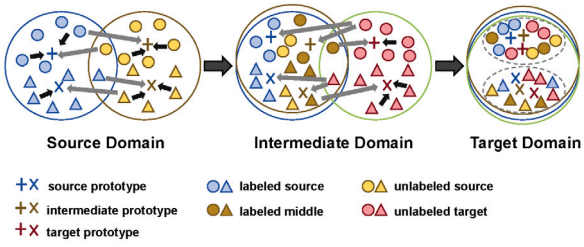


Fig. 5. Demonstration of multi-stage cross-domain adaptation process of unsupervised domain adaptation.

panoramic images in the source and target domains, ultimately better fitting the pseudo label to the final target domain.

Inspired by the Domain Antagonism Neural Network (DANN) (Gallego et al., 2020), we classify and label the target data with transformed labeled source data through representation matching. Our approach uses confrontation learning to seek a representation that makes distinguishing differences between different domains when marking samples impossible, effectively reducing feature differences during cross-domain migration and improving the model's generalization ability.

First, Cityscapes (Cordts et al., 2016) pinhole image dataset and its tag image $I^s = \left\{(x^s, y^s) \mid x^s \in \mathbb{R}^{H \times W \times 3}, y^s \in 0, 1^{H \times W \times n}\right\}$ are used as the source domain. where x^s represents the input images and y^s symbolizes the corresponding segmentation masks with n classes. The target domain is SynPass (Zhang et al., 2022) composite panoramic image dataset $I^t = \left\{\left(x^t\right) \mid x^t \in \mathbb{R}^{H \times W \times 3}\right\}$ (the label information of SynPass dataset is not used at this time). Consistency in shared categories is aimed to be ensured during model migration, assuming that the two datasets contain the same n categories. Our approach aims to adapt the model from the source domain I^s, which includes labeled data, to the target domain I^t, which lacks label information for its images, while maintaining consistency across n shared classes between the two domains. To achieve this, we employ a panoramic segmentation loss function, denoted as Eq. (5), which is used to fit the training model in the source I^s domain:

$$L_{ps}^{s}\left(y_{(i,j,n)}^{s}, p^{s}\left(y_{(i,j,n)}^{s} \mid x_{(i,j)}^{s}\right)\right) = -\sum_{i,j,n=1}^{H,W,n} y_{(i,j,n)}^{s} log\left(p_{(i,j,n)}^{s}\right) \tag{5}$$

$$L_{st}^{t}\left(\hat{y}_{(i,j,n)}^{t}, p^{t}\left(\hat{y}_{(i,j,n)}^{t} \mid x_{(i,j)}^{t}\right)\right) = -\sum_{i,j,n=1}^{H,W,n} \hat{y}_{(i,j,n)}^{t} log\left(p_{(i,j,n)}^{t}\right) \tag{6}$$

where the probability that pixel $x_{(i,j)}^{s}$ is predicted to be the nth class is denoted by $p_{(i,j,n)}^{s}$. The term $p^t\left(\hat{y}_{(i,j,n)}^{t} \mid x_{(i,j)}^{t}\right)$ denotes the probability of the predicted categories derived from training and validating the model on the source domain, encompassing the categories predicted as 1 through n, the height (i-H), and width (j-W) of the patch. During the fitting process, the conventional self-training loss calculation scheme (6) is utilized to adapt the pre-trained model to the target data. In this context, the pseudo-label of the target domain pixel, denoted by $\hat{y}_{(i,j,n)}^{t}$, is used to represent the predicted label for that pixel $x_{(i,j)}^{t}$. The pseudo-label is determined based on the most probable class in the model prediction, as represented by $\hat{y}_{(i,j,n)}^{t} = argmaxp^{t}{}_{(i,j,n)}$.

The results of the preceding stage are utilized as a prototype, employing the identical objective function calculation to adjust the original SynPass target domain dataset, which incorporates label information, into an intermediate domain. This intermediate stage prototype is consistent with the categories of Cityscapes, combining the label category vectors to ensure pseudo-label consistency. Subsequently, the unlabeled Lijiang panorama dataset is employed as the ultimate target domain in a multi-stage, multi-scale feature extraction and embedding procedure to construct the prototype. This approach enhances the robustness and expressiveness of the final target domain labels.

3.4. Human–machine adversarial technology for perceived walkability evaluation

In this study, a scoring framework for perceived walkability based on the four levels of urban walking perception is constructed. The four categories are safety, convenience, comfort, and attractiveness. These four evaluation dimensions were determined by reviewing the effective impact on pedestrian demand in previous studies (Gebel et al., 2009; Guzman et al., 2022; Farahani et al., 2023). These four levels of need represent the key aspects of pedestrian psychological choice in the walking decision process. The currently widely used evaluation standards, which are also the unified standard for the subsequent recruitment of volunteers to score the human–machine adversarial, are integrated by Fig. 6.

The evaluation criteria for the four dimensions of perceived walkability are shown by Fig. 6. Safety is a fundamental need for pedestrians, referring to the perceived threat of traffic accidents and crime. Measures such as modifying pavement morphology, controlling land development, and improving infrastructure integrity could enhance pedestrian safety. Convenience is also crucial for pedestrians when choosing destinations and planning routes, and it is closely related to land use type and infrastructure diversity. Urban street design should prioritize creating a comfortable walking environment and addressing pedestrians' concerns. Pedestrian comfort is expressed through their satisfaction with the street environment, including elements such as greenery, cleanliness, and unobstructed views of the sky. Attractiveness is also important for pedestrians to find the street environment pleasant

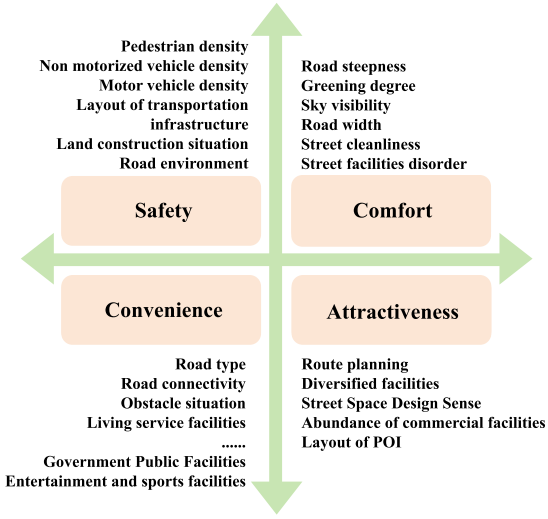


Fig. 6. Evaluation Criteria for Perceived Walkbility.

and appealing, including clear directional signage, abundant public spaces, and various commercial facilities. Optimizing these aspects could enhance the appeal and functionality of the street environment.

To assess urban walkability perception in our study area, we developed a web application[2] for online assessment, data collection, and model training, leveraging the concept of group decision-making. This application was used by volunteers familiar with the region's socio-economic context or possessing knowledge in urban street design research within a human–machine adversarial scoring framework.

The working mechanism of evaluating perceived walkability is illustrated in Fig. 7. We enlisted 40 volunteers to rate the walkability perception of city streets, standardizing their evaluation criteria through initial training to minimize discrepancies in their assessments. The criteria were made readily accessible within the application interface to guide volunteers during their evaluations. The assessment involved rating images from the PSVIs dataset of Lijiang City's ancient area, displayed randomly to each volunteer on a scale from 0 to 100 across four walkability criteria.

Our framework incorporates a Random Forest (RF) model to correlate visual elements in images with user ratings. This is achieved by using feature vectors obtained through semantic segmentation via our TDDPassNet. The system then uses the ratings from the first 50 images to refine the RF model, predicting subsequent images' scores. The model is recalibrated if there are discrepancies between predicted and volunteer scores exceeding 10 points across more than five images.

Images rated multiple times are assigned a median score to ensure reliability. The RF model uses two-thirds of its samples for training and one-third for out-of-bag error assessment. A successful human-machine compromise is reached when the difference between predicted and actual scores stays within ±5 points without exceeding a 10-point discrepancy in five consecutive ratings. Upon achieving an out-of-bag verification error below 5 points after evaluating 500–1000 images, the process concludes, generating the final perceived walkability index results.

$$\rho_{x,y} = \frac{Cov(X,Y)}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^{n} (x_i - \mu_x)(y_i - \mu_y)}{(n-1)\sigma_x \sigma_y} \tag{7}$$

where $Cov(X,Y)$ represents the covariance between variables x and y, and σ_x and σ_y represent the standard deviation of variables x and y, respectively, Eq. (7) defines the correlation coefficient between two variables, x and y. This coefficient is calculated using the Pearson correlation coefficient formula, which is obtained by dividing the covariance between x and y by the product of their standard deviations. To analyze the correlation between the effects of the four perceptual dimensions and the 19 visual elements on evaluation results, we extracted the parameters of the RF model. This calculation allows for an objective interpretation of the correlation between specific perceptual dimensions and visual elements in the street environment, aiding urban planners and designers in making more informed improvements at a detailed level.

3.5. GIS-based physical walkability evaluation

The Common Walk Score system (Carr et al., 2011) is a free, public, web-based tool for measuring local walkability available in the United States, Canada, Australia, and New Zealand. Walk Score is scored based on the accessibility of various POI facilities at each address (Horak et al., 2022). If these facilities are within 400 m, they receive a full score, which decreases as the distance increases until facilities within 2000 m no longer count towards the score. On this basis, the score is calculated based on the network distance from the address to the destination and the facility's weight. Each facility has a different weight, and the number of facilities taken into account varies (see Table 4). The weights and number of facilities were determined based on previous walkability studies. The weights of each facility were then divided, and a weighted sum was applied to them to obtain a standardized score ranging from 0 to 100.

In this study, physical walkability is analyzed using the weight values of POI facilities as independent variables, considering the effect of EULUC on walkability in different regions. First, a GIS network dataset was constructed from POI data, OSM road network data and EULUC data, and based on the facility classification validated by Kim and Jin (2023), the classification of POI facility data. As a result, each facility in Table 4 has slightly different facility accessibility values and weights. We then calculate the number of facility types within walking distance of 400/800/1200/1600 and 2000 m from each sampling point on the road. At the same time, referring to Horak et al. (2022), different EULUCs, as well as road network data types, are assigned gradually decreasing distance attenuation coefficients for different walking distances to measure the walking speed and walkability of each sampling point to the POI facilities within the different site types. Finally, a physical walking index for each sample point is obtained by multiplying the corresponding distance attenuation factors and overlapping the weights.

$$PhyWI_i = \sum_{i,j=1}^{n} N_{ij} W_j f(d) l(k) \tag{8}$$

where $PhyWI_i$ is the physical walkability index of the sample point, i is the type of classified facilities, N_{ij} is the number of facilities, W_j is the weight of class j facilities, $|d|$ is the distance between the facilities and the sample point, and $f(d)$ is the corresponding distance attenuation coefficient. $l(k)$ refers to the impact of land use type k on the walking score of sampling point i.

3.6. Comprehensive Walkability Index (CWI) evaluation

To calculate the Comprehensive Walkability Index (CWI), the results of multiple indices and variables are first normalized to 1–10. Then, an AHP approach is used to calculate weightings for each indicator by using pairwise comparison through a questionnaire. The Consistency Ratio (CR) in AHP serves as an indicator for evaluating the consistency of judgment matrices, and it is calculated by comparing the obtained Consistency Index (CI) with a Random Index (RI). If the computed CR value is less than or equal to 0.1, it indicates that the judgment matrices

[2] Human–Machine Adversarial Web Application: https://github.com/Qingkongmu/Human-machine-adversarial-aid-evaluation.

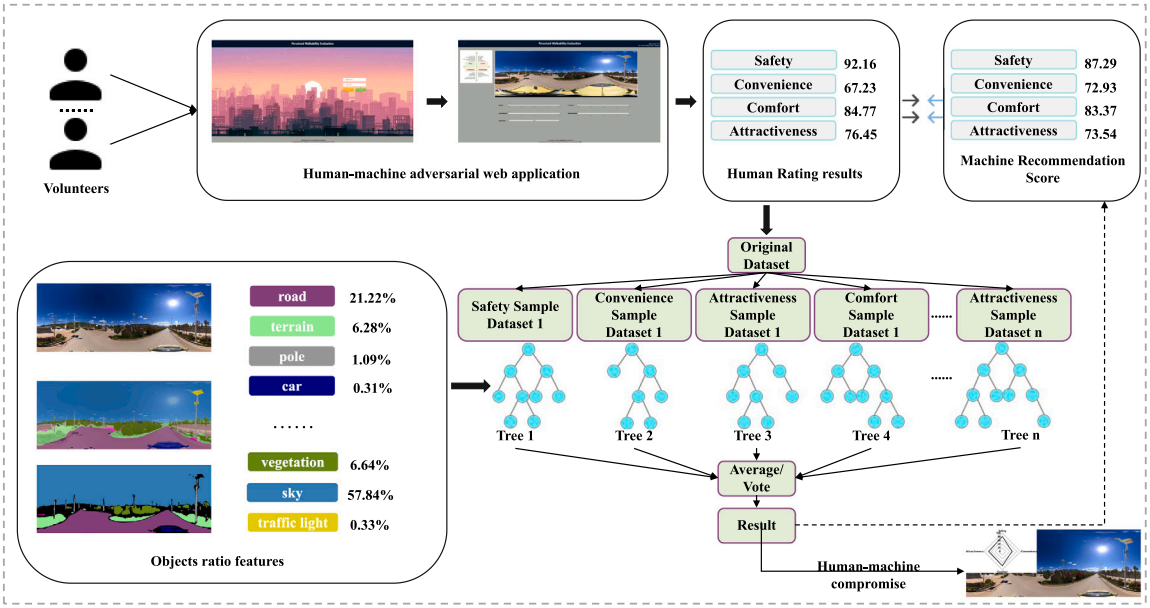


Fig. 7. Perceived Walkability Evaluation Framework Based on Human–Machine Adversarial Techniques.

pass the consistency test; otherwise, subjective judgments need to be modified.

$$CWI = w_a SEQ_{\text{index}} + w_b PhyWI + w_c PerWI = w_1 x_1 + w_2 x_2 + \cdots + w_n x_n \tag{9}$$

where x represents the detailed influencing variable in each dimension, w represents the weight corresponding to the variable determined using Analytic Hierarchy Process (AHP), and n represents the total number of influencing variables, Eq. (9) defines the Comprehensive Walkability Index (CWI). This index is calculated by summing the weighted influencing index of the three dimensions: the Street Ecological Quality Index (SEQ_{index}), the Physical Walkability Index ($PhyWI$), and the Perceived Walkability Index ($PerWI$). Each dimension may encompass multiple variables affecting it. The results of each streetscape sampling point in the study area were input into ArcGIS for visualization and analysis.

4. Experiment

Our method was applied to conduct experiments in the ancient city of Lijiang, China. Section 4.1 provides a basic description of the study area; Section 4.2 presents the results and analysis of CWI. In Section 4.3, experiments on the proposed TDDPassNet model are conducted. Section 4.4 verifies the effectiveness of the proposed domain adaptation method. Finally, Section 4.5 explores the relationship between the perception dimension and the visual elements in the street building environment.

4.1. Study area

The paper focuses on the ancient city of Lijiang City area, Yunnan Province, China, known for its history and scenery. On the left side, the administrative location of Lijiang in China is displayed by Fig. 8, while on the right side, the layout and basic distribution of land use in the ancient city are shown.

In recent years, the government of the ancient city district has focused on developing the tourism industry and building an ecological civilization city based on its geographical conditions (Gao et al., 2020). A series of urban renovations and upgrades have been carried out to improve street infrastructure, enhance environmental functions, and optimize the cityscape. The improvement of the walkability index of the ancient city could enhance the well-being and habitability of residents and tourists. Furthermore, the region has ample PSVI data, which could increase the reliability of our findings, making it a suitable target city for our study.

4.2. Comprehensive Walkability Index (CWI) results and analysis

A questionnaire was administered to 40 volunteers who were familiar with the historical and cultural background of the study area or had experience in urban planning research. The questionnaire provided a detailed description of each variable and asked respondents to score and compare all variables under the three indices dimensions: Green View Index (GVI), Sky View Factor (SVF), Relative Road Width (RRW), Physical Walking Index (PhyWI), safety, convenience, comfort, and attractiveness. Respondents were asked to rate each variable on a scale of 1 to 10 to generate the final score matrix. The geometric average of 40 matrices is taken as a summary matrix, and AHP analysis is performed. The resulting variable weight results are shown in Table 5. It is important to note that the model passed the consistency judgment (Consistency Ratio (CR) = 0.0264< 0.1), indicating that the weight obtained is reliable.

4.2.1. Results of the three-dimensional study area Walkability index

The results of various influential variables in the study area are depicted in Fig. 9. The sampling points of the streets are marked with three different colors to indicate their respective scores. The Street Ecological Quality Index (SEQ_{index}) is obtained by segmenting PSVIs of the study area through the TDDPassNet. This index includes GVI (Green View Index), SVF (Sky View Factor), and RRW (Roadway Ratio Weighted). The Fig. 9 shows that the vegetation coverage in the outer ring area of the ancient city of Lijiang City area is not higher than that in the city center area from the pedestrians' point of view. This is because the surrounding areas ①, ②, ③ are mainly mountainous and Gobi terrain. Sporadic areas with a HIGH rating are located in the central section ④, indicating that the degree of urban greenery in this area needs further enhancement. The distribution of high-grade areas for SVF and RRW is mainly concentrated in the periphery ①, ②, ③, while low-grade areas are primarily located in the city center ④. This may be due to the high density of buildings obstructing pedestrian sightlines in the city center, while the outer ring consists mostly of road lots. Although these roads are wider and have better sightlines,

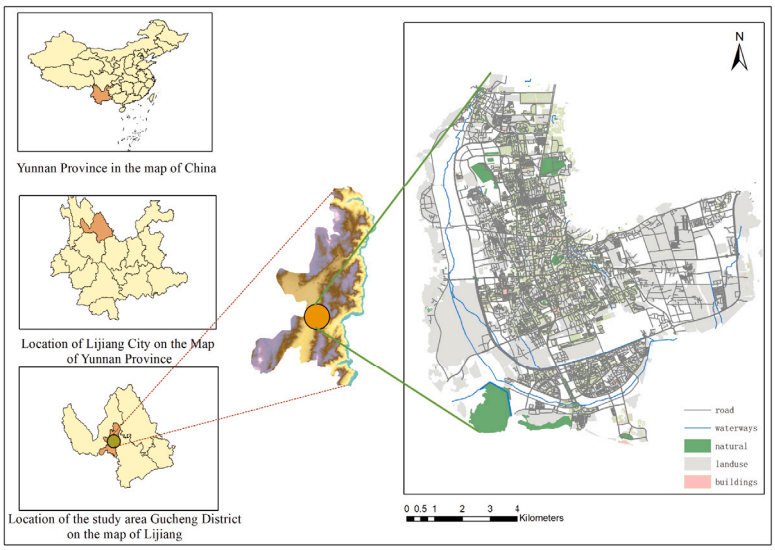


Fig. 8. Schematic Diagram of Study Area.

Table 5
Weights and consistency evaluation results of the CWI using AHP. (λ_{max} = 4.0708, CI = 0.0349, RI = 1.32, CR = 0.0264).

Indices	Street Ecological Quality Index			Physical Walkability Index	Perceived Walkability Index			
Variable	GVI	SVF	RRW	PhyWI	Safety	Convenience	Comfort	Attractiveness
Weight	0.1086	0.1937	0.0904	0.2540	0.0902	0.0836	0.0640	0.1155

they may not be suitable for walking. Based on the results of the AHP analysis and subsequent walkability assessment, it can be concluded that higher scores on certain variables do not necessarily have a greater impact on pedestrian travel.

The Physical Walkability Index (PhyWI) indicates that higher scores are typically found in areas with better topography and a concentration of businesses ①. These areas often contain parks, hospitals, stadiums, and various food and drink facilities, suggesting that streets near large public services have better physical value. These areas generally have better location and spatial quality and are home to relatively high densities of people who prioritize the physical walkability of the street. Lower scoring areas are concentrated in the periphery of the city ②, ③, ④ and share common characteristics of complex road network patterns and low accessibility. The government could increase the deployment and construction of various amenities in lower-scoring areas.

The Perceived Walkability Index (PerWI) showed high subjectivity ratings for pedestrians when using the human–machine adversarial technique. Streets with higher scores for Safety and Comfort are mainly concentrated in the Midwest ① and ②, with scattered locations in the North ③. High perception scores for the Convenience and Attractiveness dimensions are concentrated in the South Central region ① and ②, and dispersed in the East ③ and North ④ regions. Considering the characteristics of the streets, it can be inferred that areas with high Safety scores generally have low traffic and good transport facilities. Similarly, areas with high Comfort scores feature landscape elements that enhance the vibrancy of the space and create a pleasant atmosphere. Finally, streets with high Convenience scores tend to have a relatively large number of buildings, pedestrian amenities, and walkable environments. Areas with high attractiveness often contain scenic views, upscale buildings, or landscapes, while low-score areas have featureless landscapes and homogeneous street compositions. Identifying different types of perceptions can provide valuable information for targeted urban planning and pedestrian space improvements.

4.2.2. Comprehensive Walkability Index (CWI) results for the study area

The results of the spatial distribution map of CWI obtained after weighted calculation are shown in Fig. 10. The mean CWI score for the study area was 62.229, with a standard deviation of 12.738. The CWI was classified into low, middle, and high grades using the natural discontinuity grading method. The rating results indicate that 33.99% of the sampling sites were rated as low, while 44.02% and 21.98% of the sampling sites were rated as middle and high, respectively. To elaborate on the specifics of the CWI scores, the city center area was intercepted for the case study, and some examples of the measurements are provided in Fig. 11. The overall CWI score profile for this area is excellent (area ① in Fig. 11). The area could be divided into six main neighborhoods, and we randomly selected a sampling point in each neighborhood. A panoramic street image was presented with a description of the numerical results. The numerical results comprise all factors that affect walkability and the final results obtained for the CWI.

In residential areas within the urban area, streets with high scores are mainly shown in Fig. 10, particularly in the main business circles ① and residential areas ②. The area ① was selected for the case measurement experiment and is included in Fig. 11. Streets adjacent to the traffic trunk line in the east ③ and west ④ with low perceived walkability scores have low overall CWI scores. The middle ⑤ and south ⑥ areas of the city have a high density of streets, but they also have good physical accessibility scores. The traffic flow and pedestrian flow in these areas are relatively balanced, and the CWI comprehensive evaluation is relatively high. Additionally, in the north ⑦, where the building density is low and the ecological quality score of the street is average, the CWI of the entire street is mostly medium. When implementing pedestrian-friendly infrastructure, the government should consider the spatial distribution of the CWI scores and take appropriate measures for different street situations. This will enhance pedestrian facilities in lower-score areas and encourage local residents to walk, ultimately improving the city's overall livability.

4.3. Comparison of Panoramic Semantic Segmentation Models

Existing studies on urban walkability based on SVIs have been found to utilize traditional models for pinhole images without proper model selection, as demonstrated by Table 2. Additionally, Section 2.2 mentions research programs that use undisclosed indicators in their

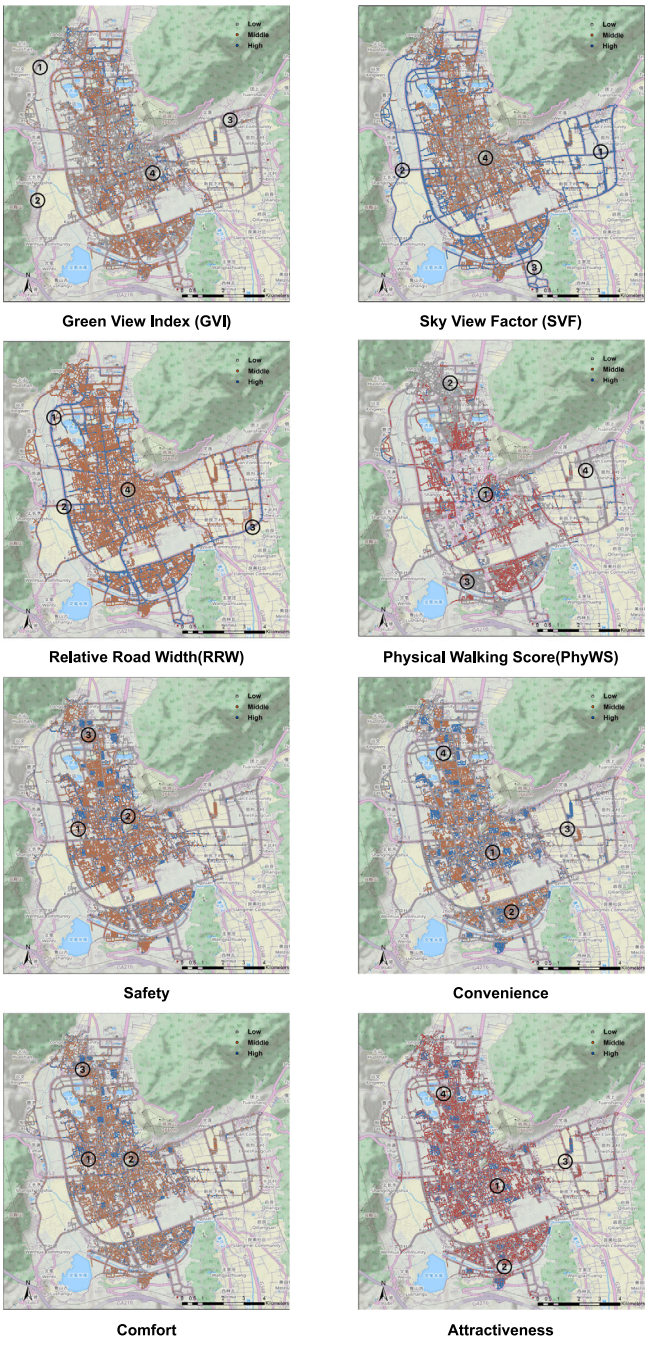


Fig. 9. Spatial distribution of the results of the Street Ecological Quality Index (GVI, SVF, RRW), the Physical Walkability Index (PhyWI), and the Perceived Walkability Index (Safety, Convenience, Comfort, Attractiveness), zoomed in to obtain a better view.

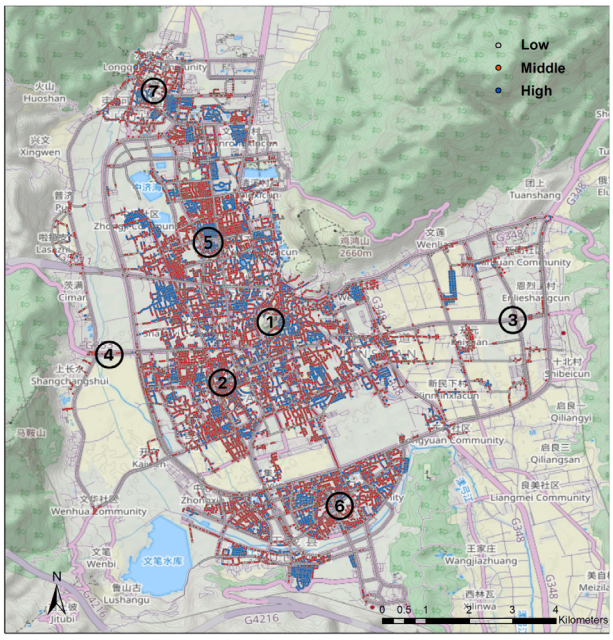


Fig. 10. Spatial distribution of Comprehensive Walkability Index (CWI) results.

Fig. 11. Results of panoramic street images with metrics data with random sampling conducted in the city center area.

models. In particular, in the PSVIs walkability study, the chosen models were not trained on panoramic image data and ignored the image and object distortions present in panoramic images. A decrease in the predictive ability of the models used can seriously affect the accuracy of the results of urban-related studies. To ensure credibility, following a general application paradigm that includes model testing, comparison, and selection is necessary.

4.3.1. Experimental setup for modeling methods

All algorithmic modeling was conducted on a single server, utilizing an NVIDIA GeForce RTX 3090 and an Intel(R) Xeon(R) Gold 6226R CPU @ 2.90 GHz. During training, a single GPU was used, with the epoch set to 200 rounds and the initial learning rate set to 5e-5. A cosine decay schedule was employed, incorporating a warm-up period of 5 epochs. The AdamW optimizer (Loshchilov and Hutter, 2017) was utilized, featuring a weight decay rate of 1e-4 and a batch size of 4 on the GPU. For image preprocessing, random adjustments ranging from 0.5–2.0, random horizontal flipping, and random cropping to 512×512 were employed.

Three main metrics are employed to compare model performance in extracting visual elements and to evaluate the accuracy of element classification results. (1) Mean Intersection over Union (mIoU) is the most commonly used evaluation metric in semantic segmentation. It measures the accuracy of the pixel positions of the visual elements. (2) Pixel Accuracy (pixAcc) calculates the accuracy of the exact matching ratio between predicted pixels and real labeled pixels. The accuracy of all semantic segmentation results is measured without distinguishing between categories. (3) Mean Accuracy (mAcc) is used to calculate the prediction accuracy of each category. Attention is focused on the discrepancies between the different categories, and the results of all

Table 6
Statistics on Dataset Information.

Data set information	Cityscapes	Synpass(Total	Cloudy	Foggy	Rainy	Sunny)	Pass
Train set	2975	5700	1420	1420	1420	1440	271
Test set	1525	1290	430	420	420	420	113
Val set	500	1290	420	430	430	410	116
Number of street elements	19	22	22	22	22	22	8

categories are averaged. The calculation formula is presented below:

$$mIoU_i = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{TP_i}{TP_i + FN_i + FP_i}\right) \tag{10}$$

$$pixAcc_i = \frac{\sum_{i=1}^{N} TP_i}{\sum_{i=1}^{N}(TP_i + FN_i + FP_i)} \tag{11}$$

$$mAcc_i = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{TP_i}{TP_i + FP_i}\right) \tag{12}$$

where N represents the number of image samples, TP_i represents the number of pixels in image i that are predicted to be positive samples and are positive samples, FP_i represents the number of pixels in image i that are predicted to be positive samples but are negative samples, and FN_i represents the number of pixels in image i that are predicted to be negative samples but are positive samples.

Additionally, the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) gauge the accuracy of estimating overall pixel values and the absolute difference between predictions and actual observations, respectively. The smaller value between the two indicates higher model accuracy when averaged over the test samples.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(TP_i - PP_i)^2} \tag{13}$$

$$MAE_i = \frac{1}{m}\sum_{i=1}^{N}\left|\frac{1}{N}\sum_{i=1}^{N} PP_i - TP_i\right| \tag{14}$$

where the variable N represents the number of pixels in a single image, while m represents the number of sample images. TP_i denotes the number of correctly identified element pixels in image i, and PP_i represents the number of predicted element pixels in image i.

4.3.2. Dataset

The information regarding the dataset used for the experiment is displayed in Table 6. The experiments were conducted following the original literature's division of datasets into training, test, and validation sets.

The Cityscapes dataset (Cordts et al., 2016) comprises 5000 high-definition pinhole street view images with a resolution of 1024 × 2048 pixels. The dataset covers a variety of weather and road scenes in 50 different cities and provides pixel-level annotations in 19 categories. This allows researchers to obtain fine-grained semantic segmentation labels for scene understanding and object detection. Each pixel is labeled with a specific category.

The SynPass dataset (Zhang et al., 2022) comprises 9080 synthetic panoramic street images generated using the CARLA simulator. The images provide 22 classes of labels and have a resolution of 1024 × 2048. To further improve the diversity of data, authors modulate the collected weather conditions. The weather conditions consist of sunny (25%), cloudy (25%), foggy (25%), and rainy (25%) conditions.

The Pass dataset (Yang et al., 2019) comprises 500 annotated panoramas from 25 cities across multiple continents for evaluation purposes. Additionally, it includes 2000 unlabeled panoramas from 40 cities that can be used to facilitate domain adaptation and create pseudo-labels.

4.3.3. Results and analysis of model indicators

To validate our proposed TDDPassNet in walkability research related to SVIs, particularly PSVIs, we followed the proposed research paradigm. Five expensive models were trained under different datasets in the same environment, referencing the common models used in various scenarios in Table 2. The available code provided in some of the original papers is accessed, and the performance of the models used in these scenarios is compared with that of TDDPassNet.

The variation curves of mean Intersection over Union (mIoU) for different method models on various datasets are shown in Fig. 12. mIoU is a widely accepted metric in the field of semantic segmentation. Upon observing the change curves, it is evident that TDDPassNet achieves higher mIoU values than other models under the same environment, indicating its superior accuracy and precision in image semantic segmentation tasks. TDDPassNet consistently performs across various image datasets, highlighting its stability and generalization capabilities. This further embodies its advantages in the field of panoramic image semantic segmentation.

The overall performance of the models is summarized, and key insights are provided in Table 7. The mean absolute error (MAE) and root mean square error (RMSE) increase as the distortion of the image changes increase, while accuracy-related metrics decrease. This may be due to distortion and object deformation inherent in panoramic images, resulting in an increase in average prediction error and a decrease in prediction accuracy. Regarding the models on the Pass dataset, there is an improvement in segmentation accuracy compared to the synthetic dataset Synpass. The discrepancy between the Pass and SynPass datasets may be attributed to the limited sample size of the former and the influence of varying weather conditions on the latter.

The consistent outperformance of TDDPassNet over the other models is illustrated by Table 7, showing an average increase in mIoU of 5.3% across all datasets. In addition, the results in the synthetic panorama dataset Synpass indicate that our model exhibits superior segmentation under extreme weather conditions compared to the other methods. Furthermore, the performance on Cityscapes, a pinhole image dataset, and Pass, a real panoramic image dataset, is commendable. Specifically, when evaluated on the pass dataset, compared to the widely used DeepLabV3+ model in urban walkability research, TDDPassNet exhibits a 4.81% decrease in MAE, a 3.22% decrease in RMSE, a 4.4% increase in mIoU, a 2.71% increase in pixAcc, and a 1.32% increase in mAcc. Additionally, the prediction time has been reduced by 14 ms. The superior performance of the TDDPassNet model could be attributed to its ability to extract image features at multiple scales, consider the diverse characteristics of image spatial structure and channel variations through the dual attention mechanism, and perform transformation mapping and deformation slicing processing of each element object in the image using the deformable MLP module. By effectively integrating these elements, the TDDPassNet model delivers impressive results on pinhole image data while achieving more accurate semantic segmentation of panoramic images.

4.3.4. Model ablation experiment and parameter experiment

Ablation experiments are conducted on the SynPass dataset to investigate the effectiveness of the key modules in the TDDPassNet model. To ensure fairness, all comparison methods used in the ablation experiments employ the same data enhancement methods and parameters. The ablation focuses on three modules, and the results in Table 8 demonstrate that the TDDPassNet model achieves the highest

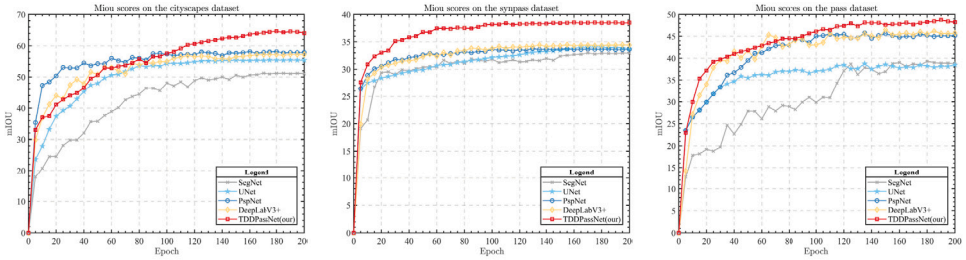


Fig. 12. Results of panoramic street images with metrics data with random sampling conducted in the city center area.

Table 7
Performance of TDDPassNet and Common Method Baseline Models on Different Datasets.

Dataset	Method	MAE[%]	RMSE[%]	mIoU[%]	pixAcc[%]	mAcc[%]	Predict[ms/img]
Cityscapes	SegNet	7.02	7.58	51.06	89.99	60.63	94
	UNet	6.73	6.42	55.37	91.05	67.03	83
	PspNet	5.54	5.62	57.89	91.87	68.87	76
	VPLR	4.37	4.42	57.57	92.64	72.06	69
	DeepLabV3+	4.66	4.78	57.21	92.75	70.3	72
	Ours	**2.39**	**2.63**	**64.63**	**93.92**	**74.45**	**48**
Synpass	SegNet	14.18	6.4	32.91	88.97	42.62	242
	UNet	13.21	6.01	33.36	90.73	44.44	186
	PspNet	11.41	5.24	33.75	91.34	45.96	168
	VPLR	11.73	3.85	34.26	92.08	46.83	169
	DeepLabV3+	11.06	3.28	34.44	92.37	46.39	172
	Ours	**8.65**	**2.17**	**38.52**	**95.53**	**49.16**	**96**
Pass	SegNet	21.58	10.07	38.69	89.65	51.03	58
	UNet	21.35	9.98	38.51	89.49	52.24	60
	PspNet	16.6	7.93	45.2	91.82	53.57	54
	VPLR	16.83	7.01	45.62	92.27	54.16	56
	DeepLabV3+	18.44	6.35	44.84	91.91	53.95	51
	Ours	**13.63**	**3.13**	**48.44**	**94.62**	**55.27**	**37**

segmentation IoU scores in 15 out of the 22 categories of segmentation benchmarks from SynPass. The model demonstrates a significant improvement of approximately 3.0% in extracting visual elements such as buildings, poles, roads, vegetation, vehicles, sky, and traffic lights. These elements pose challenges and are crucial for various fields of urban studies.

The first modification eliminates the multi-scale feature extraction module inserted in the middle of the Transformer structure. Only the second scale is used for feature extraction while retaining the deformable MLP module and dual attention. The decrease in the model's mIoU by 5.11% after the feature fusion is lost and only a single scale is utilized compared to TDDPassNet is demonstrated by Table 8. The multi-scale feature extraction and fusion module allows the model to access different layers of semantic perceptual information and spatial structural details in the panoramic image data.

Experiments are conducted by removing the deformable MLP module in each scale to investigate the effectiveness of element reprojection transformation processing for zigzag elements in panoramic images with feature re-learning. A significant decrease in mIoU compared to both TDDPassNet and the No_DA model, which eliminates the Dual Attention module, is demonstrated in Table 8. The decrease in performance may be attributed to the deformable MLP's flexibility in handling projection transformations. This prioritizes distortion awareness over fixing an offset when dealing with the panoramic image distortion problem.

Additionally, the impact of the Dual Attention mechanism on the model's performance is analyzed. The model without the Dual Attention module (No_DA) significantly underperforms TDDPassNet in terms of mIoU, as shown in Table 8. Although No_DA improves some visual elements, it decreases overall performance in 16 categories. This suggests that enhancing local semantics in panoramic data and exploring spatial-dimensional local correlation in multiscale spatial features is reasonable. Improving local semantics in feature maps creates local features with distinct spatial and channel dimensions. This aids in creating more representative semantic centers and reveals spatial-dimensional correlations in feature maps.

A series of parameter ablation experiments were conducted on the main dataset Synpass to determine the impact of all control parameters on the performance of TDDPassNet. The parameters that were adjusted during the experiment, including Epochs, Initial Learning Rate, and so forth, are presented in Table 9. The model's pixel accuracy (pixAcc) performance varies under different parameter settings. The bold numbers indicate the optimal pixel accuracy achieved on the test set, which is 95.53%.

4.3.5. Prediction results and qualitative analysis

The trained model is used for semantic segmentation prediction tasks. The unlabeled panorama dataset LJPass of the ancient city of Lijiang City area is used as the validation set of prediction. Multiple representative examples of qualitative results are presented in Fig. 13. In each case, the proposed TDDPassNet model significantly improves the segmentation boundaries of areas such as sky, buildings, and vegetation. In example (a), the model performs well in segmenting and predicting distorted city gates. In example (d), the model also outperforms the baseline in recognizing road surfaces affected by lighting. In example (b), the model has a significant advantage in segmenting road edges. The effect is also clearer for detailed segmentation of small objects, such as traffic lights and signs in example (e) and poles in all examples. These qualitative cases support the findings of our quantitative evaluation, demonstrating the effectiveness of the multi-scale feature extraction and embedding deformable MLP modules and Dual Attention mechanisms that we utilize.

To conduct a further qualitative analysis, we acquired the DenseP-ASS dataset (Gao et al., 2022). This dataset comprises 100 panoramic images and annotations for evaluation. When comparing labels, it becomes apparent that the baseline model struggles to differentiate distorted objects due to its inability to recognize global features and perceive distortion in panoramic images. When faced with distortion

Table 8
Results of model ablation experiment on SynPass dataset.

Method	mIoU	Building	Fence	Other	Pedestrian	Pole	RoadLine	Road	Side Walk	Vegetation	Vehicles	Wall	Traffic Sign	Sky	Ground	Bridge	Rail Track	Ground Rail	Traffic Light	Static	Dynamic	Water	Terrain
No_Multi-scale structure	33.41	77.27	25.12	2.14	25.64	23.64	50.42	83.21	47.15	63.04	45.75	3.33	11.75	92.93	8.2	4.35	12.58	53.96	19.41	22.18	8.67	5.46	48.9
No_DMLP	36.54	81.59	31.72	5.11	33.02	27.86	51.23	86.55	52.1	70.87	52.81	4.31	13.21	93.69	8.63	5.37	15.33	58.2	18.64	21.74	9.53	8.34	**54.15**
No_DA	37.55	78.52	**33.15**	6.06	35.16	27.51	**55.9**	87.08	**52.81**	70.22	53.36	5.34	12.73	93.45	10.07	6.69	**18.59**	62.66	17.9	**24.47**	10.64	**10.06**	53.79
TDDPassNet	**38.52**	**82.08**	32.48	**6.12**	**35.85**	**29.9**	54.64	**90.96**	52.36	**72.81**	**56.06**	**5.47**	**14.88**	**95.35**	**10.23**	**7.04**	17.76	**64.24**	**20.69**	23.66	**11.79**	9.13	54.03

Table 9
Parameter experiment on SynPass dataset. The altered parameters be indicated by the use of underlining.

Epochs	Initial learning rate	Learning rate schedule	Weight decay rate	pixAcc[%]
100/300	5e−5	Cosine Decay, Warm-up: 5 epochs	1e−4	89.35/95.26
200	1e−5/5e−8	Cosine Decay, Warm-up: 5 epochs	1e−4	92.67/94.82
200	5e−5	Linear Decay, No Warm-up/10 epochs	1e−4	91.92/93.95
200	5e−5	Cosine Decay, Warm-up: 5 epochs	1e−3/1e−6	94.31/94.59
200	5e−5	Cosine Decay, Warm-up: 5 epochs	1e−4	**95.53**

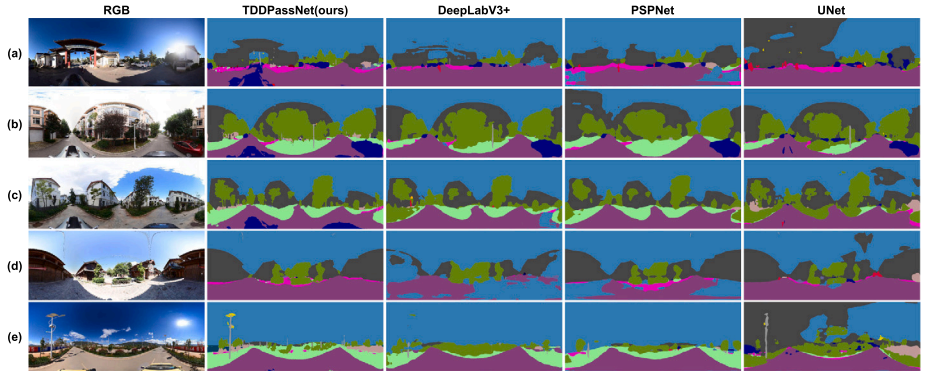


Fig. 13. Qualitative comparison of semantic segmentation prediction results between the TDDPassNet model and commonly used models in walkability research on the LJPass dataset. The figure presents several examples.

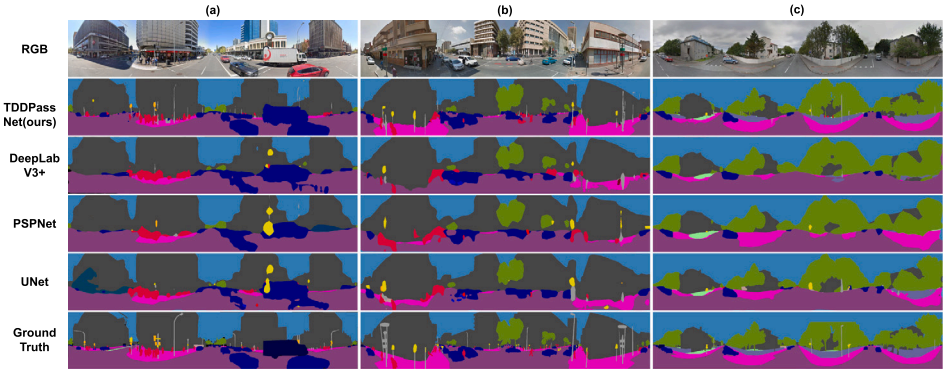


Fig. 14. Qualitative comparison between TDDPassNet model and commonly used models for walkability research on semantic partition prediction results under the densepass dataset with labeled data.

(as shown in Fig. 14), the baseline model struggles to accurately recognize challenging objects such as pedestrians, poles, and vehicles. However, our TDDPassNet can segment these objects more accurately without significant differences from the labeled image.

4.4. Analysis of results for domain adaptive methods

Panoramic images could provide a more comprehensive and distinct global context than traditional pinhole images. For instance, they allow for simultaneously observing roads and sidewalks in different directions. This feature enhances the quality of research ideas and model effects for urban researchers and facilitates the acquisition of sample data for training purposes. In the following experiment, we tested the benefits of the proposed unsupervised domain adaptive multi-stage prototype method. Our method produced high-quality sample data.

4.4.1. Feature embedding comparison of unsupervised domain adaptive multi-stage prototype methods

The t-SNE visualization of feature embedding during domain adaptation from pinhole images to panoramic images of urban scenes is shown in Fig. 15 (Van der Maaten and Hinton, 2008). Each dot represents all pixels sharing the same class in its image from the validation set of the respective domain. The source, intermediate, and target domain feature embeddings in the multi-stage outdoor domain adaptation process are shown in Fig. 15a, b, and c.

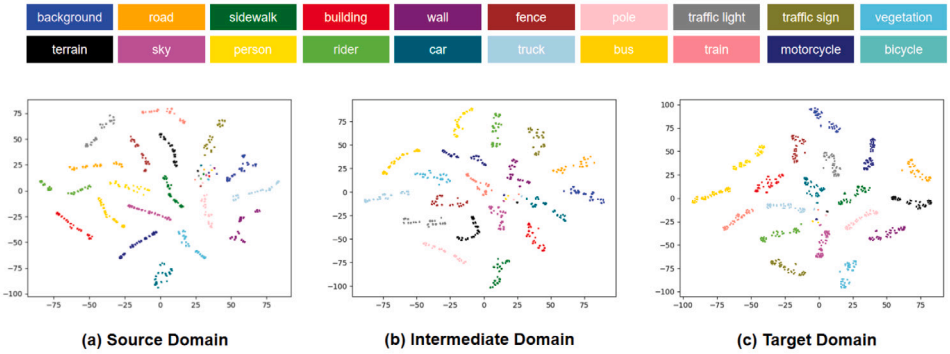


Fig. 15. t-SNE visualization of multi-stage domain adaptation process in outdoor scene.

By comparing Fig. 15a, b, and c, it is evident that the multi-stage adaptation process of pinhole-synthetic panorama-real panorama results in a clear tendency of clustering of the feature space as it passes through the intermediate domain. The initial source domain before domain adaptation, where the feature space distribution is relatively loose, is shown in Fig. 15a. After adopting the final target domain, the distribution of feature space categories is balanced, and there is a good pooling phenomenon for the same categories. This suggests that the proposed multi-stage prototype approach bridges the domain gap in different feature spaces and provides complementary feature alignment between domains, making their features more closely related to their prototypes.

4.4.2. Phased split ablation experiment and generate qualitative analysis of label quality

An ablation experiment was conducted to verify the multi-stage advantage of the intermediate domain prototype in our proposed unsupervised domain adaptation method. The experiment used Synpass as the intermediate domain and was divided into two parts: single-stage and multi-stage. The intermediate domain procedure was removed in the single-stage, but it was reserved in the multi-stage. Then, these two methods were utilized to generate the sample label image of the Ljpass dataset, which was then used as the validation set, and several baseline models trained in the same environment were employed for corresponding validation.

The mIoU effect of the verification set using tags generated by the domain adaptive strategy at different stages is displayed in Table 10, highlighting the gap between them. The results indicate that removing the intermediate domain significantly reduces the recognition ability of each model for the sample, with a mIoU gap of more than 2.5% for each model. This demonstrates the benefits of the intermediate domain added during the pinhole to panoramic domain adaptation process. It assists the model in learning the relationship between the source and target domains, resulting in improved fitting of the pseudo tag. Using Synpass, a composite panoramic image dataset with rich sample labels, as an intermediate domain, the model can gradually learn the conversion process from a pinhole image to a panoramic image. This allows for a better understanding of the feature differences between different domains and generates higher-quality sample label images. This method of gradual adaptation in multiple stages enhances the model's ability to accurately classify and label samples from both the source and target domains, thereby improving the model's performance and generalization in the target domain.

The proposed unsupervised domain-adaptive multistage prototype method for urban scenes is used to annotate the LJPass dataset obtained from the study area. Examples of annotations generated by our method are shown in Fig. 16. Panoramic image data without panoramic sampled cars are represented by a and b in Fig. 16, while data in the presence of panoramic sampled cars are represented by b and c. The top part is the original image, and the bottom part is the labeled image of

Table 10
Validation Results of Self-labeled LJPass Datasets Generated by Baseline Model in Single Stage and Multi-Stage.

Method	Stage	mIoU[%]	gap[%]
SegNet	Single	24.93	3.54
	Multi	28.47	
UNet	Single	31.27	2.59
	Multi	33.86	
PspNet	Single	31.61	2.74
	Multi	34.35	
VPLR	Single	31.94	2.79
	Multi	34.73	
DeepLabV3+	Single	32.14	2.81
	Multi	34.95	

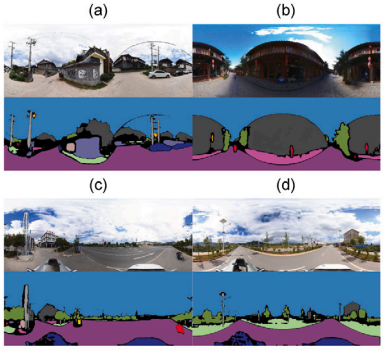


Fig. 16. Example of automatically generated annotation of panoramic image data using multi-stage domain adaptation.

the 8-bit color map we generated, with a category distribution of 0–18, where the black part is the background. It can be seen that although the generated tags are not as perfect as manual annotation, they are relatively accurate, and the definitions of different classes are relatively good.

4.5. Correlation analysis of street view visual elements and four-dimensional perceived walkability

To evaluate the impact of different visual elements on perceived walkability, the Pearson linear regression model and the parameters of the random forest model are used to compare the visual elements with the results of four perceived walkability scores. The ranking results of the independent variables with significant positive (blue bar) or negative (red bar) effects for each perceived score category are shown in Fig. 17. We constructed four regression models corresponding to four perceptual rating dimensions (safety, convention, comfort, and

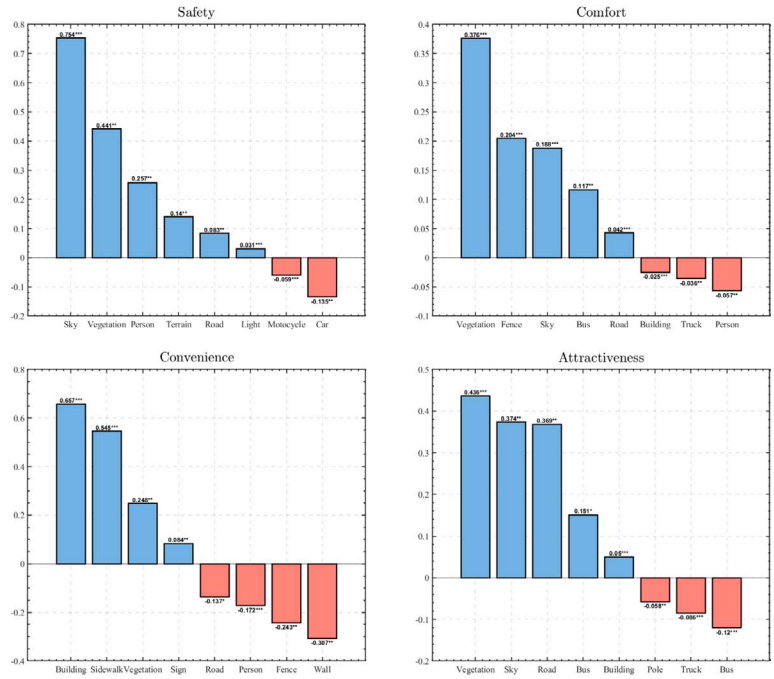


Fig. 17. The influence of visual elements on the 4-dimensional perceived walkability index is calculated step by step using Pearson's multiple linear regression model and shows the visual variables with significant influence. (Beta coefficient: * $p < 0.05$ * * $p < 0.01$ * * * $p < 0.001$).

attractiveness). Through stepwise regression, influence variables were added or excluded based on their impact on perceived walkability, retaining only the visual elements that had a significant effect on the explanation.

Based on the results in Fig. 17, the influence of visual variables such as sky, vegetation, and road on safety, comfort, and overall attractiveness is evident.Visual variables such as building, pavement, and sign have a highly sensitive positive effect on comfort. These results are consistent with previous studies that urban greenery and open views create a sense of safety, as well as affect pedestrian comfort and increase the attractiveness of streets (Li et al., 2022b). Meanwhile, spacious pedestrian spaces increase the walkability of streets, making pedestrians more comfortable and willing to travel. Convenient visual amenities also provide pedestrians with a better walking experience.

It is worth noting that the road visual element variable has a weak but negative effect on the convention and building element variables on comfort. This is consistent with previous research (Gao et al., 2022), that is, wide roads tend to have large traffic and pedestrian flows, which has a certain impact on pedestrian comfort, while the dense and high-rise building environment may put more psychological pressure on pedestrians. It should also be noted that motorcycles, lorries, cars, and buses associated with vehicles have a strong negative impact on various perceptions. This is consistent with real or perceived obstacles preventing pedestrians from walking on the street.

5. Discussion

A comprehensive analysis of walkability in Lijiang City, China, is presented with a three-dimensional urban walkability research framework from an urban planning perspective. The framework utilizes a combination of multi-form and multi-modal open data, as well as multiple machine learning methods, to extract the intrinsic environmental characteristics of the city at different scales and with different dimensions of the indicator situation to provide a comprehensive assessment of the city's walkability.

The advantages of this framework are multifaceted. Firstly, it comprehensively assesses street ecosystem characteristics, physical walkability, and perceived walkability using PSVI big data. By combining various objective and subjective factors into a single metric, a more nuanced and inclusive walkability assessment is provided by the framework, avoiding the pitfalls of one-sided evaluations based on limited data sources or singular dimensions. Secondly, the framework employs human–machine adversarial techniques, bridging the gap between purely machine-based audits and on-site assessments. More consistent results are achieved by this approach, and the relationships between different perceptual dimensions and built environment characteristics are effectively explained through correlation analyses. Such insights could assist urban planners and designers make more informed decisions and adjustments.

Exploring the factors affecting accessibility in different regions through the CWI reveals potential priority areas for urban street development. The study found that the street ecological index distribution in the ancient city district is broad and balanced, indicating high overall spatial quality. Physical walkability is high in the central area, with other good areas scattered around the central ring. Perceived walkability is prominent in the business district near the city's main road. However, traffic congestion on the main road lowers scores for attractiveness and comfort, suggesting the need for green elements to alleviate these issues to be considered. According to the spatial distribution map of CWI, the high-score regions are concentrated in the middle and south, while the low-score regions are distributed in the adjacent north and east regions. This suggests that the overall accessibility of roads in the ancient city of Lijiang City area is relatively good, but the distribution is not uniform enough. Decision-makers and urban planners should consider intervening in the built environment to improve the city's walkability.

While valuable insights into urban street walkability are provided by this study, limitations and areas for future research are identified. The small sample size and limited model scope mean that our semantic segmentation model for panoramic images, based on PSVIs, needs further optimization to better mitigate distortion effects. Enhancing the model structure in computer vision could provide more comprehensive scene perception for ultra-widefield panoramic images. Additionally, we did not account for temporal variations in urban road quality. The impact of street ecological quality on pedestrian walkability across different time dimensions could be studied by future research utilizing

seasonal PSVIs from crowdsourced maps. Furthermore, our human–machine adversarial evaluation did not fully consider different age groups and life stages. In future studies, walking-related variables should be calibrated to meet the diverse needs of various pedestrians, such as children and older adults. Finally, our research was limited to a single city, and no relocation experiments were conducted. Future experiments in diverse urban environments are necessary to validate and generalize our methodology.

6. Conclusion

This paper presents a comprehensive framework for urban walking ability, which calculates the Comprehensive Walking Ability Index (CWI) by integrating ecological, physical, and perceptual walking ability dimensions. The method is then verified through a case study of Lijiang City, China. Of particular note is the novel consideration of two major issues about panoramic images within the realm of urban research based on PSVIs: image distortion and lack of annotated training data. Addressing these challenges, a novel model, TDDPassNet, has been proposed to tackle these challenges by enhancing semantic segmentation of panoramic images, thereby bridging existing research gaps. Moreover, an unsupervised domain-adaptive multi-stage prototype method for urban scenes is introduced to partially alleviate limitations in sample size within target urban research areas, thereby providing valuable dataset support for such endeavors. Compared to previous methodologies, the proposed TDDPassNet demonstrated an average improvement of 5.3% in mIoU across all datasets. The incorporation of innovative techniques, such as deformable MLP modules and dual attention mechanisms, has enhanced the model's ability to perceive distortion and detect transformations, thereby improving the accuracy and credibility of street walkability assessment. Although traditional field investigation methods are known for their accuracy, they are also labor-intensive. In contrast, we have implemented large-scale scene analysis and complex computing based on a deep learning framework, which is suitable for a wide range of data processing, improves efficiency and scalability, and meets the requirements of modern urban research.

Looking ahead, our method has potential applications in other fields, such as real-time urban monitoring, disaster response, and smart city development. Future research could explore integrating this framework with other urban planning tools to create more dynamic and adaptive urban environments. By addressing the identified limitations and expanding the scope of the research, the utility and applicability of our proposed walkability assessment framework could be further enhanced.

CRediT authorship contribution statement

Jiaxuan Li: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Xuan Zhang:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Funding acquisition, Data curation, Conceptualization. **Linyu Li:** Writing – review & editing, Validation, Project administration, Investigation, Formal analysis. **Xu Wang:** Writing – review & editing, Supervision. **Jing Cheng:** Project administration, Funding acquisition. **Chen Gao:** Writing – review & editing, Visualization, Software, Project administration. **Jun Ling:** Visualization, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

Science Foundation of Young and Middle-aged Academic and Technical Leaders of Yunnan under Grant No. 202205AC160040; Science Foundation of Yunnan Jinzhi Expert Workstation under Grant No. 202205AF150006; Major Project of Yunnan Natural Science Foundation under Grant No. 202302AE09002003; Knowledge-driven Smart Energy Science and Technology Innovation Team of Yunnan Provincial Department of Education; Open Foundation of Yunnan Key Laboratory of Software Engineering under Grant No. 2023SE101; Yunnan University Graduate Research Innovation Foundation Project under Grant No. KC-23233953.

References

Anthony, Jr., Bokolo, 2023. The role of community engagement in urban innovation towards the co-creation of smart sustainable cities. J. Knowl. Econ. 1–33.

Arellana, Julian, Saltarín, María, Larrañaga, Ana Margarita, Alvarez, Vilma, Henao, César Augusto, 2020. Urban walkability considering pedestrians' perceptions of the built environment: a 10-year review and a case study in a medium-sized city in latin america. Transp. Rev. 40 (2), 183–203.

Badrinarayanan, Vijay, Kendall, Alex, Cipolla, Roberto, 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39 (12), 2481–2495.

Bartzokas-Tsiompras, Alexandros, Bakogiannis, Efthimios, Nikitas, Alexandros, 2023. Global microscale walkability ratings and rankings: A novel composite indicator for 59 European city centres. J. Transp. Geogr. 111, 103645.

Biljecki, Filip, Ito, Koichi, 2021. Street view imagery in urban analytics and GIS: A review. Landsc. Urban Plan. 215, 104217.

Carr, Lucas J., Dunsiger, Shira I., Marcus, Bess H., 2011. Validation of walk score for estimating access to walkable amenities. Br. J. Sports Med. 45 (14), 1144–1148.

Cerin, Ester, Leslie, Eva, du Toit, Lorinne, Owen, Neville, Frank, Lawrence D., 2007. Destinations that matter: associations with walking for transport. Health Place 13 (3), 713–724.

Chang, Wei-Lun, Wang, Hui-Po, Peng, Wen-Hsiao, Chiu, Wei-Chen, 2019. All about structure: Adapting structural information across domains for boosting semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1900–1909.

Chen, Liang-Chieh, Zhu, Yukun, Papandreou, George, Schroff, Florian, Adam, Hartwig, 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European Conference on Computer Vision. ECCV, pp. 801–818.

Cordts, Marius, Omran, Mohamed, Ramos, Sebastian, Rehfeld, Timo, Enzweiler, Markus, Benenson, Rodrigo, Franke, Uwe, Roth, Stefan, Schiele, Bernt, 2016. The cityscapes dataset for semantic urban scene understanding. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3213–3223.

Ewing, Reid, Handy, Susan, 2009. Measuring the unmeasurable: Urban design qualities related to walkability. J. Urban Design 14 (1), 65–84.

Ewing, Reid, Handy, Susan, Brownson, Ross C., Clemente, Otto, Winston, Emily, 2006. Identifying and measuring urban design qualities related to walkability. J. Physical Activity Health 3 (s1), S223–S240.

Farahani, Mahsa, Razavi-Termeh, Seyed Vahid, Sadeghi-Niaraki, Abolghasem, Choi, Soo-Mi, 2023. A hybridization of spatial modeling and deep learning for People's visual perception of urban landscapes. Sustainability 15 (13), 10403.

Gallego, Antonio-Javier, Calvo-Zaragoza, Jorge, Fisher, Robert B., 2020. Incremental unsupervised domain-adversarial training of neural networks. IEEE Trans. Neural Netw. Learn. Syst. 32 (11), 4864–4878.

Gao, Ping, Liu, Shenghe, Qi, Wei, Qi, Honggang, 2020. The nexus between poverty and the environment: A case study of lijiang, China. Sustainability 12 (3), 1066.

Gao, Wenxiu, Qian, Yuting, Chen, Hanzhe, Zhong, Zhenqian, Zhou, Min, Aminpour, Fatemeh, 2022. Assessment of sidewalk walkability: Integrating objective and subjective measures of identical context-based sidewalk features. Sustainable Cities Soc. 87, 104142.

Gebel, Klaus, Bauman, Adrian, Owen, Neville, 2009. Correlates of non-concordance between perceived and objective measures of walkability. Ann. behav. med. 37 (2), 228–238.

Gong, Peng, Chen, Bin, Li, Xuecao, Liu, Han, Wang, Jie, Bai, Yuqi, Chen, Jingming, Chen, Xi, Fang, Lei, Feng, Shuailong, et al., 2020. Mapping essential urban land use categories in China (EULUC-China): Preliminary results for 2018. Sci. Bull. 65 (3), 182–187.

Gong, Fang-Ying, Zeng, Zhao-Cheng, Zhang, Fan, Li, Xiaojiang, Ng, Edward, Norford, Leslie K., 2018. Mapping sky, tree, and building view factors of street canyons in a high-density urban environment. Build. Environ. 134, 155–167.

Goodfellow, Ian, Pouget-Abadie, Jean, Mirza, Mehdi, Xu, Bing, Warde-Farley, David, Ozair, Sherjil, Courville, Aaron, Bengio, Yoshua, 2014. Generative adversarial nets. Adv. Neural Inf. Process. Syst. 27.

Guzman, Luis A., Arellana, Julian, Castro, William Felipe, 2022. Desirable streets for pedestrians: Using a street-level index to assess walkability. Transp. Res. D: Transp. Environ. 111, 103462.

He, Xuan, He, Sylvia Y., 2023. Using open data and deep learning to explore walkability in shenzhen, China. Transp. Res. D 118, 103696.

Horak, Jiri, Kukuliac, Pavel, Maresova, Petra, Orlikova, Lucie, Kolodziej, Ondrej, 2022. Spatial pattern of the walkability index, walk score and walk score modification for elderly. ISPRS Int. J. Geo-Inf. 11 (5), 279.

Hua, Junyi, Cai, Meng, Shi, Yuan, Ren, Chao, Xie, Jing, Chung, Lamuel Chi Hay, Lu, Yi, Chen, Long, Yu, Zhaowu, Webster, Chris, 2022. Investigating pedestrian-level greenery in urban forms in a high-density city for urban planning. Sustainable Cities Soc. 80, 103755.

Huo, Xinyue, Xie, Lingxi, Hu, Hengtong, Zhou, Wengang, Li, Houqiang, Tian, Qi, 2022. Domain-agnostic prior for transfer semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7075–7085.

Jamei, Elmira, Ahmadi, Khatereh, Chau, Hing Wah, Seyedmahmoudian, Mehdi, Horan, Ben, Stojcevski, Alex, 2021. Urban design and walkability: Lessons learnt from Iranian traditional cities. Sustainability 13 (10), 5731.

Jiang, Yuxiao, Chen, Long, Grekousis, George, Xiao, Yang, Ye, Yu, Lu, Yi, 2021. Spatial disparity of individual and collective walking behaviors: A new theoretical framework. Transp. Res. D: Transp. Environ. 101, 103096.

Kang, Youngok, Kim, Jiyeon, Park, Jiyoung, Lee, Jiyoon, 2023. Assessment of perceived and physical walkability using street view images and deep learning technology. ISPRS Int. J. Geo-Inf. 12 (5), 186.

Ki, Donghwan, Lee, Sugie, 2021. Analyzing the effects of green view index of neighborhood streets on walking time using google street view and deep learning. Landsc. Urban Plan. 205, 103920.

Kim, Eun Jung, Jin, Suin, 2023. Walk score and neighborhood walkability: A case study of daegu, South Korea. Int. J. Environ. Res. Public Health 20 (5), 4246.

Kim, Sohee, Woo, Ayoung, 2022. Streetscape and business survival: Examining the impact of walkable environments on the survival of restaurant businesses in commercial areas based on street view images. J. Transp. Geogr. 105, 103480.

Kim, Eun-Sub, Yun, Seok-Hwan, Park, Chae-Yeon, Heo, Han-Kyul, Lee, Dong-Kun, 2022. Estimation of mean radiant temperature in urban canyons using google street view: A case study on seoul. Remote Sens. 14 (2), 260.

Koohsari, Mohammad Javad, McCormack, Gavin R., Shibata, Ai, Ishii, Kaori, Yasunaga, Akitomo, Nakaya, Tomoki, Oka, Koichiro, 2021. The relationship between walk score® and perceived walkability in ultrahigh density areas. Prevent. Med. Rep. 23, 101393.

Kouw, Wouter M., Loog, Marco, 2019. A review of domain adaptation without target labels. IEEE Trans. Pattern Anal. Mach. Intell. 43 (3), 766–785.

Lai, Wei-Sheng, Huang, Yujia, Joshi, Neel, Buehler, Christopher, Yang, Ming-Hsuan, Kang, Sing Bing, 2017. Semantic-driven generation of hyperlapse from 360 degree video. IEEE Trans. Visual. Comput. Graphics 24 (9), 2610–2621.

Li, Ruihuang, Li, Shuai, He, Chenhang, Zhang, Yabin, Jia, Xu, Zhang, Lei, 2022a. Class-balanced pixel-level self-labeling for domain adaptive semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11593–11603.

Li, Yunqin, Yabuki, Nobuyoshi, Fukuda, Tomohiro, 2022b. Measuring visual walkability perception using panoramic street view images, virtual reality, and deep learning. Sustainable Cities Soc. 86, 104140.

Li, Yunqin, Yabuki, Nobuyoshi, Fukuda, Tomohiro, 2023a. Integrating GIS, deep learning, and environmental sensors for multicriteria evaluation of urban street walkability. Landsc. Urban Plan. 230, 104603.

Li, Fei, Yigitcanlar, Tan, Nepal, Madhav, Nguyen, Kien, Dur, Fatih, 2023b. Machine learning and remote sensing integration for leveraging urban sustainability: A review and framework. Sustainable Cities Soc. 104653.

Liu, Pengyuan, Biljecki, Filip, 2022. A review of spatially-explicit geoai applications in urban geography. Int. J. Appl. Earth Obs. Geoinf. 112, 102936.

Long, Jonathan, Shelhamer, Evan, Darrell, Trevor, 2015. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3431–3440.

Loshchilov, Ilya, Hutter, Frank, 2017. Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101.

Mohanty, Sudatta, Pozdnukhov, Alexey, Cassidy, Michael, 2020. Region-wide congestion prediction and control using deep learning. Transp. Res. C 116, 102624.

Nagata, Shohei, Nakaya, Tomoki, Hanibuchi, Tomoya, Amagasa, Shiho, Kikuchi, Hiroyuki, Inoue, Shigeru, 2020. Objective scoring of streetscape walkability related to leisure walking: Statistical modeling approach with semantic segmentation of google street view images. Health Place 66, 102428.

Ogawa, Yoshiki, Zhao, Chenbo, Oki, Takuya, Chen, Shenglong, Sekimoto, Yoshihide, 2023. Deep learning approach for classifying the built year and structure of individual buildings by automatically linking street view images and gis building data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 16, 1740–1755.

Oza, Poojan, Sindagi, Vishwanath A., Sharmini, Vibashan Vishnukumar, Patel, Vishal M., 2023. Unsupervised domain adaptation of object detectors: A survey. IEEE Trans. Pattern Anal. Mach. Intell..

Peiravian, Farideddin, Derrible, Sybil, Ijaz, Farukh, 2014. Development and application of the pedestrian environment index (PEI). J. Transp. Geogr. 39, 73–84.

Quercia, Daniele, Schifanella, Rossano, Aiello, Luca Maria, 2014. The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. In: Proceedings of the 25th ACM Conference on Hypertext and Social Media. pp. 116–125.

Ronneberger, Olaf, Fischer, Philipp, Brox, Thomas, 2015. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer, pp. 234–241.

Saaty, Thomas L., 1988. What is the analytic hierarchy process? In: Mathematical Models for Decision Support. Springer, pp. 109–121.

Scorza, Francesco, Fortunato, Giovanni, Carbone, Raffaella, Murgante, Beniamino, Pontrandolfi, Piergiuseppe, 2021. Increasing urban walkability through citizens' participation processes. Sustainability 13 (11), 5835.

Suel, Esra, Bhatt, Samir, Brauer, Michael, Flaxman, Seth, Ezzati, Majid, 2021. Multimodal deep learning from satellite and street-level imagery for measuring income, overcrowding, and environmental deprivation in urban areas. Remote Sens. Environ. 257, 112339.

Sun, Cheng, Sun, Min, Chen, Hwann-Tzong, 2021a. Hohonet: 360 indoor holistic understanding with latent horizontal features. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2573–2582.

Sun, Yi, Wang, Xingzhi, Zhu, Jiayin, Chen, Liangjian, Jia, Yuhang, Lawrence, Jean M., Jiang, Luo-hua, Xie, Xiaohui, Wu, Jun, 2021b. Using machine learning to examine street green space types at a high spatial resolution: Application in los angeles county on socioeconomic disparities in exposure. Sci. Total Environ. 787, 147653.

Tang, Jingxian, Long, Ying, 2019. Measuring visual quality of street space and its temporal variation: Methodology and its application in the hutong area in Beijing. Landsc. Urban Plan. 191, 103436.

Tang, Yiwen, Zhang, Jiaxin, Liu, Runjiao, Li, Yunqin, 2022. Exploring the impact of built environment attributes on social followings using social media data and deep learning. ISPRS Int. J. Geo-Inf. 11 (6), 325.

Tsiompras, Alexandros Bartzokas, Photis, Yorgos N., 2017. What matters when it comes to "walk and the city"? Defining a weighted GIS-based walkability index. Transp. Res. Procedia 24, 523–530.

Van der Maaten, Laurens, Hinton, Geoffrey, 2008. Visualizing data using t-SNE. J. Mach. Learn. Res. 9 (11).

Vaswani, Ashish, Shazeer, Noam, Parmar, Niki, Uszkoreit, Jakob, Jones, Llion, Gomez, Aidan N., Kaiser, Łukasz, Polosukhin, Illia, 2017. Attention is all you need. Adv. Neural Inf. Process. Syst. 30.

Wang, Chaofeng, Antos, Sarah Elizabeth, Triveno, Luis Miguel, 2021. Automatic detection of unreinforced masonry buildings from street view images using deep learning-based image segmentation. Autom. Constr. 132, 103968.

Wang, Jing, Biljecki, Filip, 2022. Unsupervised machine learning in urban studies: A systematic review of applications. Cities 129, 103925.

Woo, Sanghyun, Park, Jongchan, Lee, Joon-Young, Kweon, In So, 2018. Cbam: Convolutional block attention module. In: Proceedings of the European Conference on Computer Vision. ECCV, pp. 3–19.

Yang, Kailun, Hu, Xinxin, Bergasa, Luis M., Romera, Eduardo, Wang, Kaiwei, 2019. Pass: Panoramic annular semantic segmentation. IEEE Trans. Intell. Transp. Syst. 21 (10), 4171–4185.

Yang, Kailun, Hu, Xinxin, Stiefelhagen, Rainer, 2021a. Is context-aware CNN ready for the surroundings? Panoramic semantic segmentation in the wild. IEEE Trans. Image Process. 30, 1866–1881.

Yang, Kailun, Zhang, Jiaming, Reiß, Simon, Hu, Xinxin, Stiefelhagen, Rainer, 2021b. Capturing omni-range context for omnidirectional segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1376–1386.

Zhang, Yang, David, Philip, Foroosh, Hassan, Gong, Boqing, 2019. A curriculum domain adaptation approach to the semantic segmentation of urban scenes. IEEE Trans. Pattern Anal. Mach. Intell. 42 (8), 1823–1841.

Zhang, Jiaming, Yang, Kailun, Shi, Hao, Reiß, Simon, Peng, Kunyu, Ma, Chaoxiang, Fu, Haodong, Torr, Philip HS, Wang, Kaiwei, Stiefelhagen, Rainer, 2022. Behind every domain there is a shift: Adapting distortion-aware vision transformers for panoramic semantic segmentation. arXiv preprint arXiv:2207.11860.

Zhao, Chenbo, Ogawa, Yoshiki, Chen, Shenglong, Oki, Takuya, Sekimoto, Yoshihide, 2023. Quantitative land price analysis via computer vision from street view images. Eng. Appl. Artif. Intell. 123, 106294.

Zhao, Hengshuang, Shi, Jianping, Qi, Xiaojuan, Wang, Xiaogang, Jia, Jiaya, 2017. Pyramid scene parsing network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2881–2890.

Zhou, Hao, He, Shenjing, Cai, Yuyang, Wang, Miao, Su, Shiliang, 2019. Social inequalities in neighborhood visual walkability: Using street view imagery and deep learning technologies to facilitate healthy city planning. Sustain. Cities Soc. 50, 101605.

Zhu, Yi, Sapra, Karan, Reda, Fitsum A., Shih, Kevin J., Newsam, Shawn, Tao, Andrew, Catanzaro, Bryan, 2019. Improving semantic segmentation via video propagation and label relaxation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8856–8865.