Exploring the association between street built environment and street vitality using deep learning methods

Article *in* Sustainable Cities and Society · December 2021

DOI: 10.1016/j.scs.2021.103656

CITATIONS
2

READS
294

3 authors, including:

Yunqin Li
Osaka University
6 PUBLICATIONS **7** CITATIONS
SEE PROFILE

Nobuyoshi Yabuki
Osaka University
247 PUBLICATIONS **1,241** CITATIONS
SEE PROFILE

Sustainable Cities and Society xxx (xxxx) 103656

Contents lists available at ScienceDirect

Sustainable Cities and Society

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

Exploring the association between street built environment and street vitality using deep learning methods

Yunqin Li, Nobuyoshi Yabuki *, Tomohiro Fukuda

Division of Sustainable Energy and Environmental Engineering, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan

ARTICLE INFO

Keywords:
Street vitality
Built environment
Pedestrian behavior and preference
Scene classification
Semantic segmentation
Multiple object tracking

ABSTRACT

Street vitality has become an essential indicator for evaluating the attractiveness and potential of the sustainable development of urban blocks, and it can be reflected by the type and the frequency of people's pedestrian activities on the street. While it is recognized that street built environment features affect pedestrian behavior and street vitality, quantifying the impact of these characteristics remains inconclusive. This paper proposes an automated deep learning approach to quantitatively explore the association between the street built environment and street vitality. First, we established a deep learning model for street vitality classification for automatic evaluation of street vitality based on the volumes and activities of pedestrians in the street through multiple object tracking and scene classification. Then, we applied semantic segmentation to measure five selected vitality-related street built environment variables. Finally, a linear regression model was applied to evaluate the built environment variables' significance and effects on street vitality. To verify our method's accuracy and applicability, we selected a commercial complex in Osaka as an illustrative example. The experimental results highlight that street width and transparency have significant positive effects on street vitality. Compared with traditional methods, our approach is feasible, reliable, transferable, and more efficient.

1. Introduction

Since the 1960s, with a rethinking of modern urban planning dominated by functional zoning, there has been a growing emphasis on diverse and vibrant urban spaces (Zeng et al., 2018). Vibrant cities and streets enhance the happiness of residents and social cohesion (Mouratidis & Poortinga, 2020; Zhang et al., 2021a). Urban street vitality has become an important indicator for assessing the attractiveness and potential of urban neighborhoods for sustainable development (Jacobs, 1961; Maas, 1984). However, with the development of urban traffic and the impact of virtual space, the social function of the street is gradually encroached by the traffic function, pedestrians are marginalized or introduced into the indoor environment, and the street space faces the crisis of gradually losing vitality or even decay (Zeng et al., 2018). As the physical and social space of human activities, the built environment of urban streets has close connections with urban vitality (Yue et al., 2021). Many design theories have been proposed to create a vibrant urban street space and a growing body of methodological and empirical research has been conducted on the role of relevant physical environment variables, such as street interface continuity, greening ratio, ground floor interface transparency, and commercial density, in fostering street activities and vitality (Buchanan, 1988; Gehl & Svarre, 2013; Lopes & Camanho, 2013; Montgomery, 1998; Sung & Lee, 2015; Zeng et al., 2019; Zhang et al., 2021a).

For a long time, there have been two perspectives on interpreting the concept of urban vitality: urban sociology and architecture. Urban sociology generally holds that economic, social, and cultural vitality are all intertwined in urban vitality, and urban spatial vitality is the spatial representation of economic, social, and cultural activities (Xia et al., 2020). In contrast, architects believe that urban space vitality can be understood as a kind of urban activity based on the form of urban space and can be created through design approaches (He et al., 2019; Marcus, 2010). The street built environment itself cannot form vitality directly but provides a place for hosting and influencing people's activities. In other words, urban vitality is an isomorphism of spatial characteristics and the social activities behind them, and spatial morphological characteristics could affect the intensity and complexity of pedestrians' activities, especially the intensity of optional activities defined by Gehl (1987). In recent years, research on urban spatial vitality has gradually become joint. More scholars believe that urban spatial vitality can be understood as a kind of urban activity based on urban spatial form (Lees, 2010). In this paper, street vitality concerns mainly its social vi-

* Corresponding author.
E-mail address: yabuki@see.eng.osaka-u.ac.jp (N. Yabuki).

https://doi.org/10.1016/j.scs.2021.103656
Received 15 August 2021; Received in revised form 29 December 2021; Accepted 29 December 2021

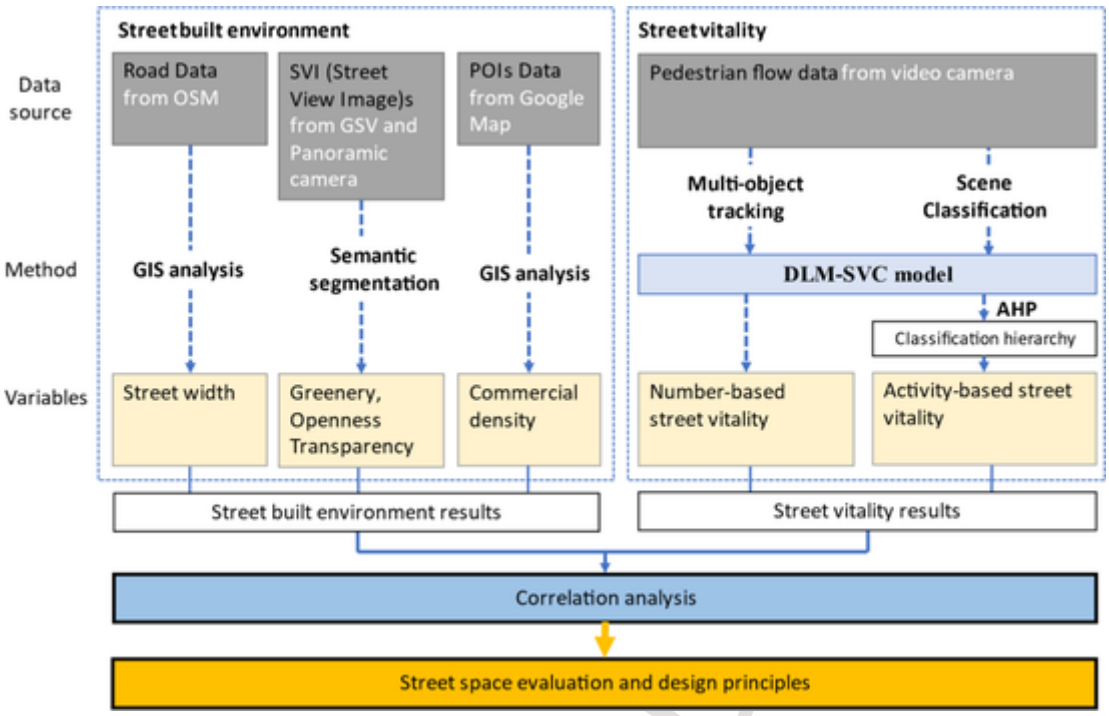

Fig. 1. The framework of the study.

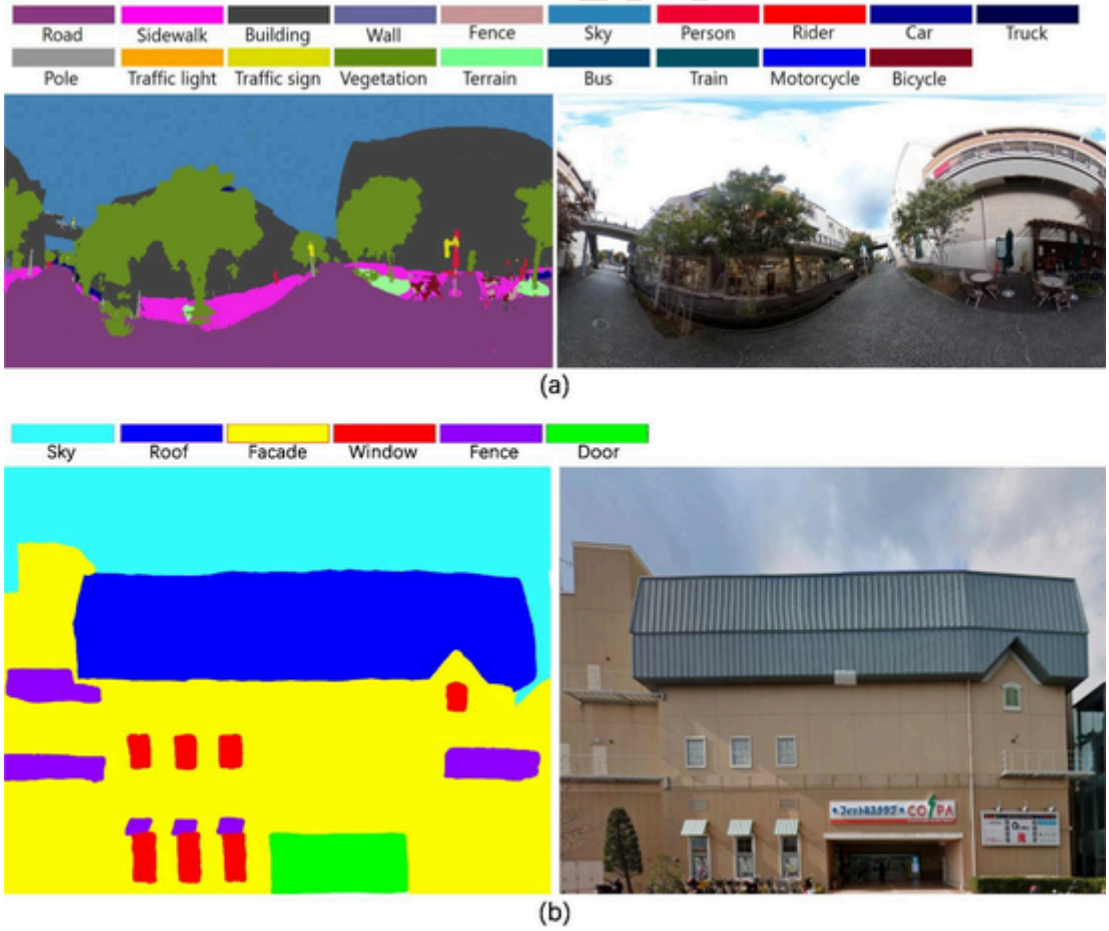

Fig. 2. Segmentation testing samples for (a) DeepLabv3 + and (b) WEEK 3-Facade parsing.

tality and is related to street spatial characteristics. It can be measured by the intensity and complexity of activities reflected in the type and frequency of activities people do on the street.

Previous works on street vitality and physical indicators of the street built environment are usually qualitative and descriptive studies, including questionnaire surveys, field surveys, and cognitive maps (Sung & Lee, 2015), or macro-scale studies based on geographic, sociological, and statistical methods (Zeng et al., 2018). Data collection and analysis on the characteristics of the built environment and the pedestrian activity level are time-consuming and labor-intensive. For example, Gehl and Svarre (2013) public space public life (PSPL) survey, one of the typical subjective perception and observation assessment methods based on field investigation, was limited by observation time and the number of observers. Moreover, most macro-scale studies have addressed only a single environmental variable for its impact on street vitality, and it is also difficult to determine which of the multiple variables are more important in promoting spatial vitality (Xia et al., 2020; Zhang et al., 2021a).

Recent advances in deep learning and big data technologies have brought about a paradigm shift in street vitality and physical indicators of the street built environment (Dong et al., 2021; Sulis et al., 2018). The increasing popularity of urban geographical information system (GIS) data and point of interest (POI) data facilitates the acquisition of basic information on urban streets and building functions, providing better conditions for large-scale measurement of the built environment. Owing to the implementation of deep learning algorithms for image processing, the street physical environment captured by remote-sensing images and street view images can now be further quantified (Li et al., 2020). Gong et al. (2018) and Yin and Wang (2016) validated the feasibility of using semantic segmentation and Google Street View images in measuring sky, tree, building, and other landscape features in the streetscape. Hu et al. (2020) proposed an image scene classification method with a multitask deep learning model and street view images to classify urban canyons accurately.

Additionally, the development of deep convolutional neural networks (CNNs) has further stimulated the interest in image content analysis. Image-sensing data allow for the measurement and visualization of people's behaviors, movements, and preferences with low cost but high accuracy. A wide range of computer vision applications in the automated analysis of video or image data has proven useful in macro-scale studies (Quintanar et al., 2021; Zhang et al., 2021b). It is widely accepted that in micro-scale studies, there needs to be a more comprehensive understanding of how pedestrians negotiate space (Dziedzic et al., 2019). For example, Hou et al. (2020) used video-based surveillance to record people's activity data, counting and gridding people in a common area to provide a new way to track and quantify how small public spaces are used. Angah and Chen (2020) analyzed construction worker activities using image sensing data through a multiple-object tracking (MOT) model. Liang et al. (2020) visualized pedestrian trajectories using video recordings and analyzed how weather and climate affect pedestrian walking speed. However, these initial explorations concentrated more on quantitatively evaluating people's behaviors,

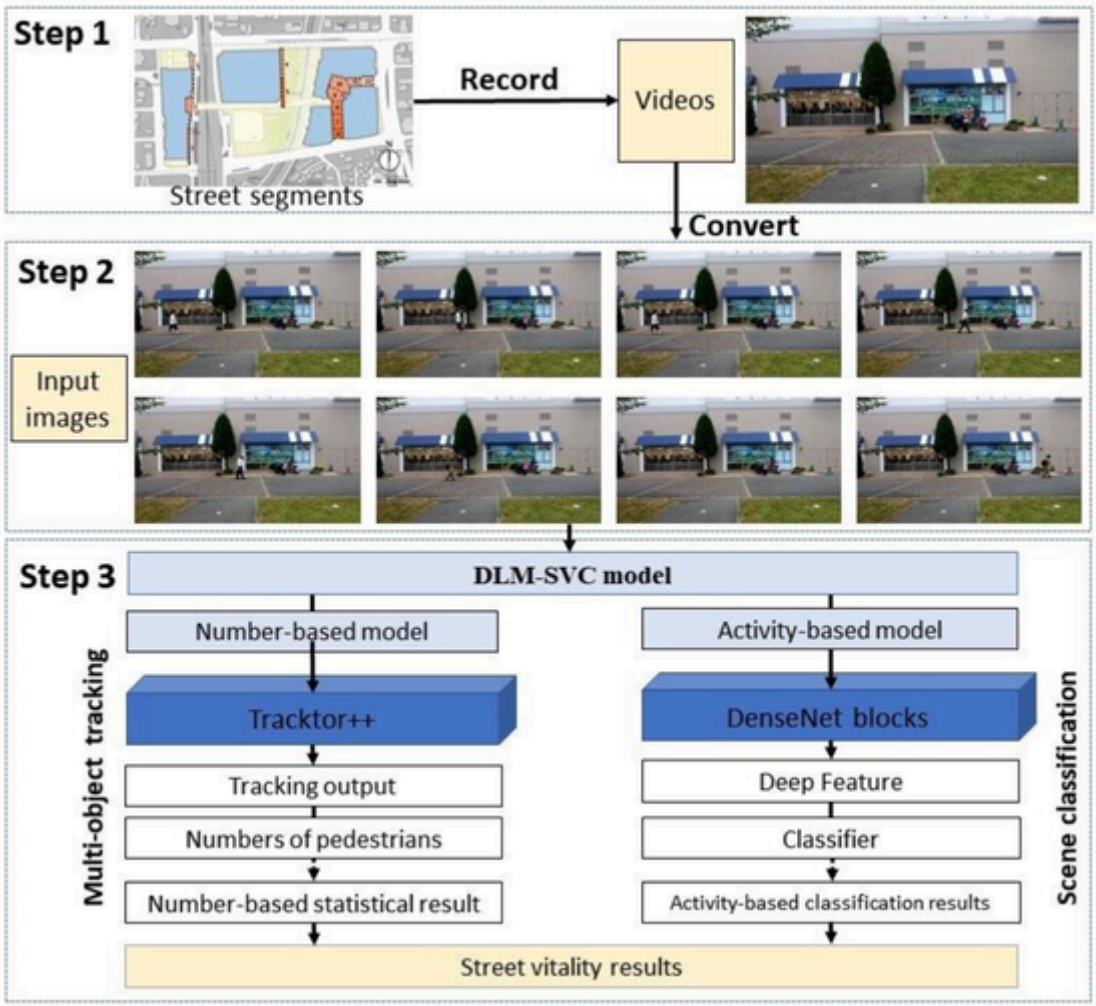

Fig. 3. Technology roadmap of the DLM-SVC model for the street vitality evaluation.

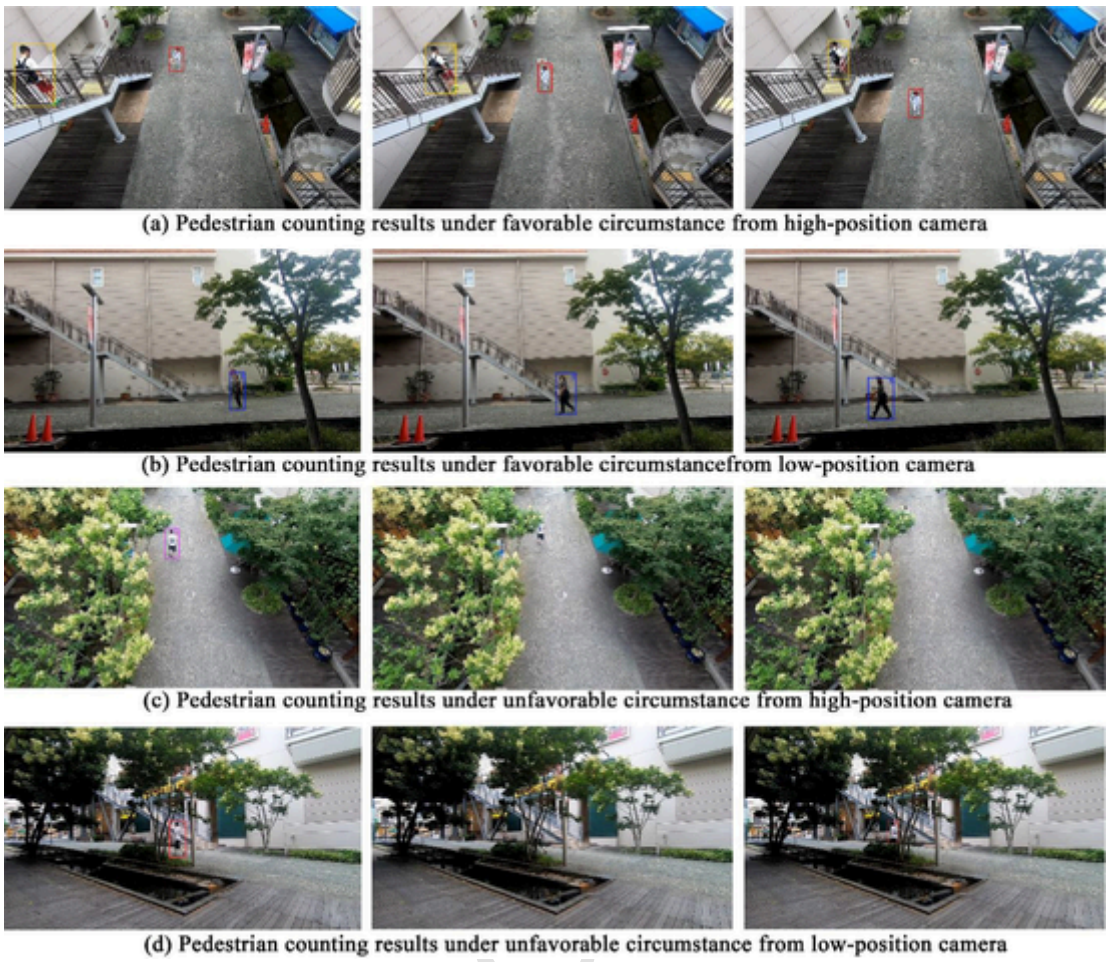

Fig. 4. Examples of pedestrian counting from high-position camera and low-position camera using multiple-object tracking.

and rarely addressed human behavioral choices and their interactions with the built environment.

In this study, we propose a new approach that applies state-of-the-art deep learning-based computer vision techniques to automatically evaluate the association between street vitality and the street built environment quantitatively. Using scene classification and MOT for video analysis based on-site auditing could help classify the intensity of pedestrian activity and examine pedestrian preferences in the physical world. Semantic segmentation for processing street view images makes the quantitative street built environment measurement possible. The purposes of this study are to (1) build a quantitative measurement method for the vitality-related street built environment features, including street width, greenery, openness, transparency, and commercial density; (2) develop an evaluation framework for the vitality of street space, integrating activity-based and pedestrian number-based vitality assessment using video-image data and deep learning; and (3) analyze the association between street vitality level and the street built environment features quantitatively.

The rest of the manuscript is organized as follows. Section 2 explains our research methods framework and details. Section 3 conducts an empirical study on the streets of a commercial complex as a test of our proposed methods. Section 4 summarizes our findings and contributions and discusses the feasibility and limitations of our methodology, and Section 5 concludes remarks.

2. Methodology

In this study, we define street vitality as the intensity of pedestrian activity, including the intensity and complexity of the activities triggered and provided by pedestrians. To achieve our research goal, we build a research framework (Fig. 1) of automatically quantifying the relationship between street vitality and the street built environment containing three main parts. First, we analyze the street built environment variables related to street vitality by combining GIS analysis and semantic segmentation. Second, we propose a deep learning model for street vitality classification (DLM-SVC) based on the intensity of pedestrian behavior and pedestrian volume. The DLM-SVC model includes two parts: an activity-based model using scene classification and a pedestrian number-based model using multiple object tracking. In the activity-based model, a classification hierarchy is designed using an analytic hierarchy process (AHP) and questionnaire. Third, a linear regression model is introduced to explore the internal correlation between the street built environment variables and street vitality.

2.1. Variables of the street built environment and street vitality

2.1.1. Variables of the street built environment

A number of researchers have confirmed the link between various built environment features and pedestrian spatial perception. Gehl (1987) found that the width and length of the street, the diversity of the interface, the flow of motor vehicles, and the difference between indoor and outdoor heights all have different degrees of influence on pedestrian activity. Ewing et al. (2015) investigated the correlation between twenty streetscape features and pedestrian traffic volumes, and found three significant influential features, specifically transparency, active street frontage, and street furniture. A study of Copenhagen's shopping streets further pointed out the direct link be-

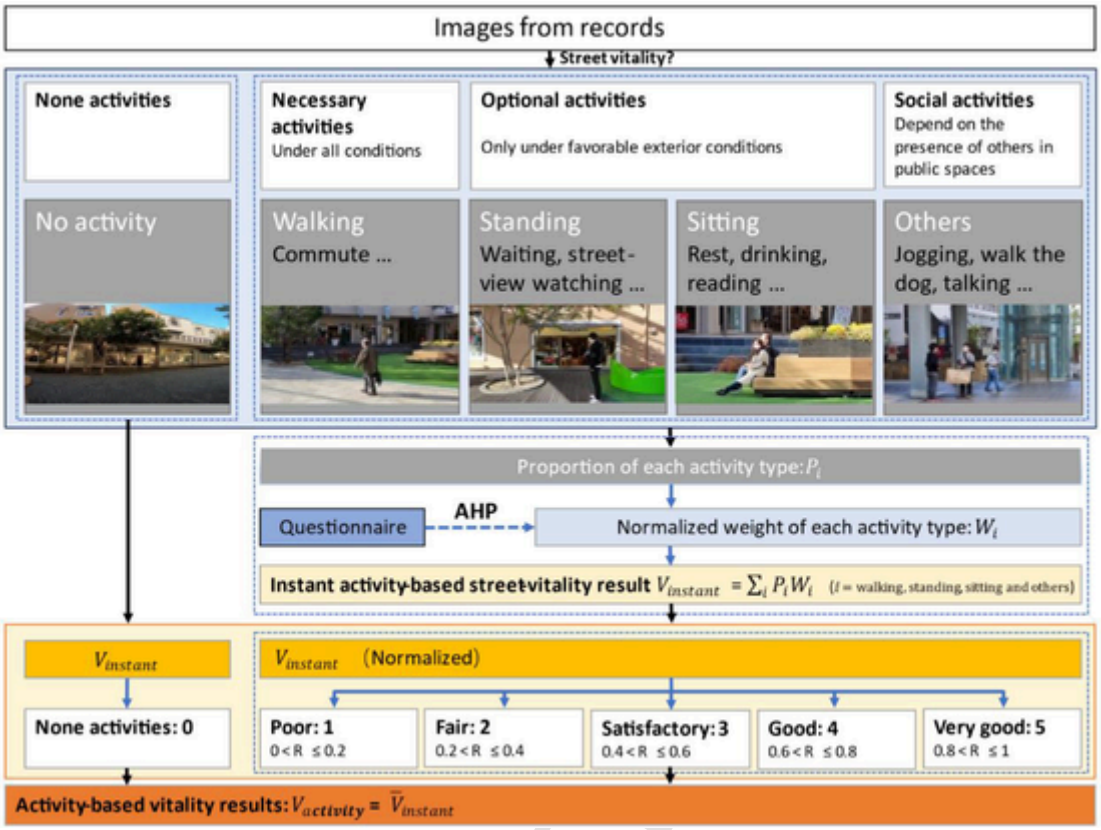

Fig. 5. Classification hierarchy for pedestrian activities of activity-based model.

Table 1
Activity type weightings for classification hierarchy of activity-based model.

Activity type	Walking	Sitting	Standing	Others
Weight	13.620%	28.238%	22.257%	35.886%

Table 2
Characterization of individuals in the questionnaire survey.

Demographics	Group	Frequency	Percent
Gender	Male	59	52.2%
	Female	54	47.8%
Education	Bachelor	34	30.0%
	Master	63	55.8%
	Doctor	16	14.2%
Major	Architecture	42	37.2%
	Urban planning	32	28.3%
	Landscape planning	21	18.5%
	Environment engineering	18	16.0%

tween the built environment and spatial dynamics, which includes the number of stores, boundary transparency, street level units, and the functional mix of street-level buildings (Gehl et al., 2006). Zhang et al. (2020) also studied the influence of the built environment on pedestrian activity in three aspects: transparency, permeability, and façade elements. Based on these studies, the present paper selects five easily quantifiable features of the built environment, namely, street width, greenery, openness, transparency, and commercial density, to investigate their association with street vitality.

Street width is a key impact variable of street-level commercial and social activities. An appropriate street width combined with street furniture arrangements helps to enrich the variety and amount of pedestrian activities. It has been explicitly stated in many urban street design guidelines to ensure pedestrian activities (Ewing et al., 2015). **Greenery** affects pedestrians' thermal comfort and sense of safety. A street space rich in greenery can increase pedestrians' desire to stay (Ewing et al., 2013). The degree of street **openness** has a positive impact on the social activities on the street. A relatively open street interface means the walkers can see more of the sky, which promotes social activities. Open spaces with space penetration tend to become nodes of social activities on the street and places for people to rest and communicate (Yin & Wang, 2016). **Transparency** is the ratio of the area of building openings to the wall area at the street interface. In contrast to the wall interface, which can act as a view blocker, street windows allow for interaction between the inside building activity and the street activity. A transparent ground floor interface is an open interface that deepens and adds layers to the street interface to a certain extent. Many researchers have found that building ground interface transparency is important in shaping excellent pedestrian street spaces (Ewing et al., 2015; Gehl et al., 2006). **Commercial density** is one of the influential primary variables of commercial activity that activates the street ground interface. Some research found that a higher commercial density results in more frequent and varied pedestrian activities (Gehl & Svarre, 2013).

2.1.2. . Variables of street vitality

Because the core of street vitality reflects the people engaged in various activities on the street, including how often and how they use the streets, the street vitality evaluation should include both the volume and activity of pedestrians. We take the total number of pedestrians per unit time on a street segment as the pedestrian volume of pedestrian number-based activity. In addition, we introduced a classification hierarchy of the activity-based vitality to evaluate the complexity of pedestrians' activities. This classification hierarchy is based on the proportion of pedestrian activities of each type and helps to eliminate significant differences in the absolute number and total behavior of pedestri-

Fig. 6. Region maps of study area.

ans between street segments caused by differences in traffic conditions and business patterns.

2.2. Data collection

A set of new data were collected to comprehensively evaluate the street built environment, including road data from OpenStreetMap (OSM), street view images from Google Street View, and POI data from Google Maps in the study area with the help of a corresponding API. When acquiring street view images from websites, we captured street view images in four directions (0, 90, 180, and 270 deg) at each sampling point of each street segment. The lateral images are then converted into panoramic images via skybox images and used for further analysis. In addition, we also collect some panorama images from a 360-deg panorama camera at a height of 2 m as a supplement for places without Street View image data.

To create a collection of street vitality classification maps with a wider range of categories, we first have to take video recordings of several streets in different places for a period of time as primary data. The video information contains detailed and realistic spatial and temporal information about pedestrian behavior and the built environment, which is valuable for classifying the street vitality. The video data are used in two deep learning tasks: scene classification and MOT. According to the specific needs of the scene classification work, the installation of the photographic device should be parallel to the street façade so that the movement and behavior of each pedestrian can be identified. In the MOT task, to improve the calculation accuracy, when installing our cameras, we sought to simplify the video scene and reduce the background complexity to avoid interference with object detection and tracking, such as glass window reflections, merchandise models, and occlusion by large obstacles. The installation height of the camera also matters, because high-position cameras are prone to the problem of foliage blocking the field of view, while low-position cameras are prone to the problem of foreground pedestrian blocking the rear pedestrian when there is a high flow of people. Depending on the site conditions and traffic density in the study area, we can set up multiple cameras at different heights and points for video data collection. The recorded videos are then converted and filtered into images by setting a time interval, which helps minimize image load while ensuring that the results reflect diverse street scenes and continuous pedestrian movement.

2.3. Street built environment measurement

Five impact variables (street width, greenery, openness, transparency, and commercial density) are selected for street built environment evaluation from the literature. Street width and commercial density measurement are achieved in GIS using the geoprocessing function. Street width is the ratio of the total area to the length of the street segment. Commercial density refers to the number of entrances and exits per 20 m of commercial units in a street segment (Ye et al., 2018). In the small-scale study, the commercial density calculation with 20 m as the unit can moderately cover the commercial information and avoid the duplications and missing of data. For calculation of greenery, openness, and transparency, we used semantic segmentation and object detection methods with street-view images. All the street-view images are interpreted as color groups using semantic segmentation models, enabling the calculation of each physical component's pixels. In the greenery and openness calculation, the DeepLabv3+ model (Chen et al., 2018) with the Cityscapes dataset are used and 19 physical components are obtained, such as roads, sidewalks, buildings, walls, vegetation, terrain, and sky. In the transparency calculation, the WEEK 3-Facade parsing model (Liu et al., 2020) with eTRIMS, ECP, and Paris Art Deco datasets are used. To boost the model performance, a region proposal generator based on object detection method is included in the WEEK 3-Facade parsing model. Physical components such as windows, walls, and doors can be better segmented. The percentages of each component in the dataset are summarized. Then the percentage of trees as greenery (Eq. (1)), the sky ratio as enclosure (Eq. (2)), and the sum of window-to-wall and door-to-wall ratio as transparency (Eq. (3)) are used for evaluation. Fig. 2 shows a testing sample of these two segmentation models.

$$P_{greenery} = \sum_{i=1}^{4} GP_i / \sum_{i=1}^{4} P_i, \tag{1}$$

$$P_{openness} = \sum_{i=1}^{4} SP_i / \sum_{i=1}^{4} P_i, \tag{2}$$

$$P_{transparency} = \sum_{i=1}^{4} \left(WP_i + DP_i\right) / \sum_{i=1}^{4} FP_i, \tag{3}$$

Here, GP_i is the number of greenery pixels in image i, SP_i is the number of sky pixels in image i, WP_i is the number of window pixels in image i, DP_i is the number of door pixels in image i, FP_i is the number of wall pixels in image i, and P_i is the total number of pixels in image i.

2.4. Street vitality evaluation and the DLM-SVC model

In recent years, deep learning-based models have been applied in a variety of fields with great success, including image recognition, object

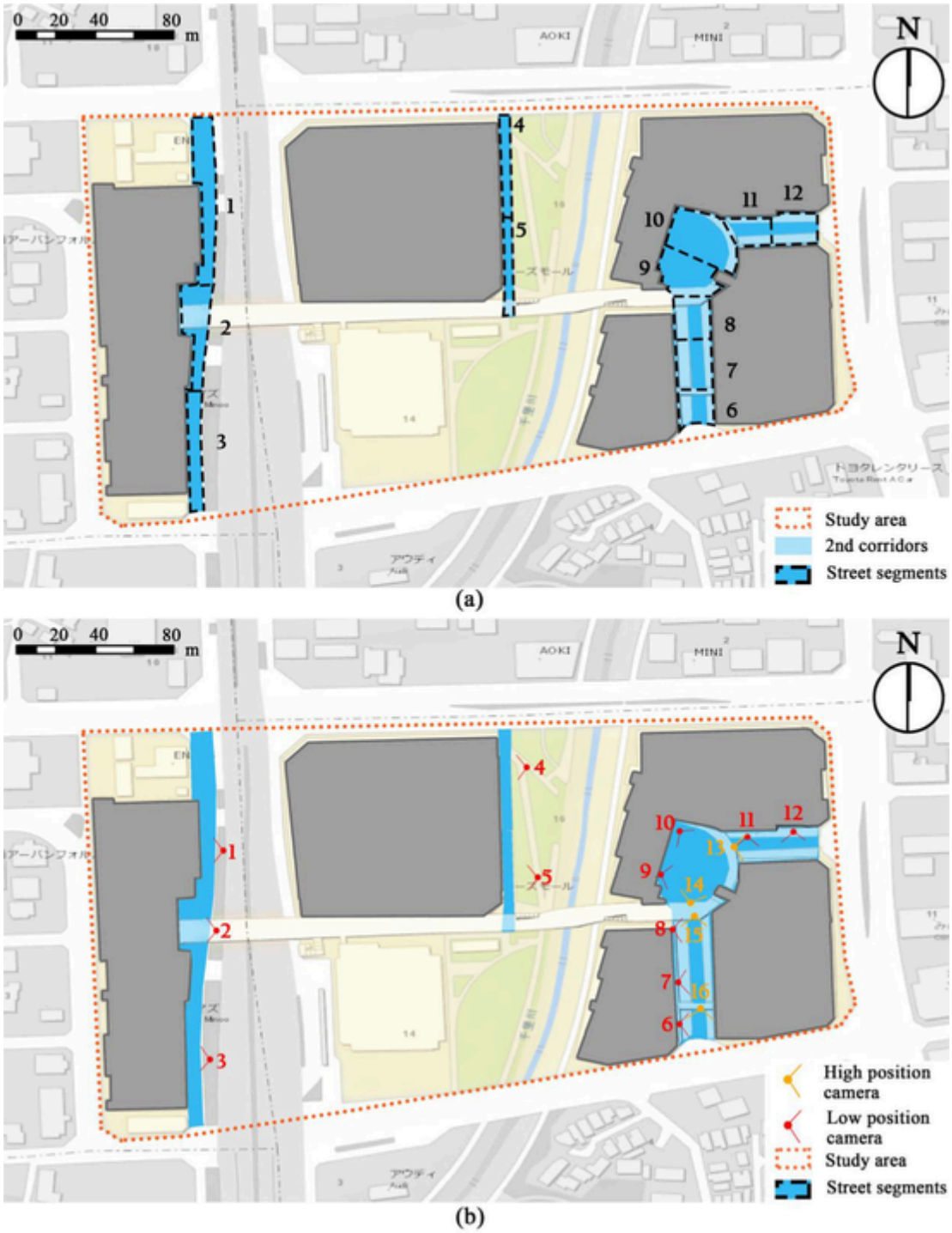

Fig. 7. (a) Distribution of 12 street segments and (b) 16 shooting locations in the study area.

detection and tracking, and image scene segmentation. Inspired by these works, we propose a DLM-SVC model that includes a pedestrian number-based model and an activity-based model that is capable of inferring street vitality from two different aspects: pedestrian numbers and pedestrian activity classification (Fig. 3). First, we need to record videos during survey time in street segments according to the data collection rules in Section 2.2. Then, these videos are converted into images at certain time intervals as raw data of the DLM-SVC model. Finally, these input images are used in number-based model and activity-based model for street vitality evaluation.

2.4.1. Pedestrian number-based model

We use MOT to evaluate pedestrian number-based vitality based on a Faster R-CNN detector, enabling detection, tracking, and ID recordings of multiple objects of interest in a video simultaneously. The standard approach in MOT is tracking by detection, which usually includes four parts: detection, feature extraction or motion prediction, affinity, and association (Ciaparrone et al., 2020).

The MOT model "Tracktor + +" we used was based on deep features proposed by Bergmann et al. (2019) and presented well in most of the easy tracking scenarios. Tracktor + + is a new tracking paradigm that converts an object detector to a tractor and exploits a CNN-based re-

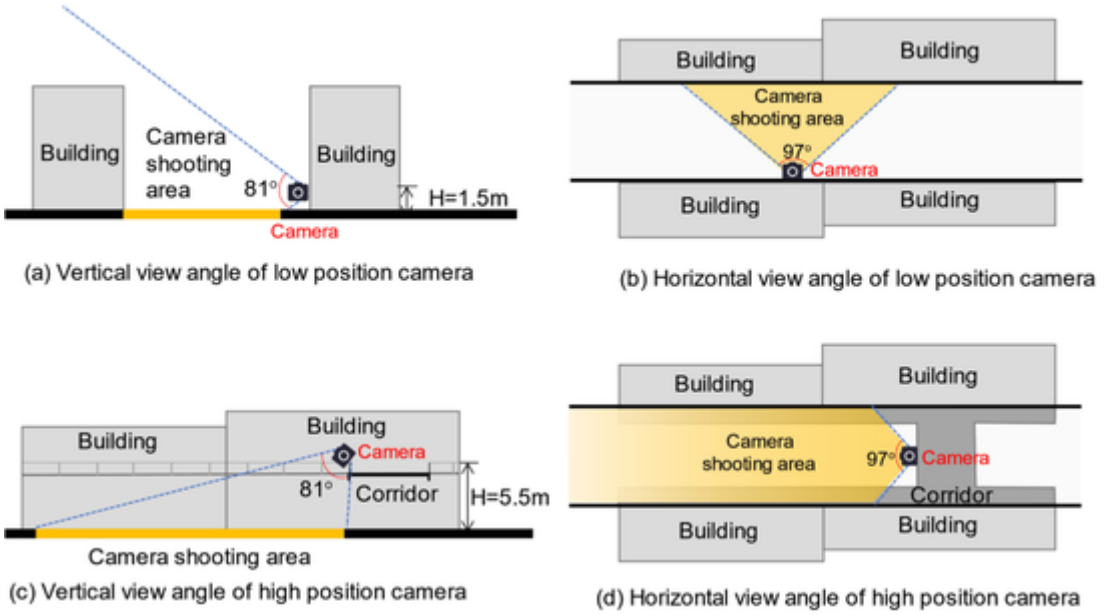

Fig. 8. Installed height and viewing angle of video cameras in low and high positions.

Table 3
Hyperparameter settings of our pre-trained number-based model.

Hyperparameter	Value
How similar do image and old track need to be considered the same person	0.8
How much IoU do track and image need to be considered for matching	0.8
How much time-steps dead tracks are kept and considered for reid	20

Table 4
Pedestrian count results and accuracy of 10 video clip samples.

Video	Camera position	Clip (mm: ss)	Ground truth	Prediction	Accuracy	Overall accuracy
1	High	08:00–11: 00	10	10	100.0%	88.6%
2	High	02:00–05: 00	12	13	84.6%	
3	High	06:10–09: 10	20	21	95.0%	
4	High	04:00–07: 00	6	6	100.0%	
5	Low	10:30–13: 30	5	5	100.0%	
6	Low	00:20–03: 20	24	30	75.0%	
7	Low	12:00–15: 00	21	25	70.8%	
8	Low	01:50–04: 50	13	15	84.6%	
9	Low	06:00–09: 00	16	18	87.5%	
10	High	00:50–03: 50	26	29	88.4%	

gressor for bounding box refinement in the tracking task (Azimi et al., 2021). In other words, the detector is used not only for classification of the target and background, but also for further correction and prediction of the target using regression. The object detector was trained based on a Faster R-CNN with ResNet-101 and Feature Pyramid Networks. Fig. 4 shows an example of pedestrian detection and counting with the proposed model.

Previous studies have shown that the upper limit of a pedestrian street's permissible density is 10–15 people per minute per street width

Table 5
Video image classification accuracy, precision, recall, and F1-Score of the activity-based model in DLM-SVC model.

Overall accuracy	Type	# of Test samples	Precision	Recall	F1-Score
88.7%	Poor	287	92.4%	91.2%	0.91
	Fair	312	84.5%	83.7%	0.84
	Satisfactory	293	90.1%	89.1%	0.90
	Good	305	67.8%	70.6%	0.69
	Very Good	303	93.7%	96.6%	0.95

in meters (Gehl, 1987). Therefore, we quantified the number-based vitality results of a video by Eq. (4) in a range of 0 to 5.

$$V_n = 5 \cdot \left(1 - \left|N_m - \frac{N}{T}\right| / N_m\right), \quad (4)$$

where V_n is the pedestrian number-based street vitality of the site, N_{max} is the maximum number of people per minute (to ensure a more comfortable activity experience of the street, N_{max} in this study is 10), N is the total number of people passing through the video during the survey time, and T is the number of video minutes.

2.4.2. Activity-based model and classification hierarchy

For the activity-based vitality, we aim at integrating existing street vitality classification metrics and propose a classification method based on pedestrian activity using video images to refine the classification of street vitality. This is not the first attempt to use pedestrian activity for vitality evaluation. In the 1980s, Gehl classified pedestrian activities into necessary activities, optional activities, and social activities and found that spontaneous stay-over activities are the leading cause of urban vitality (Gehl, 1987). Many other scholars subsequently did empirical research that has shown some support for Gehl's theory (Sung & Lee, 2015; Wu et al., 2018). Following Gehl's classic work, we used five classifications: no activities, walking, standing, sitting, and other activities (e.g., jogging and dog walking) (Fig. 5). The proportions of each activity type in each video image are the foundation of the classification criteria. An AHP approach calculates weightings for each activity type using pairwise comparison by a questionnaire. AHP requires a number of assessors to use pairwise comparisons to determine the relative importance of indicators, reducing potential decision errors with more reli-

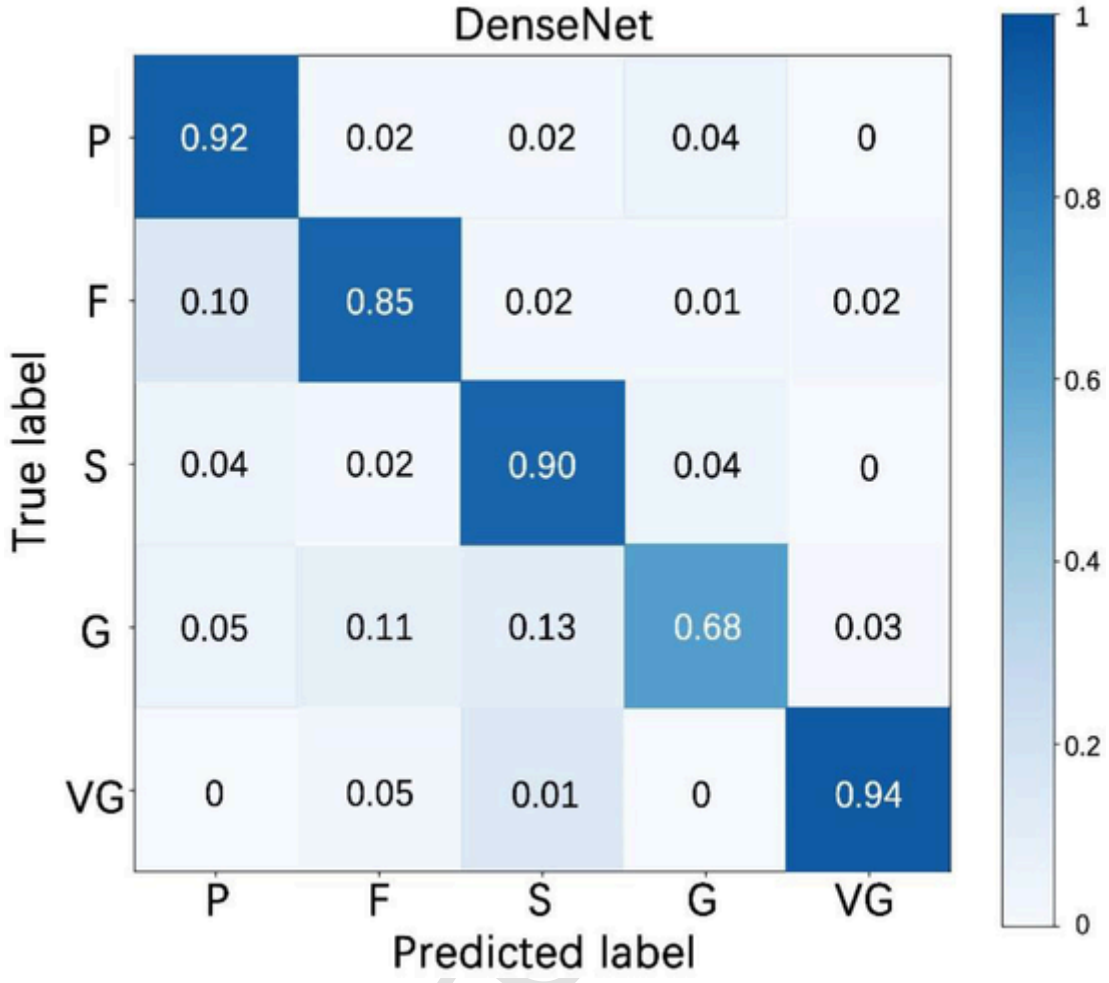

Fig. 9. Confusion matrix for the activity-based classification task.

Table 6
Video image classification accuracy, precision, recall, and F1-Score of activity-based model in DLM-SVC model for 200 image samples.

Overall accuracy	Type	# of Test samples	Precision	Recall	F1-Score
85.8%	Poor	62	93.7%	89.1%	0.91
	Fair	38	82.1%	85.6%	0.84
	Satisfactory	44	87.2%	78.9%	0.83
	Good	18	64.5%	74.1%	0.69
	Very Good	38	85.3%	95.3%	0.90

able results. One hundred thirteen students with an urban design education background were involved in the questionnaire survey, the results of which are presented in Table 1. Table 2 shows the characteristics of the individuals who participated in the survey. In the questionnaire, respondents were asked to consider indicator weights for activity types and perform pairwise comparisons on each pair, including walking, standing, sitting, and other activities. With these weights, an instant activity-based street vitality result of each image could be obtained by Eq. (5):

$$V_{instant} = \sum_i P_i W_i, \tag{5}$$

where $V_{instant}$ is the instant activity-based street vitality of an image, i is the activity type, P_i is the proportion of each activity type in an image, and W_i is the weight of each activity type.

To better classify the activity-based street vitality, we normalized the instant activity-based street vitality results. Theoretically, we can subdivide the categories of video images infinitely. Considering the time and labor costs, we then ranked them into five subcategories, namely P (poor), F (fair), S (satisfactory), G (good), and VG (very good), scores from 1 to 5. If an image contains no activity, then it is assigned a score of 0. The mean value of each instant activity-based street vitality of the site during the survey time is the activity-based street vitality (Eq. (6)). Therefore, we can build a corresponding benchmark dataset containing video images to train the scene classifier for activity-based vitality.

$$V_a = \bar{V}_{instant}, \tag{6}$$

where V_a is activity-based street vitality of the site, and the $\bar{V}_{instant}$ is mean value of each the instant activity-based street vitality of the site during the survey time.

Similar to the image scene understanding task, we propose a scene classification method with DenseNet architecture based on a deep CNN model that does not require statistical pedestrian behavior data (Zhu & Newsam, 2017). Instead, the built DenseNet learns the deep features automatically to support the classification task based on the classification hierarchy and the corresponding benchmark dataset. Therefore, we can obtain both pedestrian number-based street vitality and activity-based street vitality. To quantify street vitality comprehensively, we standardize and average the pedestrian number-based results and the activity-based results on a continuous scale ranging from 0 to 5 (Eq. (7)).

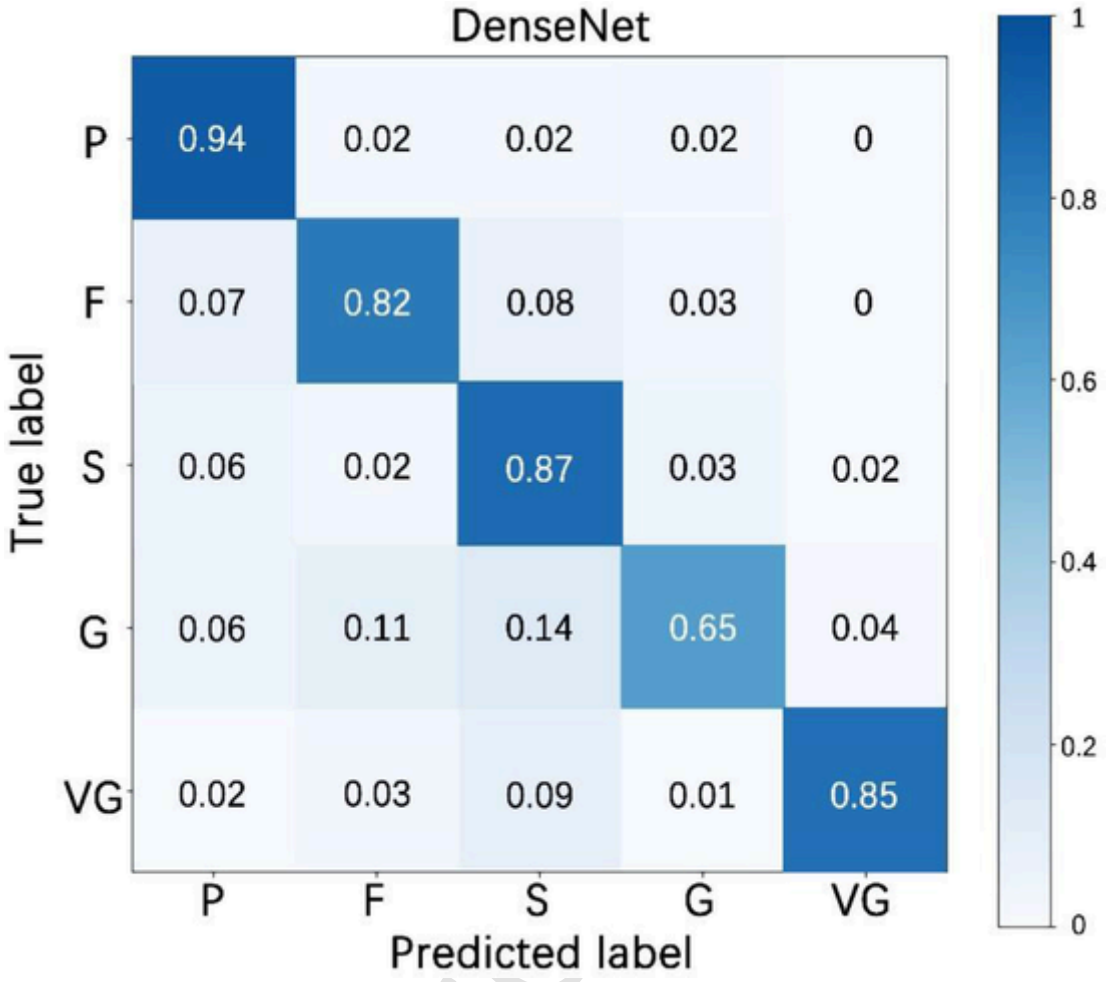

Fig. 10. Confusion matrix for the activity-based classification task of 200 image samples.

Table 7
Street vitality evaluation results for 12 street segments.

Street segment number	1	2	3	4	5	6	7	8	9	10	11	12
street vitality	0.9	1.7	0.8	0.7	2.2	2.8	3.6	4.2	4.5	4.3	3.7	2.4

Note: The street vitality results were obtained by Eqs. (4)–(7).

$$V = \left(V_n + V_a\right)/2, \tag{7}$$

where V is the street vitality of the site, the V_n is number-based street vitality of the site, and the V_a is activity-based street vitality of the site.

2.5. Correlation analysis of the street built environment and street vitality

We use a linear regression model to analyze the relationship between the independent variables and street vitality to explore the effect of the urban street built environment features on pedestrian activities. In this paper, five street built environment features (street width, greenery, openness, transparency, and commercial density) are selected as independent variables, and the street vitality is the dependent variable. To compare the degree of influence of each spatial variable on street vitality, we first standardize the five independent variables to eliminate the effect of different units of covariates. The linear regression model's expression is as follows:

$$Y = a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + a_5X_5 + \mathrm{b},$$

where Y is the vitality in the street segment; X_1, X_2, X_3, X_4, and X_5 represent the independent street vitality variables street width, greenery, openness, transparency, and commercial density, respectively; and a_i is the regression coefficient of the independent variables.

3. Experiments and results

In this section, we conducted an experimental study on the streets of an anonymous commercial complex using our proposed methods. We first introduce the study area and experimental preparations including data collection and model training. Then we validate the DLM-SVC model accuracy. Finally, we present the evaluation results of street built environment and street vitality, and their correlation analysis results.

3.1. Study area

The proposed framework was implemented in the streets of an anonymous commercial complex in Osaka, and 12 street segments were selected in this area, which measured about 0.53 km^2 (Figs. 6 and 7a). The length of each street segment is around 10 m. From the field survey, the research area is a large commercial complex in the surrounding living blocks. It has various space interface features, different street width, transparency, and other built environment elements, and its certain typicality is suitable for correlation analysis. The pedestrian flow here is complicated, and pedestrian behaviors and the street built environment features are diverse. The streets are all pedestrian streets, and the street width is mostly within 5–20 m. Compared with other street blocks, commercial complexes pay more attention to improving the

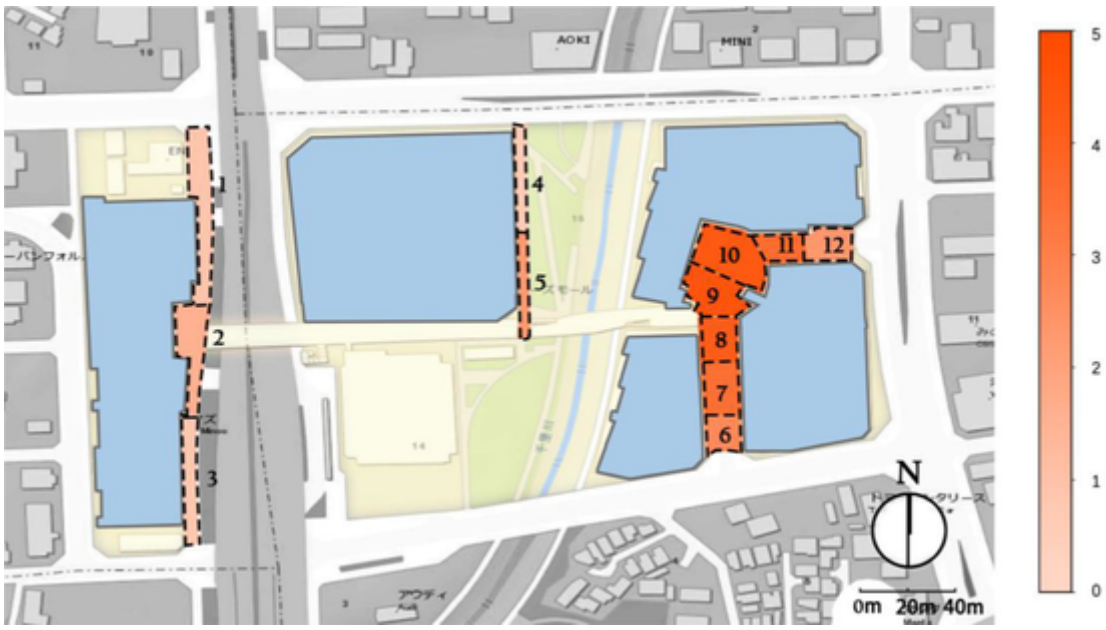

Fig. 11. Visualization results of street vitality evaluation in the study area.

construction of street built environment vitality for driving economic and other vitalities.

In this study, streets with different features in the same commercial complex are selected as research objects because these streets have the same location conditions and thus have basically the same traffic conditions, commercial classes, pedestrian types, and pedestrian access purposes. This helps to weaken the influence of other macro variables related to commercial service level and pedestrian characteristics. Meanwhile, to reduce the influence of large pedestrian flows of commuting purposes around tram stops on the evaluation of commercial vitality, the commercial complexes around tram stops were not selected for the study. In addition, to assimilate as much as possible the non-built environment elements of each survey sample, the data collection was conducted in the same time period with similar weather.

3.2. Experimental setup

3.2.1. Data collection

We first obtained road data from OpenStreetMap, 68 facility POI data from Google Maps, and 12 panoramic camera street-view images in the study area based on the methodology in Section 2.2.

In the pre-survey of the study area, we found that the level of pedestrian activity is low and fewer activity types are present in the morning. In the afternoon, pedestrian activities increased significantly, but after 19:30, the pedestrian flow and the activity characteristics dropped sharply. Based on this background and to ensure the camera's shooting performance, our measurements were conducted during 12:00–18:00 on four weekends (10–11 and 24–25, October 2020; 9–11 and 17–18, April 2021) and eight weekdays (20–23, October 2020; 5–8, April 2021) for a total of 16 days, when the climate and weather were similar and suitable for outdoor activities.

We used four cameras to take six consecutive shots at 16 shooting locations in 12 street segments based on the camera installation principles in Section 2.2, with each shot lasting 15 min per location at 1-hour intervals. Because the pedestrian density in the study area is somewhat low, low-position cameras (height: 1.5 m) were deployed at each street segment for MOT and scene classification tasks. We also set up four shooting locations for high-position cameras (height: 5.5 m) in areas with second-floor pedestrian corridors to collect supplementary video data for the MOT task to improve the accuracy of pedestrian counting. During the experimental time, we acquired a total of 17,280 and 5760 min of video from 12 low-position cameras and 4 high-position cameras at 16 shooting locations, respectively. Figs. 7b and 8 illustrate the location, height, and viewing angle of shooting cameras.

3.2.2. DLM-SVC model training

The DLM-SVC model was trained as follows. We used the MOT17 dataset (Milan et al., 2016) in the pedestrian number-based model training, and Table 3 shows the settings of our pre-trained number-

Table 8
Statistic results of street built environment variables for 12 street segments.

Street segment number	Street width (m)	Greenery	Openness	Transparency	Commercial density (per 20 m)
1	6.0	0.151	0.367	0.427	0.12
2	6.5	0.123	0.125	0.297	0.04
3	5.1	0.140	0.382	0.287	0.03
4	3.2	0.383	0.152	0.196	0.05
5	3.2	0.392	0.182	0.231	0.20
6	10.5	0.218	0.245	0.774	0.11
7	10.3	0.223	0.231	0.732	0.15
8	10.5	0.208	0.243	0.671	0.06
9	15.6	0.124	0.318	0.701	0.18
10	16.2	0.131	0.327	0.742	0.23
11	12.5	0.262	0.224	0.647	0.17
12	10.8	0.227	0.206	0.704	0.06

Table 9
Statistic results of the standardized street built environment variables for 12 street segments.

Street segment number	Standardized street width (m)	Standardized street width Greenery	Standardized street width Openness	Standardized street width Transparency	Standardized street width Commercial density (per 10 m)
1	−0.97189	−0.71534	1.47440	−0.54844	−0.97189
2	−1.42045	−1.02749	−1.57956	−1.17158	−1.42045
3	−1.42045	−0.83797	1.66369	−1.21952	−1.42045
4	−0.68858	1.87103	−1.23883	−1.65572	−0.68858
5	−0.66497	1.97136	−0.86024	−1.06613	−0.66497
6	0.30298	0.03159	−0.06520	1.11486	0.30298
7	0.25576	0.08733	−0.24188	0.91354	0.25576
8	0.30298	−0.07989	−0.09044	0.62114	0.30298
9	1.50701	−1.01634	0.85603	0.76494	1.50701
10	1.64866	−0.93830	0.96961	0.96147	1.64866
11	0.77515	0.52210	−0.33021	0.50610	0.77515
12	0.37380	0.13192	−0.55737	0.77932	0.37380

based model. For the activity-based model training, we take one frame every 20 s from the 17,280 min of video collected from the low-position camera as a dataset. A total of 51,840 images were processed from videos. From these, we randomly collected 7500 images and then manually grouped them into five classes according to the classification hierarchy in Section 2.4 and the number of labeled images in each class was around 1500. 80% and 20% of these 7500 video images are used as a training subset and a testing subset, respectively. The learned activity-based model was tested on the testing subset after the training process. Finally, the well-trained model successfully finished the prediction of the remaining 44,340 pictures out of 51,840 images.

3.3. Accuracy verification of the DLM-SVC model

3.3.1. Pedestrian number-based model accuracy

Various classic metrics can comprehensively reflect the performance of the MOT model, such as MOT Accuracy, Identification F1, and Mostly Tracked trajectories. Because we focus on pedestrian counting in the number-based model, we did not use classic metrics to validate the number-based model accuracy. Instead, we randomly selected 10 video clip samples based on field surveys during 12:00–18:00 in the study area and compared the predicted pedestrian counts from the pre-trained model with manual statistics results (Kim, 2020). The choice of 10 video clip samples is the result of balancing the representativeness of the sampled video and the collection cost of sampling video. The accuracy of number-based model counting results is 88.6%, as shown in Table 4. Videos taken with high-position cameras and videos with a low total number of pedestrians have a higher accuracy rate.

3.3.2. Activity-based model accuracy

Classification accuracy of the activity-based model. For the activity-based model, the classification accuracy is shown in Table 5. The percentage of correct classifications is used as the classification accuracy of the activity-based model. If the predicted result of a video image from the trained model and the labeled type is consistent, it is considered as a correct classification; otherwise, it is considered incorrect. The overall classification accuracy of the activity-based model is 88.7% for five classification types. Compared with other subcategories, due to the small test samples, the precision of the "good" subcategory is the lowest (67.8%). Because of the sample's complexity and diversity, this low-precision result is acceptable.

To assess the performance of the model in distinguishing pairs of similar categories, the confusion matrices of the activity-based model are presented in Fig. 9. The percentage of samples from one category correctly categorized into another category by the model is represented by the values in the matrix. Some categories are more easily misclassified into each other, especially when an image contains a large number of different objects or overlapping objects.

Validation based on field survey data. To further validate the accuracy of the activity-based model, a random sample of 200 images based on field surveys in the study area were selected and labeled with activity-based vitality results. This sample size can balance time cost and accuracy in the model validation (Hu et al., 2020; Zhang et al., 2021c). The overall classification accuracy of the activity-based

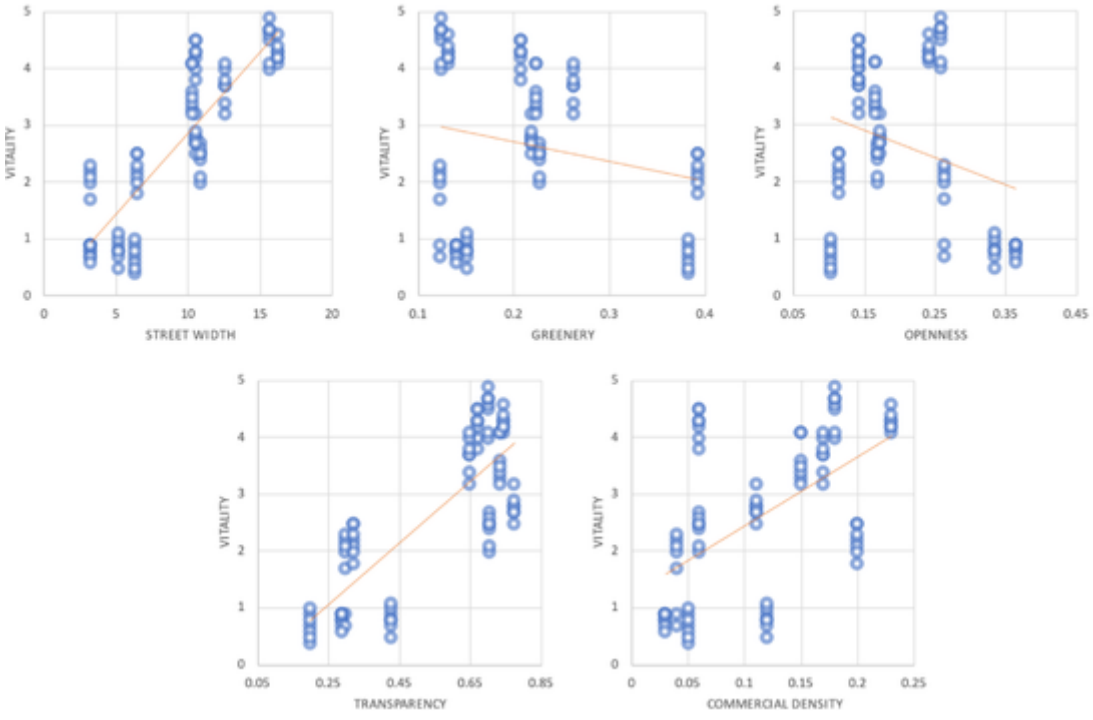

Fig. 12. Scatter plot analysis of street vibrancy and street built environment variables.

Table 10
Linear regression results of street vitality and street built environment variables.

	Regression coefficient	Standardization coefficient	t	P	VIF	R^2
Constant	2.647	–	41.888	0.000	–	0.814
Street width	0.761	0.545	5.287	0.000***	3.941	
Greenery	−0.278	−0.199	−3.210	0.002**	1.864	
Openness	−0.338	−0.242	−4.098	0.000***	1.688	
Transparency	0.372	0.266	2.918	0.004**	3.029	
Commercial density	0.239	0.171	2.616	0.010*	2.073	

D-W Value: 1.800.
*$p < 0.1$, **$p < 0.05$, ***$p < 0.001$.

model is 85.8%. Table 6 shows the classification accuracy of the 200 image samples, and Fig. 10 presents the confusion matrix of the 200 image samples. These results are similar to those in Table 3 and Fig. 8 which proved the trained activity-based model's performance; that is, the model's prediction results, are normally consistent with the verification results based on random samples.

3.4. Street built environment and street vitality results

Table 7 and Fig. 11 show the street vitality results for the 12 street segments evaluated by the DLM-SVC model, and Table 8 illustrates the street built environment variables results of the proposed method. The diversity of pedestrian activities is higher at segments 9 and 10, with a higher proportion of rest and viewing. The road is broad, and the street space is composed of a small circular square, an oriented, open, and pleasant space. At the same time, the place lies at the traffic node of three buildings in the commercial complex, where the flow of the pedestrian movement is high, providing the possibility for people to see and be seen with each other. The commercial form of street segments 4 and 5 is high-quality interior mega-commercial with low transparency of 0.196 and 0.231, respectively. Although they are adjacent to a landscaped park and the greenery is high, 0.383 and 0.392, respectively, the overall activity is low. This indicates that transparency is an important variable in attracting activities for streets within commercial complexes. Because small units of highly transparent commercial stores are informative about their products, people may approach goods and businesses in transparent stores, but interior commercial forms visually block people from engaging with goods and businesses.

3.5. Correlation analysis results

Due to the different magnitudes and units of the independent variables in this study, we first standardized the raw data of street built environment variables (Table 9). After organizing the above data and data of the average vitality values per hour into five scatter plots of 72 points each (Fig. 12), we found the following: (1) The fitted line in the scatter plots of street width, commercial density, and street vitality showed a significant increasing trend, with street vitality decreasing as the street width increased. (2) Transparency has a very strong influence on street vitality, and the fitted line shows a significant upward trend. (3) Openness, greenery, and commercial density all show a steady trend with the fitted line of the scatter plot of vitality value.

A linear regression model was built to compute the weight coefficients of each variable so that the effect of these five selected built-environment variables on street vitality could be compared. In the regression model, mean vitality values for each hour of each day on 12 street segments during the experimental time and standardized street environment variable data comprised 1152 sets of samples. The regression results are shown in Table 10. The fitting degree (R^2 value) of the model was 0.814, which means that these five selected street built environment variables can explain 81.4% of the variation in street vitality. The model passed the F-test ($F = 78.757$, $p < 0.001$), which means that at least one of the features street width, greenery, openness, and transparency has an effect on vitality. All the variance inflation factor (VIF) values in the model are less than 5 and sample number is 1152, which means that there is no cointegration problem (O'Brien, 2007). The Durban–Watson (D-W) value is around 2, which means that the model has no autocorrelation and the sample data have no correlation; thus, the model is good. The smaller the P-value is, the more significant the correlation is. The regression results show that street width has a significant positive effect on street vitality, followed by transparency with positive correlation. Meanwhile, openness and greenery have a negative relationship with street vitality. Commercial density has a positive impact on street vitality.

4. Discussion

This study investigated a method to quantitatively analyze the association between street vitality level and the street built environment features by quantitative measurement of the vitality-related street built environment features (street width, greenery, openness, transparency, and commercial density) and a vitality evaluation framework based on proposed DLM-SVC model. The street built environment variables were measured by GIS analysis and semantic segmentation using street view images. The DLM-SVC model could analyze street vitality based on the intensity of pedestrian behavior and pedestrian volume using video-image data. The linear regression model results with raw data standardization can represent the relationship between street vitality and street built environment variables in the experimental area. Street view images and video images were reliable, efficient and cost-effective data sources in analyzing street built environment and pedestrian activities. In the experiments, we carefully selected 16 camera locations in public spaces that did not interfere with pedestrian activity, and all shots were taken in compliance with local laws.

In general, the main contributions are listed below.

- Compared with the methods measured through traditional manual on-site surveys, in this paper, we developed an automatic measurement method for street built environment variable evaluation and pedestrian activity-based street vitality evaluation using state-of-the-art deep learning models, street view image, and video data, while combining regression models to analyze the relationship between street vitality and street built environment variables, greatly improving the efficiency of data analysis.
- We applied our method to a commercial complex. The validation results show that our method is effective and scalable. In our proposed DLM-SVC model, both the quantity-based and activity-based sub-models have good accuracy.
- We built a tailored video image dataset for training a scene classifier for street vitality in the activity-based sub-model of the DLM-SVC model.

Our research provides feasibility and reliability by using deep learning methods to explore the relationship between typical street built environment features and street vitality. Most of the findings are consistent with previous studies (Ewing et al., 2015; Hamidi & Moazzeni, 2019), and the findings suggest that:

- Street width is one of the primary variables affecting street activity. High-vitality pedestrian activities are spread over streets 15–20 m wide, which means stay-over activities need a certain street width to avoid adjacent pedestrian flow.
- The transparency of the building interface along the street has a potentially beneficial role in pedestrian activities (Fig. 13a).

Fig. 13. Field survey images of street segments: (a) Low vitality with high greenery and low transparency in segment 4; (b) Poorly placed greenery in segment 12; (c) High openness but centripetal space in segment 10; (d) High vitality with high greenery but centripetal-free space in segment 1.

However, when the transparency exceeds 70%, the street dynamics no longer increases significantly. This suggests that a transparent and open ground floor interface can, on the one hand, attract commercial stay-over activities. Still, on the other hand, it weakens the boundary effect of the ground floor and discourages certain social stay-over activities that require a sense of enclosure from boundaries.

- The effect of greenery and openness on street vitality shows a degree of negative correlation and may not be a simple linear relationship. While a certain amount of greenery is beneficial to social activities, such as sitting, poorly placed greenery may obstruct street commercial interfaces and signs, causing commercial activities to suffer (Fig. 13b).
- Openness shows a potentially significant negative effect on street vitality. Open and centripetal-free space can cause pedestrians to lose their sense of security and enclosure, which is also detrimental to the boundary effect and pedestrian activities (Fig. 13c). In addition, in this experiment, the larger openness tends to suggest that the street segment has a discordant street height to width ratio and is dominated by traffic functions (Fig. 13d).
- Among the five examined variables, commercial density shows the weakest impact on street vitality, which may be because the study area is located in a commercial complex where commercial distribution is dense and less differentiated. Further research can be conducted to refine this association by selecting different commercial complex streets and expanding the study area.

Based on the previous analysis, we could find some suitable values of street built environment features to maximize street vitality in the study area and propose some practical suggestions concisely and effectively for architects and urban planners. These urban street design recommendations are to a certain extent applicable to urban streets with similar characteristics as the experimental study area. Since there were only 12 street segments in the experiment, it is difficult to generalize the existing findings (15–20 m wide, 70% of transparency, etc.) to the usual cases. However, more streets can be studied using our methodology, thus providing more site-specific suggestions for street design in terms of improving vitality rather than one-size-fits-all theorems.

It is important to note that in practice, these five street built environment variables are mutually influential in some locations, and there might be some autocorrelation phenomena. For example, street segments with many storefronts are often those with a wide street and high interface penetration. It is the effective overlapping and complementarity of these elements together that contributes to a vibrant street life. In addition, the same street built environment features may have different effects on different activity types. Therefore, while the overall control of the street interface attribute is important, the space design of each node in the street segment and the rational layout of street facilities need to be considered in detail.

Compared with traditional methods, such as the PSPL survey, our proposed framework is transferable and efficient and is suitable for this big data era. First, the DLM-SVC model is easier to transfer to other scenarios with the model training process. Second, although the accuracy of the traditional methods is higher, it relies heavily on labor-intensive acquisition of field survey data. Our methods could analyze a large number of scenes and achieve complex calculations with big data. Furthermore, with the development of smart cities, our proposed model has data-driven potential, and the accuracy of the model could be further improved by an updated large training set of street-level pedestrian behavior data provided from Internet of Things devices.

This study has some limitations. First, the current activity-based classification model in the DLM-SVC model does not consider the video's temporal order. The accuracy of the street vitality evaluation can be further improved by expanding the training set and adding a model with temporal sequencing. Using an action recognition model is another choice to obtain precise activity-based vitality results rather than the classified results. Second, the selected street built environment variables are not all-inclusive. It must be acknowledged that many variables affect street vitality, such as building façade details, merchandise display content, and street furniture. These variables can have different degrees of influence but are limited by data quantification and data collection. Future research could include more relevant environment variables, such as street continuity, to comprehensively explore what kind

of space can attract people's activities. In addition, although our approach could be transferred to other streets, the current experimental study was limited to buildings in a commercial complex. If the method is applied to other streets, the correlation results may be different and need further discussion.

5. Conclusions

Few studies have applied the deep learning approach to explore the association between street vitality and the street built environment. Even fewer studies have evaluated street vitality with video image data. Some previous investigations have used the proportion of activity type as an indicator to infer a street's vitality. Therefore, the proportion of activity type in a video image is a suitable and similar evaluation of the activity-based vitality.

This article established a research framework to quantify the relationship between the built environment and street vitality and validated our approach with an experiment in a commercial complex, which can be referenced as a toolbox for further studies and practical projects with other types of streets. Specifically, we use MOT and scene segmentation to build a DLM-SVC model that automatically evaluates street vitality based on the number of people and activities in the street from video data, then propose to quantify the elements of the built environment using GIS analysis and semantic segmentation, and finally, introduce these variables into a regression model. With the deep learning algorithms and urban digital data, our method is more data-driven and can be applied to the streets of other cities. The validation results indicate that the proposed DLM-SVC model is feasible and reliable for classifying street vitality and can be applied to study the influence of street built environment variables on street vitality. The revealed associations between street vitality and the street built environment in an example commercial complex show that street width, greenery, openness and transparency are significant variables in improving street vitality. These results can help put forward opinions and suggestions for urban street planning and street-level architectural design in similar street blocks.

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.

References

Angah, O., & Chen, A. Y. (2020). *Tracking multiple construction workers through deep learning and the gradient based method with re-matching based on multi-object tracking accuracy. Automation in Construction*, *119*, 103308. https://doi.org/10.1016/j.autcon.2020.103308.

Azimi, S. M., Kraus, M., Bahmanyar, R., & Reinartz, P. (2021). *Multiple pedestrians and vehicles tracking in aerial imagery using a convolutional neural network. Remote Sensing*, *13*(10), 1953. https://doi.org/10.3390/rs1310195.

Bergmann, P., Meinhardt, T., & Leal-Taixe, L. (2019). Tracking without bells and whistles. *Proceedings of the IEEE international conference on computer vision* (pp. 941–951).

Buchanan, P. (1988). *What city? A plea for place in the public realm. The Architectural Review*, *184*(1101), 31–41.

Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. *Proceedings of the European conference on computer vision (ECCV)* (pp. 801–818).

Ciaparrone, G., Luque Sánchez, F., Tabik, S., Troiano, L., Tagliaferri, R., & Herrera, F. (2020). *Deep learning in video multi-object tracking: A survey*. *Neurocomputing*, *381*, 61–88. https://doi.org/10.1016/j.neucom.2019.11.023.

Dong, Y. H., Peng, F. L., & Guo, T. F. (2021). *Quantitative assessment method on urban vitality of metro-led underground space based on multi-source data: A case study of Shanghai Inner Ring area. Tunnelling and Underground Space Technology*, *116*, 104108. https://doi.org/10.1016/j.tust.2021.104108.

Dziedzic, J. W., Da, Y., & Novakovic, V. (2019). *Indoor occupant behaviour monitoring with the use of a depth registration camera. Building and Environment*, *148*, 44–54. https://doi.org/10.1016/j.buildenv.2018.10.032.

Ewing, R., Connors, M. B., Goates, J. P., Hajrasouliha, A., Neckerman, K., Nelson, A. C., et al. (2013). Validating urban design measures. *Presented at the Transportation Research Board 92nd Annual MeetingTransportation Research Board*. Retrieved from https://trid.trb.org/view/1241107.

Ewing, R., Hajrasouliha, A., Neckerman, K., Purciel-Hill, M., & Greene, W. (2015). *Streetscape features related to pedestrian activity*. *Journal of Planning Education and Research*, 36. https://doi.org/10.1177/0739456X15591585.

Gehl, J. (1987). *Life between buildings: Using public space*. Island press.

Gehl, J., Kaefer, L., & Reigstad, S. (2006). *Close encounters with buildings*. *Urban Design International*, *11*, 29–47. https://doi.org/10.1057/palgrave.udi.9000162.

Gehl, J., & Svarre, B. (2013). *Public space, public life: An interaction. How to study public life* (pp. 1–8). Springer.

Gong, F. Y., Zeng, Z. C., Zhang, F., Li, X., Ng, E., & Norford, L. K. (2018). *Mapping sky, tree, and building view factors of street canyons in a high-density urban environment. Building and Environment*, *134*, 155–167. https://doi.org/10.1016/j.buildenv.2018.02.042.

Hamidi, S., & Moazzeni, S. (2019). *Examining the relationship between urban design qualities and walking behavior: Empirical evidence from dallas, TX*. *Sustainability*, *11*, 2720. https://doi.org/10.3390/su11102720.

He, S., Yu, S., Wei, P., & Fang, C. (2019). *A spatial design network analysis of street networks and the locations of leisure entertainment activities: A case study of Wuhan, China. Sustainable Cities and Society*, *44*, 880–887. https://doi.org/10.1016/j.scs.2018.11.007.

Hou, J., Chen, L., Zhang, E., Jia, H., & Long, Y. (2020). *Quantifying the usage of small public spaces using deep convolutional neural network*. *PloS One*, *15*(10), e0239390. https://doi.org/10.1371/journal.pone.0239390.

Hu, C. B., Zhang, F., Gong, F. Y., Ratti, C., & Li, X. (2020). *Classification and mapping of urban canyon geometry using google street view images and deep multitask learning. Building and Environment*, *167*, 106424. https://doi.org/10.1016/j.buildenv.2019.106424.

Jacobs, J. (1961). *The death and life of great American cities*. Vintage.

Kim, D. (2020). *Pedestrian and bicycle volume data collection using drone technology*. *Journal of Urban Technology*, *27*(2), 45–60. https://doi.org/10.1080/10630732.2020.1715158.

Lees, L. (2010). *Commentary*. *Environment and Planning A: Economy and Space*, *42*(10), 2302–2308. https://doi.org/10.1068/a43360.

Li, Y., Yabuki, N., Fukuda, T., & Zhang, J. (2020). September 16. *A big data evaluation of urban street walkability using deep learning and environmental sensors-a case study around osaka university suita campus: 2* (pp. 319–328). Berlin, Germany: TU Berlin.

Liang, S., Leng, H., Yuan, Q., Wang, B., & Yuan, C. (2020). *How does weather and climate affect pedestrian walking speed during cool and cold seasons in severely cold areas? Building and Environment*, *175*, 106811. https://doi.org/10.1016/j.buildenv.2020.106811.

Liu, H., Xu, Y., Zhang, J., Zhu, J., Li, Y., & Hoi, C. S. (2020). *DeepFacade: A deep learning approach to facade parsing with symmetric loss. IEEE Transactions on Multimedia*.

Lopes, M. N., & Camanho, A. S. (2013). *Public green space use and consequences on urban vitality: An assessment of European cities*. *Social Indicators Research*, *113*(3), 751–767. https://doi.org/10.1007/s11205-012-0106-9.

Maas, P. R. (1984). *Towards a theory of urban vitality*. University of British Columbia. PhD Thesis.

Marcus, L. (2010). *Spatial capital. The Journal of Space Syntax*, *1*(1), 30–40.

Milan, A., Leal-Taixé, L., Reid, I., Roth, S., & Schindler, K. (2016). *MOT16: A benchmark for multi-object tracking. ArXiv Preprint ArXiv:1603.00831*.

Montgomery, J. (1998). *Making a city: Urbanity, vitality and urban design*. *Journal of Urban Design*, *3*(1), 93–116. https://doi.org/10.1080/13574809808724418.

Mouratidis, K., & Poortinga, W. (2020). *Built environment, urban vitality and social cohesion: Do vibrant neighborhoods foster strong communities? Landscape and Urban Planning*, *204*, 103951. https://doi.org/10.1016/j.landurbplan.2020.103951.

O'Brien, R. M. (2007). *A caution regarding rules of thumb for variance inflation factors*. *Quality & Quantity*, *41*(5), 673–690. https://doi.org/10.1007/s11135-006-9018-6.

Quintanar, A., Fernández-Llorca, D., Parra, I., Izquierdo, R., & Sotelo, M. A. (2021). *Predicting vehicles trajectories in urban scenarios with transformer networks and augmented information. ArXiv Preprint ArXiv:2106.00559*.

Sulis, P., Manley, E., Zhong, C., & Batty, M. (2018). *Using mobility data as proxy for measuring urban vitality*. *Journal of Spatial Information Science*, *16*, 137–162.

Sung, H., & Lee, S. (2015). *Residential built environment and walking activity: Empirical evidence of Jane Jacobs' urban vitality*. *Transportation Research Part D: Transport and Environment*, *41*, 318–329. https://doi.org/10.1016/j.trd.2015.09.009.

Wu, J., Ta, N., Song, Y., Lin, J., & Chai, Y. (2018). *Urban form breeds neighborhood vibrancy: A case study using a GPS-based activity survey in suburban Beijing. Cities (London, England)*, *74*, 100–108. https://doi.org/10.1016/j.cities.2017.11.008.

Xia, C., Yeh, A. G. O., & Zhang, A. (2020). *Analyzing spatial relationships between urban land use intensity and urban vitality at street block level: A case study of five Chinese megacities. Landscape and Urban Planning*, *193*, 103669. https://doi.org/10.1016/j.landurbplan.2019.103669.

Ye, Y., Li, D., & Liu, X. (2018). *How block density and typology affect urban vitality: An exploratory analysis in Shenzhen, China*. *Urban Geography*, *39*(4), 631–652. https://doi.org/10.1080/02723638.2017.1381536.

Yin, L., & Wang, Z. (2016). *Measuring visual enclosure for street walkability: Using machine learning algorithms and google street view imagery*. *Applied Geography*, *76*, 147–153. https://doi.org/10.1016/j.apgeog.2016.09.024.

Yue, W., Chen, Y., Thy, P. T. M., Fan, P., Liu, Y., & Zhang, W. (2021). *Identifying urban vitality in metropolitan areas of developing countries from a comparative perspective: Ho Chi Minh City versus Shanghai. Sustainable Cities and Society*, *65*, 102609. https://doi.org/10.1016/j.scs.2020.102609.

Zeng, C., Song, Y., He, Q., & Shen, F. (2018). *Spatially explicit assessment on urban vitality: Case studies in Chicago and Wuhan. Sustainable Cities and Society*, *40*, 296–306. https://doi.org/10.1016/j.scs.2018.04.021.

Zeng, Q., Zhang, M., & Yuan, F. (2019). The influence of the interface characteristics of street in historical and cultural blocks on the vitality of the block. *Proceedings of the international conference on computer, network, communication and information systems (CNCI 2019)*. Atlantis Press. https://doi.org/10.2991/cnci-19.2019.60.

Zhang, A., Li, W., Wu, J., Lin, J., Chu, J., & Xia, C. (2021a). *How can the urban landscape affect urban vitality at the street block level? A case study of 15 metropolises in China. Environment and Planning B: Urban Analytics and City Science*, *48*(5), 1245–1262. https://doi.org/10.1177/2399808320924425.

Zhang, C., Berger, C., & Dozza, M. (2021b). *Social-IWSTCNN: A social interaction-weighted spatio-temporal convolutional neural network for pedestrian trajectory prediction in urban traffic scenarios. ArXiv Preprint ArXiv:2105.12436.*

Zhang, J., Fukuda, T., & Yabuki, N. (2021c). *Development of a city-scale approach for façade color measurement with building functional classification using deep learning and street view images. ISPRS International Journal of Geo-Information*, *10*(8), 551. https://doi.org/10.3390/ijgi10080551.

Zhang, L., Zhang, R., & Yin, B. (2020). *The impact of the built-up environment of streets on pedestrian activities in the historical area. Alexandria Engineering Journal*. https://doi.org/10.1016/j.aej.2020.08.008.

Zhu, Y., & Newsam, S. (2017). Densenet for dense flow. *Proceedings of the IEEE international conference on image processing (ICIP)* (pp. 790–794). IEEE. https://doi.org/10.1109/ICIP.2017.8296389.