

Article

Street Design for Hedonistic Sustainability through AI and Human Co-Operative Evaluation

Kanyou Sou ¹, Hiroya Shiokawa ², Kento Yoh ^{1,*}  and Kenji Doi ¹

¹ Department of Civil Engineering, Division of Global Architecture, Graduate School of Engineering, Osaka University, Osaka 5650871, Japan; sou.kanyou@civil.eng.osaka-u.ac.jp (K.S.); doi@civil.eng.osaka-u.ac.jp (K.D.)

² Department of Engineering, Tokyo Electric Power Company Holdings (TEPCO), Inc., Tokyo 1080023, Japan; shiokawa.hiroya@tepcoco.jp

* Correspondence: yoh.kento@civil.eng.osaka-u.ac.jp

Abstract: Recently, there has been an increasing emphasis on community development centered on the well-being and quality of life of citizens, while pursuing sustainability. This study proposes an AI and human co-operative evaluation (AIHCE) framework that facilitates communication design between designers and stakeholders based on human emotions and values and is an evaluation method for street space. AIHCE is an evaluation method based on image recognition technology that performs deep learning of the facial expressions of both people and the city; namely, it consists of a facial expression recognition model (FERM) and a street image evaluation model (SIEM). The former evaluates the street space based on the emotions and values of the pedestrian's facial expression, and the latter evaluates the target street space from the prepared street space image. AIHCE is an integrated framework for these two models, enabling continuous and objective evaluation of space with simultaneous subjective emotional evaluation, showing the possibility of reflecting it in the design. It is expected to contribute to fostering people's awareness that streets are public goods reflecting the basic functions of public spaces and the values and regional characteristics of residents, contributing to the improvement of the sustainability of the entire city.

Keywords: street and urban design; co-operative design; deep learning; new normal era; walkable city; lingerability; facial expression; happiness level; well-being



Citation: Sou, K.; Shiokawa, H.; Yoh, K.; Doi, K. Street Design for Hedonistic Sustainability through AI and Human Co-Operative Evaluation. *Sustainability* **2021**, *13*, 9066. <https://doi.org/10.3390/su13169066>

Academic Editor: Tomio Miwa

Received: 14 June 2021

Accepted: 9 August 2021

Published: 13 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Research Background

In recent years, health-oriented and carbon neutral-oriented lifestyles are being pursued, and the Ministry of Land, Infrastructure, Transport, and Tourism in Japan has set “a city where you want to walk comfortably” as the direction of future urban development in 2019. In accordance with this policy direction, 288 municipalities have been designated as “Walkability Promotion Cities”, and efforts are being made towards forming a city that makes people feel comfortable and want to walk until the end of 2020 [1]. However, with the spread of the COVID-19 infection, the frequency of going out on foot or by bicycle has increased because of increased anxiety about public transport use, while the number of activity opportunities around home has increased, in avoiding long-distance travels. Therefore, neighborhood street spaces, especially spaces for pedestrians, are becoming increasingly important [2]. In this context, street space is positioned as an important public good for improving the quality of life and well-being of citizens. To improve the sustainability of the city, the effective use of the street space as a unique public good that reflects people's sense of values and local context is expected; therefore, a method for reflecting diverse opinions and values of the residents and stakeholders in street design and its evaluation method are required.

1.2. Conceptual Framework

This section first presents a meta-design framework for improving people's happiness and pursuing hedonistic sustainability, while responding to the demands of the carbon-neutral and new-normal era. Then, a conceptual framework of AI and human co-operative evaluation (AIHCE) is developed to deploy the meta-design. In the future, particularly in Asian countries that are aging rapidly, the improvement of quality of mobility in response to outbreaks, including infectious disease pandemics, natural climate disasters, and the quality of everyday mobility is becoming a common issue. Under such circumstances, there is a need to shift from the conventional mass and fast transportation system to a safe, secure, resilient, and sustainable mobility system that includes people from various positions and values in society.

Foreseeing the upcoming need, the authors have conducted the Osaka University "Tran-Support" project and developed the JST-JICA Smart Transport Strategy for the Thailand 4.0 project as part of Science and Technology Research Partnership for Sustainable Development. Meta-design used here is an act called "designing the design process" that emphasizes value rationality and creates technical and social conditions to encourage broader participation, and is named in References [3,4]. Hedonistic sustainability is defined as a requirement that new local and public spaces should have under the New Normal regime. In addition, happiness is defined here as a subjective outcome that reflects a state in which physical, linguistic, and mental factors are intertwined as well-being.

Figure 1 illustrates the framework for evaluating happiness and well-being for street spaces in cities. Here, in the evaluation of well-being, two indexes, walkability and lingerability [5–7] for citizens, are used, and both are quantified based on the paired linguistic data and image data of street images. However, happiness is quantified based on facial expression data when walking on a street.

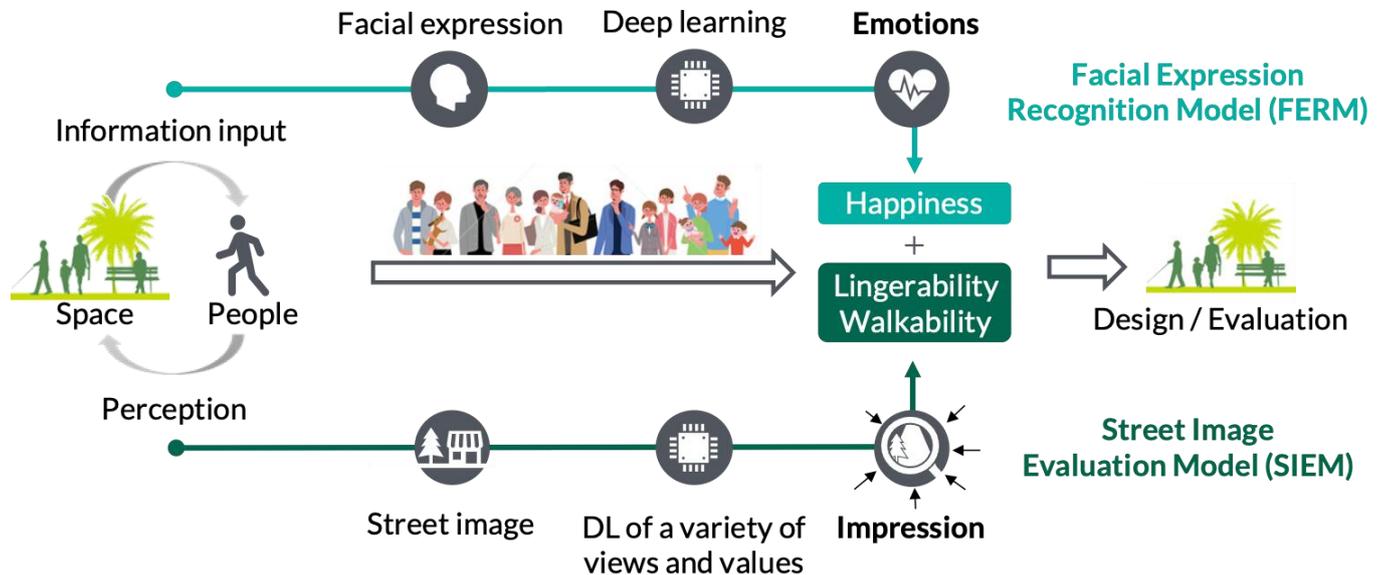


Figure 1. AI and human co-operative evaluation.

As shown in Figure 1, AIHCE consists of (a) the street image evaluation model (SIEM) that directly judges and evaluates the impression of the street from the data that is a set of the image of the street and its impression, and (b) the facial expression recognition model (FERM), which estimates emotions from the relationship between street images and facial expression data when walking on the street, and indirectly judges and evaluates the street. The former is a method of evaluating the walkability and lingerability of the street from the image of the street space, while the latter estimates emotions, such as happiness from the facial expressions of people walking in the street space, and evaluates the "comfort" of

the street. Evaluation of the street space from both sides enables continuous and objective evaluation of the value of the street and subjective emotional evaluation of the space.

This study signifies that the utilization of a new street space evaluation framework, AIHCE, that facilitates smooth communication between designers and stakeholders, based on human emotions and values, will change the concept process of street design and enable collaboration between AI and people and social co-creation. Furthermore, if AIHCE is extended and applied to include living spaces, streets, public plazas, and transportation hubs, it will be possible to contribute to the living quality of citizens and the pursuit of well-being by encouraging changes in the overall lifestyle in various scenes of life so as to enhance the livability in the new normal era (Figure 2).

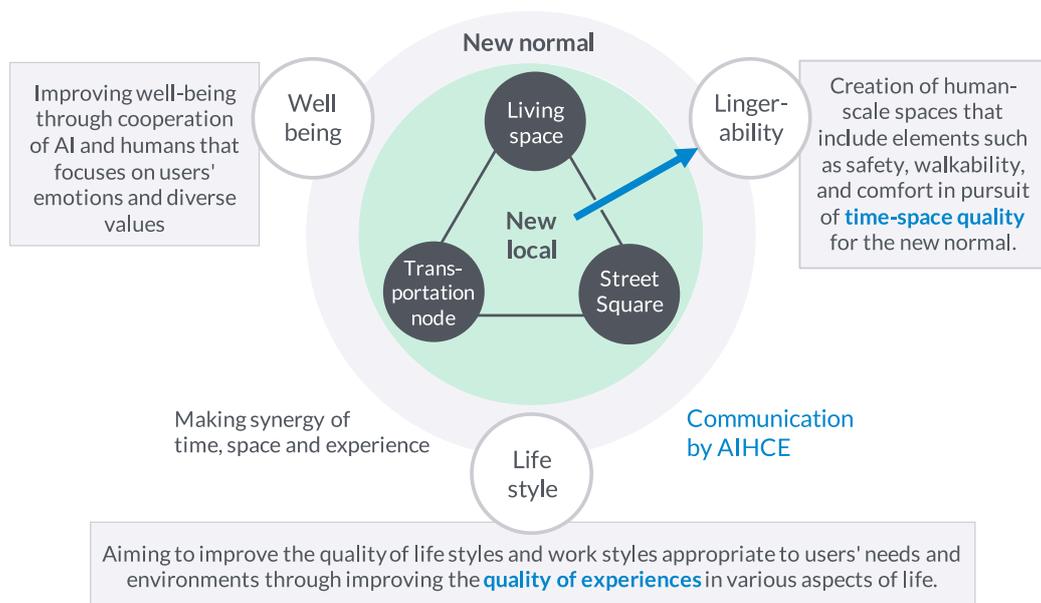


Figure 2. Public space design with AIHCE.

1.3. Objectives

This study aims to develop SIEM and FERM of AIHCE and demonstrate practical examples, albeit with limited conditions. In particular, among the public spaces mentioned above, they are applied to streets, public plazas, and transportation hub spaces (hereinafter collectively referred to as street spaces), and their usefulness is examined. To build a human-centered space that promotes hedonic well-being for users who use such street spaces, it is important to understand how spatial performance influences pedestrians' behaviors, facial expressions, emotions, and internal values.

In the development of SIEM, we examined how the evaluation results on walkability and lingerability by using AI constructed by the training data reflecting the opinions of the users who use the street directly evaluates the image of the street space is improved compared to the one that does not reflect it. Meanwhile, in the development of FERM, the relationship between pedestrians' facial expressions inferred by AI and the emotions estimated by interview surveys, hereinafter simply referred to as estimated emotions, is clarified. We further examined how this relationship is affected by spatial performance.

2. Literature Review

2.1. Co-Operative Design of Street Space

The conventional evaluation of street space began with the study of the level of service (LOS) [8] as one of the road performance standards in American Road Capacity Manuals in 1965, followed by space syntax theory by Hillier et al. [9] based on pedestrians' behaviors and street structures in the 1970s, and walking audit system by Davies and Clark [10] in

2009. However, these methods are problematic because of different viewpoints of designers and stakeholders.

However, the Pattern Language [11] was proposed by Alexander at the same time as the Space Syntax theory. Pattern language is the origin of co-creation design. This language tries to reflect the values of citizens in the design by using the common pattern of comfortable cities as a common language for communicating with citizens. However, there was a problem in that both the designers and the citizens were heavily burdened in its practice (Pattern Language practice).

In promoting future human-centered community development, it is indispensable to design streets based not only on the opinions of either designers or users/citizens but also on the common perspectives of both, such as by pattern language. A highly transparent and rational street space evaluation method easy for citizens to understand is desired. In other words, an evaluation method that reduces the burden on co-creation actors by efficiently connecting pattern information and linguistic information is required. This is what AI is most good at, and the significance of developing and utilizing AIHCE is explained in detail in the next section.

As shown in Figure 3, this study captures the performance of street space, which is a place for various activities, with three layers of legibility, walkability, and lingerability based on human perception and cognitive patterns. Lingerability is the time and spatial performance of the street for pedestrians to stay, and specific impressions of such a space include calmness, attachment, familiarity, and coziness. Walkability is a spatial performance to walk comfortably, and the impression of such a space includes moving safely, comfortably, and smoothly. In addition, legibility is the functional capability of using space and is in a state where it is easy to grasp the space.

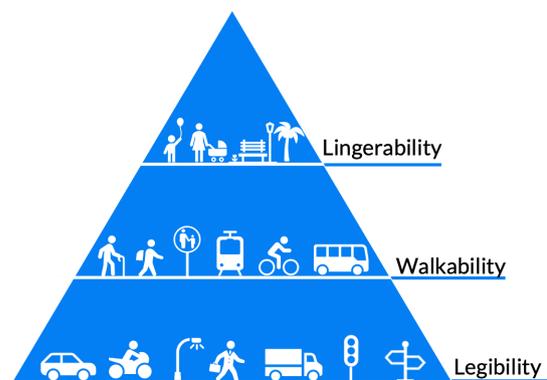


Figure 3. Understanding the hierarchy of street's spatial performance of lingerability, walkability, and legibility.

Figure 4 summarizes the concept of well-being when considering the correspondence between pedestrians and happiness.

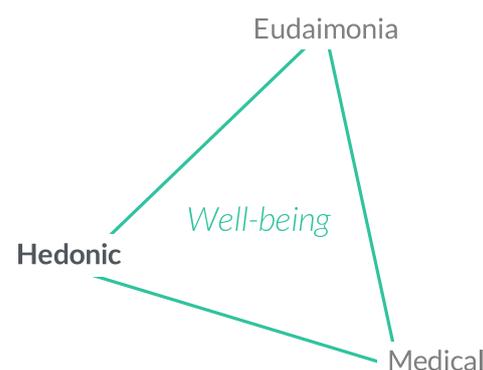


Figure 4. Pedestrian well-being.

Calvo et al. [12] classified well-being into three major categories: medical well-being, which refers to well-being where there is no dysfunction; hedonic well-being, which refers to an experience of positive emotions; and a third category that refers to well-being as a discovery of significance and potential. The medical approach has been studied for many years in the field of medicine to treat diseases, disorders, and illnesses. This approach is effective as a curative and preventive medicine but is not as effective in promoting well-being proactively. Therefore, a hedonic approach, as well as a eudaimonic approach, have been considered to promote well-being [12]. The former focuses on happiness and defines well-being in terms of pleasure attainment and pain avoidance [13]. Many studies on the hedonic approach have assessed subjective well-being (SWB) [14]. SWB consists of three components: life satisfaction, the presence of a positive mood, and the absence of negative mood. The latter defines well-being in terms of how fully a person functions and focuses on meaning and self-actualization [13]. Several existing studies have shown that well-being is considered a multidimensional construct that involves both hedonic and eudaimonic aspects [15,16].

Jan Gehl classifies people's activities in public spaces into three categories: necessary activities, voluntary activities, and social activities. Of these, necessary and voluntary activities are passive actions [17]. Actions included in passive behavior actions are comprised of looking at street settings and landscapes, looking at people who are active, walking around the promenade, and stopping to feel the city closer. Since these factors influence pedestrians' emotions in a short span that changes from moment to moment, the happiness felt by pedestrians is positioned as momentary hedonic well-being, which is short-term comfort, and, thus, can be measured by FERM.

Human beings can accumulate experiences and memories by perceiving and recognizing the interaction with the environment that occurs in the city [18], and experience linked to the momentary hedonic well-being in the street space is supposed to accumulate and lead to the construction of values based on individual memories. Thus, momentary hedonic well-being is expected to be measured by SIEM.

2.2. Evaluation Based on AI Image Processing

Recent advanced methods of pattern recognition by machine learning and deep learning attempt to extract and classify certain rules and meanings from data, such as a large number of images and sounds. The development of convolutional neural networks (CNNs) [19] has significantly improved the accuracy of pattern recognition and has been applied to various social issues and solutions. In this relation, we will refer to (a) a method of evaluating spatial performance from street space images and (b) a method of evaluating street performance through emotions from pedestrians' facial expressions using CNN.

With regard to the former, there has been a gradual increase in research that uses CNN to evaluate users' perceptions and impressions from spatial images. To assess the visual quality of urban air, Ye et al. [20] defined the key factors affecting the visual quality of six streets in the central ring area of Shanghai and performed image segmentation using street view images and SegNet, and then measured the ratio of each factor. Furthermore, the evaluation model was learned by ANN using images ranked by urban design experts as training data, and the visual quality of the entire city was evaluated.

Yin and Wan [21] provided a method for objectively measuring major street-level urban design functions related to walkability that were subjectively measured using image segmentation techniques. Li et al. [22] aimed to create an evaluation framework for walkability and proposed a physical walkability index using image segmentation technology for green, enclosure, and relative walking width among the typical elements. Liu et al. [23] focused on the visual quality of façades and the visual continuity of road walls and developed a machine learning method that automatically evaluates the visual environment of large cities by applying CNN to expertly evaluated datasets of images collected through Street View in Beijing. It is also evidenced that the machine learning algorithm can provide a good approximation of the visual experience of the general public.

Regarding computer vision techniques to quantify the perception of the urban environment, Dubey et al. [24] created a dataset using new crowdsourced data, including 110,988 images from 56 cities and 1.17 million pairs of image comparison data on six perceptual attributes, such as safety, liveliness, boredom, wealth, melancholy, and beauty. Therefore, it was possible to predict human judgment for a pair of image comparisons using a CNN that learns a combination of classification loss and rank loss.

Fan et al. [25] introduced a DCNN model and achieved a high accuracy in predicting the six human perceptual indicators in Chinese cities. Furthermore, a series of statistical analyses were conducted to identify the visual elements that could cause a place to be perceived differently. Haohao et al. [26] proposed a novel classification-then-regression strategy based on CNN and random forest to evaluate human perceptions of urban space. Meanwhile, multi-source data were employed to investigate the associations between human perceptions and the indicators of the built and socio-economic environment.

Seresinhe et al. [27] combined cloud source data generated from more than 200,000 images with the ability to extract hundreds of features from images using CNN-Places365, and then identified the configuration of a beautiful outdoor space. Meanwhile, there are other urban design approaches using CNN, and, among them, Yamada and Ono [28] developed an AI that estimates street names and willingness to visit cityscape images. It clarified the causal relationship, between the selection of spatial features recognized through vision and impressions of spaces, and applied it to the evaluation of urban design.

Among these studies, street evaluation [20–22] using image segmentation technology, which is an expression of perceptual information, focuses on the area ratio of physical elements in street space. It is applied to street evaluation by quantifying perceptual elements from physical elements based on the presence or absence of objects and facilities. However, studies [24–27] that quantify the perception of the urban environment focus on the continuous relationship of all information in the image, such as objects, facilities, and backgrounds, rather than the existence of individual objects. These studies are highly effective in street space evaluation by extracting specific patterns that give rise to emotions and impressions of the street space using CNN.

Therefore, in this study, we adopted a street space evaluation method using CNN for impressions and patterns rather than the presence-absence or abundance of objects. Thus, SIEM focuses on emotional factors, such as comfort, and collected image data for learning that reflect the opinions of the general public. Therefore, extracting patterns of human values and emotional elements from street images was possible by conducting a questionnaire survey of users and feedback their opinions to deep learning.

However, the conventional evaluation approach, with emphasis on psychological factors of street users obtained by the questionnaire survey, has unavoidable problems, such as being affected by the psychological burden and psychological state of respondents. Therefore, Fudamoto et al. [29] devised a new methodology that uses the smile rate of pedestrians in the cross-section of the street as an index of the comfort of the street space. Various algorithms have been developed for facial expression recognition using deep learning. Some of these methods aimed to accurately recognize facial expressions, even though the facial images included sudden illumination changes, for example, a video taken in an outdoor environment [30,31]. In this study, facial expression images were taken under a controlled environment, as explained later, and, therefore, did not require dedicated techniques for fitting to the data with unwanted noise.

In terms of street space evaluation using facial expression recognition, Noji and Kishimoto [32] used the “FaceAPI”, which is part of the Cognitive Service developed by Microsoft, to judge the percentage of pedestrians with smiling facial expressions on weekdays, Saturdays, and Sundays on each street in Shibuya. The spatial distribution of the smiling percentage is shown on the map, and the influential factors are identified. Compared with these previous studies, our FERM has an advantage as it enables continuous evaluation of space by capturing time-series facial expression changes focusing on

individual pedestrians and further mentions the relationship with spatial performance through the emotions of pedestrians.

3. Methods

3.1. Data Collection of Street Space Images

The training data used in SIEM and FERM is the image data frequently searched on the Internet regarding a keyword, collected by the web scraping method, which is an automatic image collection method on the Internet. It is assumed that the users' opinions are reflected in the search frequency; in other words, the web scraping methods bridge the language data and the image data based on people's (massive number of Internet users) thoughts or values.

In SIEM, web scraping is used to collect street images corresponding to lingerability and walkability as training data. The words (1) "cozy-street" and (2) "dirty-street" were adopted as the impression words corresponding to lingerability, and the words (3) "walkable-street" and (4) "unwalkable-street" were used to search for walkable street images. The image data were collected by scraping using the keywords from (1) to (4), and then 30 images for each keyword were collected deleting the noisy images. Additionally, another learning dataset was created to adjust the training data to the specific users' needs. A questionnaire survey was conducted with 24 college student examinees. The examinees watched the collected 30 images corresponding to (1) to (4) and were asked to choose the five most suitable images for each word. Based on the questionnaire survey, the top 20 images that the respondents chose most were selected as the training data reflecting the examinees' opinions.

In FERM, the words "human-happy-face" and "human-sad-face" were adopted as keywords, and 200 images of each facial expression were collected by web scraping as the training data. The noisy images were removed, and, finally, 50 images for each impression word were used as training data for FERM.

There does not exist any accessible dataset that includes street images labeled with human impressions, such as "cozy" and "walkable". Therefore, we prepared the dataset by web scraping. Regarding the dataset of facial expressions labeled with human emotions, we could use the existing one. However, to match the method of collecting the data in SIEM and FERM, we created the dataset by web-scraping.

3.2. Development of CNN Models

The collected dataset was divided into training data and test data, and the training data accounted for 60% of the collected data, while the test data was 40%. Using these data, a deep-learning model was constructed. Figures 5 and 6 show the training processes of SIEM and FERM. Figures 5 and 6 show the training processes of SIEM and FERM. In both models, the feature value extraction process in the center of the figures includes the convolution neural network (CNN) and max pooling processes. CNN extracts the features of the image, and max pooling is a method to increase the classification efficiency of neural networks by selecting the maximum value of the extracted features. In the fully connected layers on the right of the figures, the results obtained in the process are given as the selection probability (0–1) for each classification item. In Figures 5 and 6, the top number of each layer represents the processing size (length \times width \times channel), and the bottom number represents the output size after processing. In both methods, the ReLU was used as activation functions, and the mini-batch size was fixed at 32 for 30 epochs. In the inference phase, the model outputs a probability value from 0 to 1, which is determined according to how the input image has feature values identified in the convolution process. Then, RMSProp was used as the optimizer, and the learning rate was 0.00005. In the SIEM, the model accuracy was 82.8%, and the validation accuracy was 81.3%. In the FERM, the model accuracy was 98.5%, and the validation accuracy was 65.0%. Because the accuracy of the facial expression recognition model depends on the nature of the facial expression dataset used for training, there is no clear criterion for determining a reasonable accuracy

of the model. If we consider the fact that the accuracy of facial expression judgment for the general public, instead of specific actors, remains at a maximum of around 70% [33], it cannot be said that the FERF has a critical problem in accuracy.

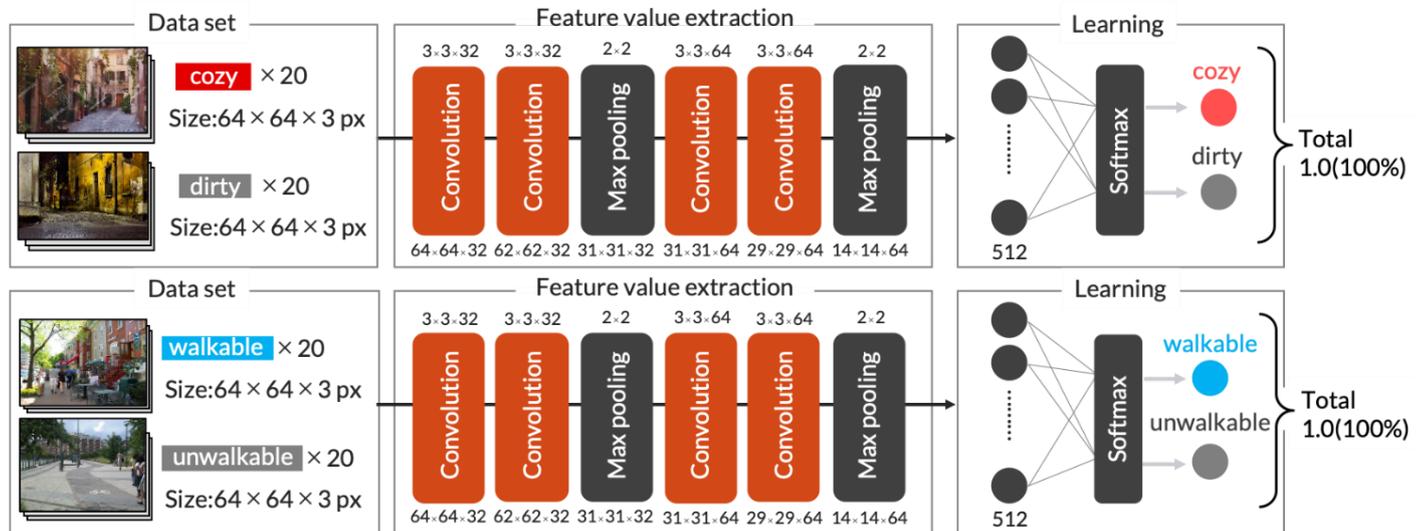
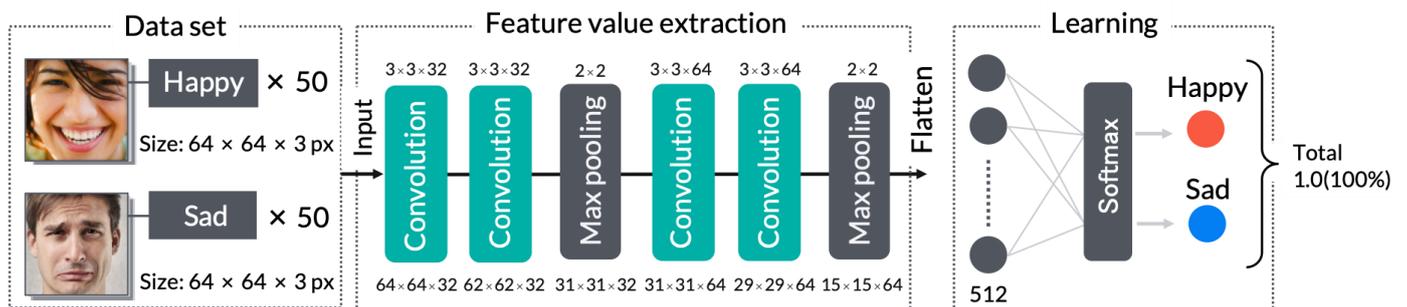


Figure 5. Street image evaluation model (with feedback).

Learning processing phase



Inference processing phase

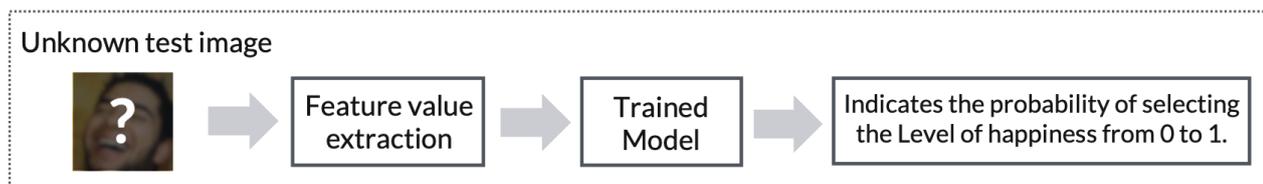


Figure 6. Facial expression recognition model.

In this study, the output probability is regarded as the lingerability and walkability of the street in SIEM and the degree of happiness in FERF. The closer the output value is to 1, the more comfortable and walkable the input street image is evaluated in SIEM, and the higher the level of happiness the input facial image is in FERF.

3.3. Experiment Using Street Image Evaluation Model (SIEM)

By using the two types of training datasets, SIEM was calibrated and applied to Higashimachi Street in Kobe City, Hyogo Prefecture, Japan, located near the city center railway station. Then, the results of SIEM with web-scraped training dataset and those with dataset including the questionnaire survey data were compared to examine the effectiveness of introducing the questionnaire survey.

First, the video of the street was taken at the eye level while walking along the street. Second, the video was divided into images by picking them every 1 s. Finally, the images were input to the SIEM, and the lingerability level score or walkability level score was calculated. To examine the differences in the results, gradient-weighted class activation mapping (Grad-CAM) [34], which visualizes the factors that contribute to the judgment results of images by CNN, was used. Grad-CAM is a method that uses gradient information in the final layer of convolution to calculate the influence of the feature map on the predicted labels. This allows us to visualize the regions of interest of the CNN model on the input image using a heat map. The elements affecting the level of coziness and walkability identified by Grad-CAM were examined by referring to previous studies on psychological well-being. Subsequently, the validity of the SIEM model was examined.

3.4. Experiment Using Facial Expression Recognition Model (FERM)

In FERM, two methods, direct and indirect, were used. The direct method uses a small video camera attached to an examinee walking on a street, and the video camera captures the videos of both the front scape and his/her facial expression. The indirect method was conducted indoors. An examinee watches a video of the streetscape at the eye level on the monitor, and his/her facial expression was captured by a video camera. These captured facial expressions picked up from the video every 0.1 s were input into FERM, and the probability, or happiness level of the moment walking on a street, was calculated. The comparison of the two methods showed that the results of the direct methods were strongly influenced by environmental factors, such as lighting. Therefore, an indirect method was adopted in this study. The 17 college students who participated in the experiment were asked to watch short videos of five streets: St1: a busy shopping street, St2: a street under the overpass with few people, St3: a lively main street, St4: a street in a lush park, and St5: a street in a commercial district (Figure 7). In the first 5 s after the start of the video viewing, the facial expressions were assumed to be influenced by the condition immediately before viewing; therefore, the facial expressions in that part were not used for the evaluation.



Figure 7. 5 spatial images of selected streets: St1: a busy shopping street (Link: <https://youtu.be/BDW9P60NgHs>) (accessed on 28 July 2021), St2: a street under the overpass with few people (Link: <https://youtu.be/1xQMolBXgsE>) (accessed on 28 July 2021), St3: a lively main street (Link: https://youtu.be/uKNMr_FD048) (accessed on 28 July 2021), St4: a street in a lush park (Link: <https://youtu.be/C6QaVf2oY3A>) (accessed on 28 July 2021), and St5: a street in a commercial district (Link: <https://youtu.be/SUJdXXH88EQ>) (accessed on 28 July 2021).

The analysis revealed that seven samples had a constant happiness level of 0 (zero). The cause of this was visually analyzed using Grad-CAM, and it showed that the influence of hairstyles and clothing hiding part of the face (or facial contour) was more dominant than facial expression in judging the level of happiness. Therefore, the samples were eliminated from the analysis. To visualize the characteristics of the street, the following procedures were performed: The results of each examinee for each street scene were divided into 5-s intervals. The happiness levels of every 0.1 s calculated as the probability was cumulated by the 0.25-happiness-level intervals. Furthermore, every examinee's cumulative data were averaged for 5-s.

Additionally, an eye-tracker device (Tobii Eye Tracker 4C, Tobii Technology K.K., Tokyo, Japan) tracking what the examinees see was attached to the monitor the examinees looked at. The validity of the FERM model was examined by comparing the level of happiness calculated by FERM with the points that the examinees saw.

3.5. Supplemental Questionnaire Survey to Elaborate the FERM

The relationship between the happiness level by facial expressions and respondents' opinions by the questionnaire survey was examined. Russell [35] explained the emotion by two axes: the "aroused-unaroused" (vertical) axis and the "pleasant-unpleasant" (horizontal) axis. By combining the two axes, he classified the emotion into eight categories: lively, exciting (both located in arousal and pleasant area), irritating, anxious (arousal and unpleasant area), calm, peaceful (unaroused and pleasant area), boring, and tiring (unaroused and unpleasant area), as shown in Figure 8.

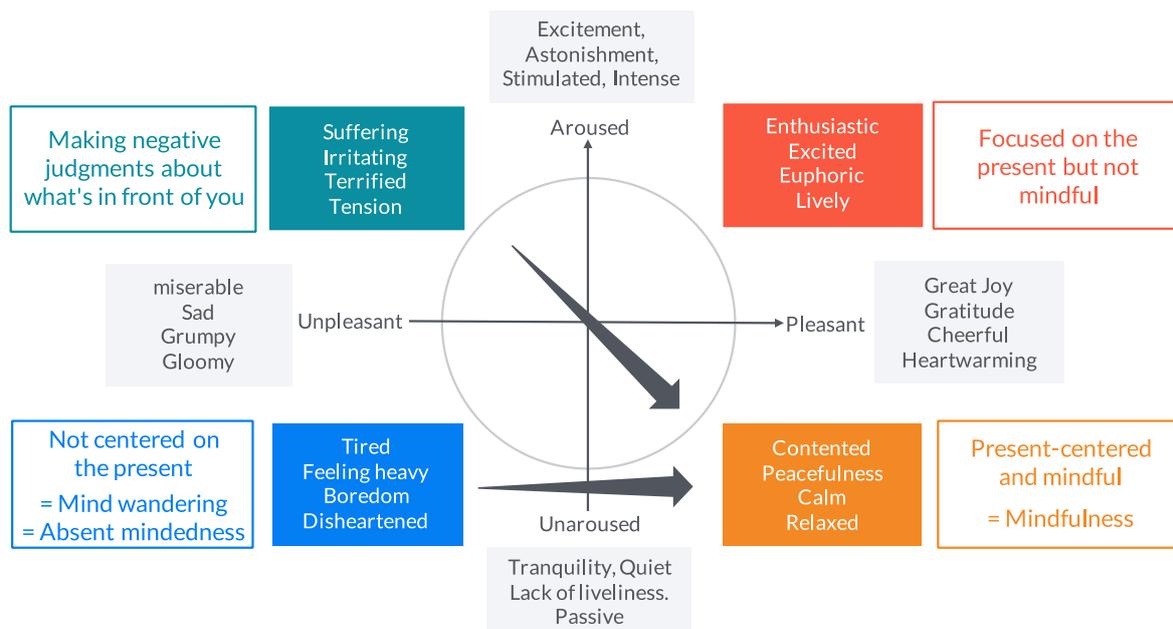


Figure 8. Two-dimensional emotion model.

The respondents were asked about their impressions of the video immediately after they watched it. They rated the degrees of the respective eight emotions and the degrees of walkability and lingerability on a six-point scale (1 = strongly disagree to 6 = strongly agree).

With regard to the relationship between hedonic well-being and human emotional fluctuations, Christie et al. [36] analyzed the relationship and confirmed that mindfulness, expressed as both "conscious awareness" and "non-judgment", had a significant indirect effect on hedonic well-being. Rowland et al. [37] found momentary mindfulness to be positively associated with low arousal positive affect inertia, a lower switching propensity to negative affect, and less instability.

Fredrickson et al. [38] listed joy and contentment as specific types of positive emotions. The Positive and Negative Affect Schedule [39], which is the most widely and frequently used scale for assessing positive and negative emotions, includes the following measures of positive emotions: excited and enthusiastic. Thus, emotions in the "pleasant" areas can be regarded as positive. Previous studies [36–39] have shown that momentary mindfulness corresponds to "unaroused and pleasant" areas of the affective model.

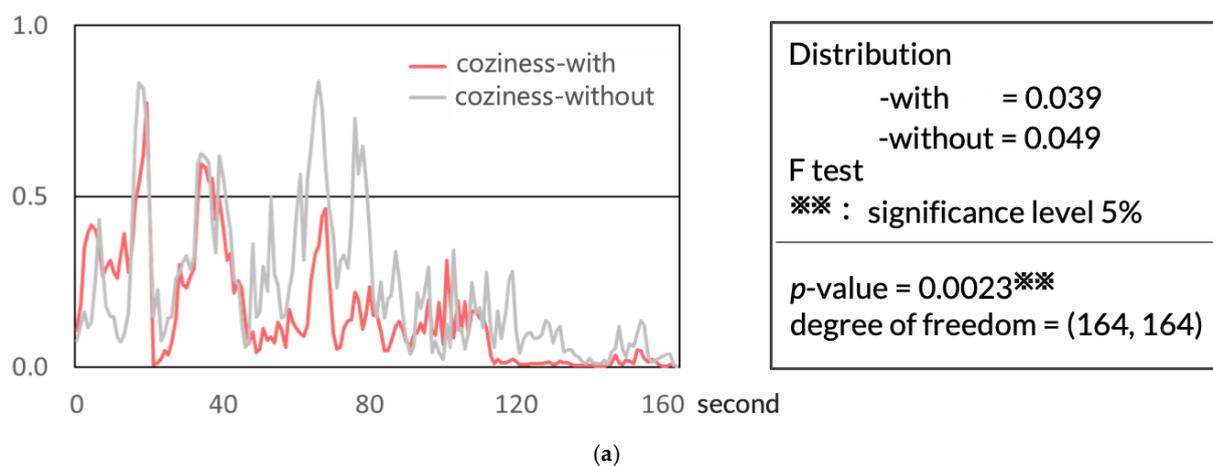
Therefore, in our study, the emotion of calm and peaceful ("unaroused and pleasant" areas) is considered to improve hedonic well-being, and the arrows in Figure 8 indicate the direction of better hedonic well-being. In addition, the results of the FERM, questionnaire survey, and streetscape on the screen were compared.

4. Results

4.1. Model Behavior Based on User Feedback

Here, the application results of the image judgment method are presented first. Figure 9 shows the results of the continuous evaluation of the coziness and walkability along the target street, including information on the variance of the evaluation values and the results of the F-test. “Coziness-with” and “walkability-with” are the evaluation results of coziness and walkability by CNN using training data reflecting the questionnaire results, respectively; “coziness-without” and “walkability-without” correspond to the evaluation result using the initial training data only, i.e., without any feedbacks of questionnaire results. The model without the questionnaire reflects the opinions of unspecified internet users, while that with the questionnaire considers specific respondents’, as well. The CNN was evaluated three times, and the average value was used as the final evaluation value. Regarding coziness, the model reflecting the results of the questionnaire survey showed a smaller variance in the evaluation values, and more stable results were obtained.

Level of coziness



Level of walkability

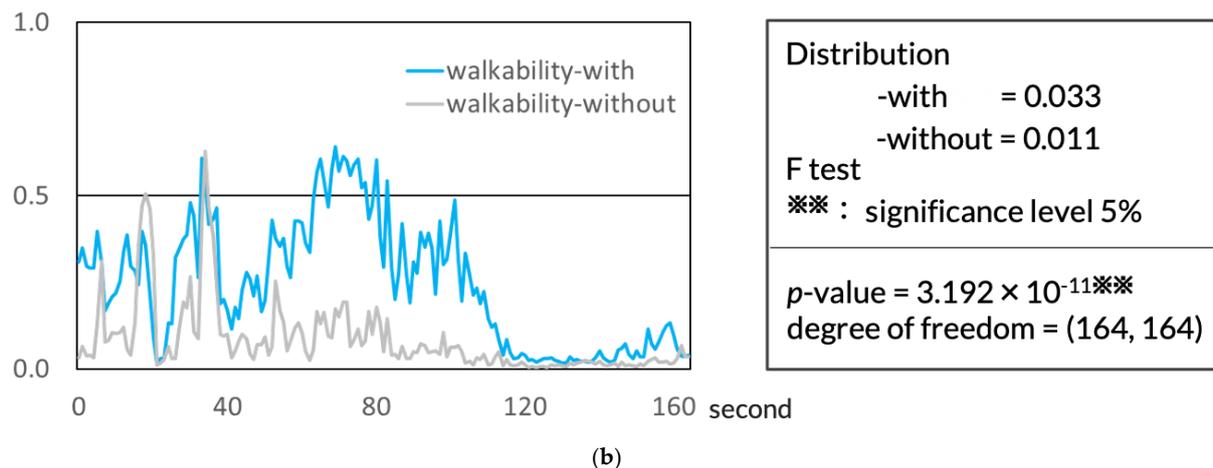


Figure 9. Effect on evaluation results based on the with/without questionnaire comparison, (a) level of coziness, (b) level of walkability, (c) spatial images of Higashimachi Street obtained from the video from 40 s to 120 s with the significant difference between the results with and without feedback.

Comparing Figure 9a,b, it can be seen that the former shows a strong positive correlation between the with and without cases, while the latter does not. The difference seems to be caused by a large gap between 50 and 110 s, as shown in Figure 9b. In the street image, during this time, black guardrails were seen along the sidewalk, and the parking space for motorcycles separated the sidewalk from the roadway. Grad-CAM identified that the surface conditions of the sidewalks increased the level of walkability (Figure 10b). From this result, it can be inferred that the SIEM model evaluates clearly separated sidewalks as walkable.

Figure 9a shows that the level of coziness in the case without the questionnaire was higher than that of the case with the questionnaire. A large gap can be observed between 50 and 80 s. During this period, a wall and plants were present along the left side of the sidewalk, and they increased the coziness in the case without the questionnaire; this increase was less in the case with the questionnaire. This could be because the respondents of the questionnaire tended to evaluate street images with stereoscopic depth as cozier streets; thus, street images with trees and plants had a relatively strong impact on coziness. It was observed that the coziness values were mostly higher in the without-case than in the with-case.

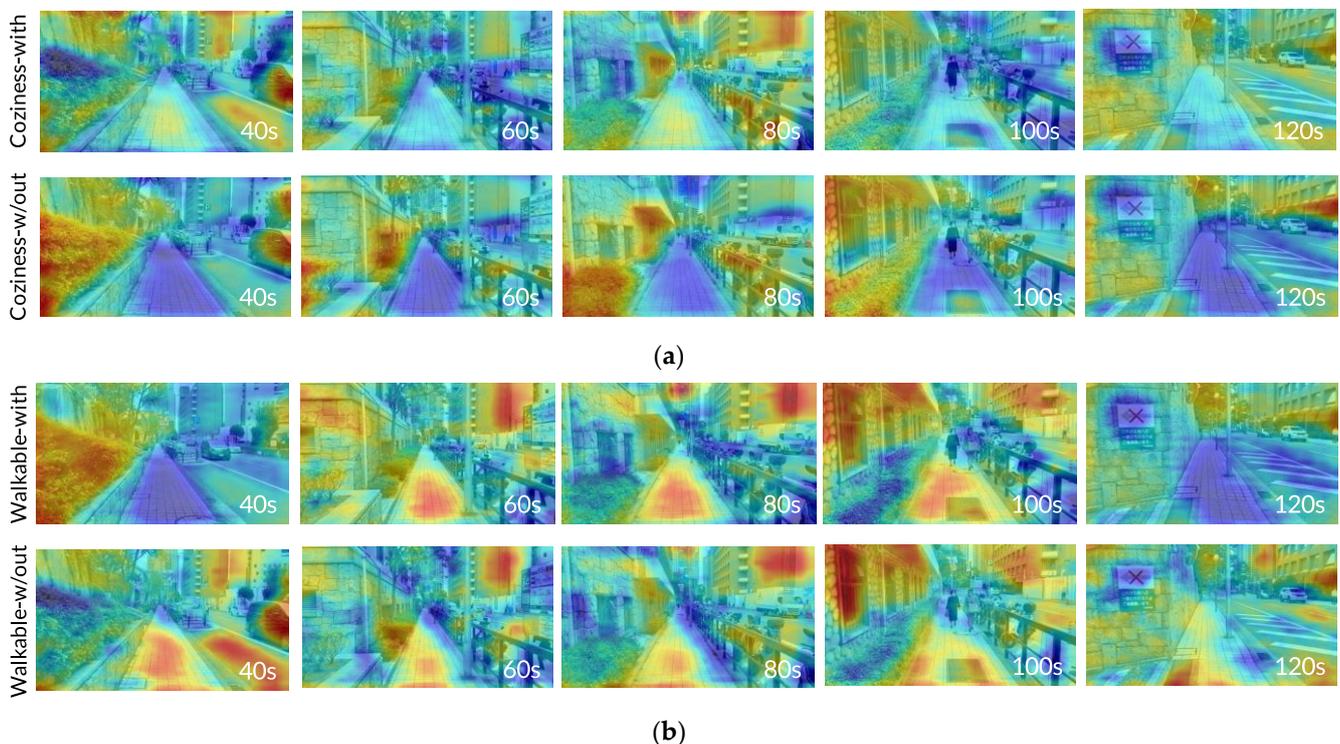


Figure 10. Comparison of coziness and walkable, -with and -without, based on Grad-CAM, (a) comparison of coziness-with and -without based on Grad-CAM from 40 s to 120 s corresponding to Figure 9c, (b) comparison of walkable-with and -without based on Grad-CAM from 40 s to 120 s corresponding to Figure 9c.

Figure 11 shows the analysis results of the factors that influence the evaluation of street space using Grad-CAM, focusing on two-time sections: (a) 19 s with the highest evaluation value on the target street and (b) 21 s with a sharp drop in the evaluation value from there. In the 19 s image, it can be seen that the yellow furniture is the criterion for judging the coziness of AI. However, in the 21 s image, although AI evaluated the coziness as low, it was found that the yellow display board in the image had a reaction. To analyze the effect of such color (yellow) elements on the evaluation results, three patterns of yellow in the 19 s and 21 s images, three alternative colors (yellow, red, and light blue) were provided to the furniture and display board and compared. The evaluation results of the level of coziness were 0.775 at 9 s, 0.007 at 21 s in the case of yellow color, 0.631 at 19 s, 0.008 at 21 s

4.2. Evaluation of Streets Space Performance Based on Estimated Facial Expressions and Judged Emotions

To simply illustrate how the FERM is applied to evaluate street space performance, the analytical result of a respondent is given in Figure 13. A time-series graph in the central part of Figure 13 shows the temporal development of facial expressions while watching the videos. The horizontal axis shows the time in seconds, and the vertical axis displays the level of happiness. The surrounding pictures present how the respondents' facial expressions and street scenery change over time. The sceneries of Street No.2 (St2) and Street No.4 (St4) are given. When focusing on the results of one respondent, St4 has a stable and high level of happiness, while St2 shows a relatively low level of happiness until the end of the video.

Next, we discuss the results of a street evaluation performed by 10 respondents. Using the results of St4 illustrated in Figure 14, the evaluation results of the other streets are discussed without graphs. The happiness levels of the 10 respondents are shown in Figure 14. The lines in the figure are colored to distinguish the examinees. The time series (horizontal axis) was sampled at 5-s intervals, and the happiness level (vertical axis) was sampled at 0.25 units. The ratios of respondents distributed in each level of happiness are presented in Table 1, showing the changes in respondents' happiness toward St 4 measured every 5 s.

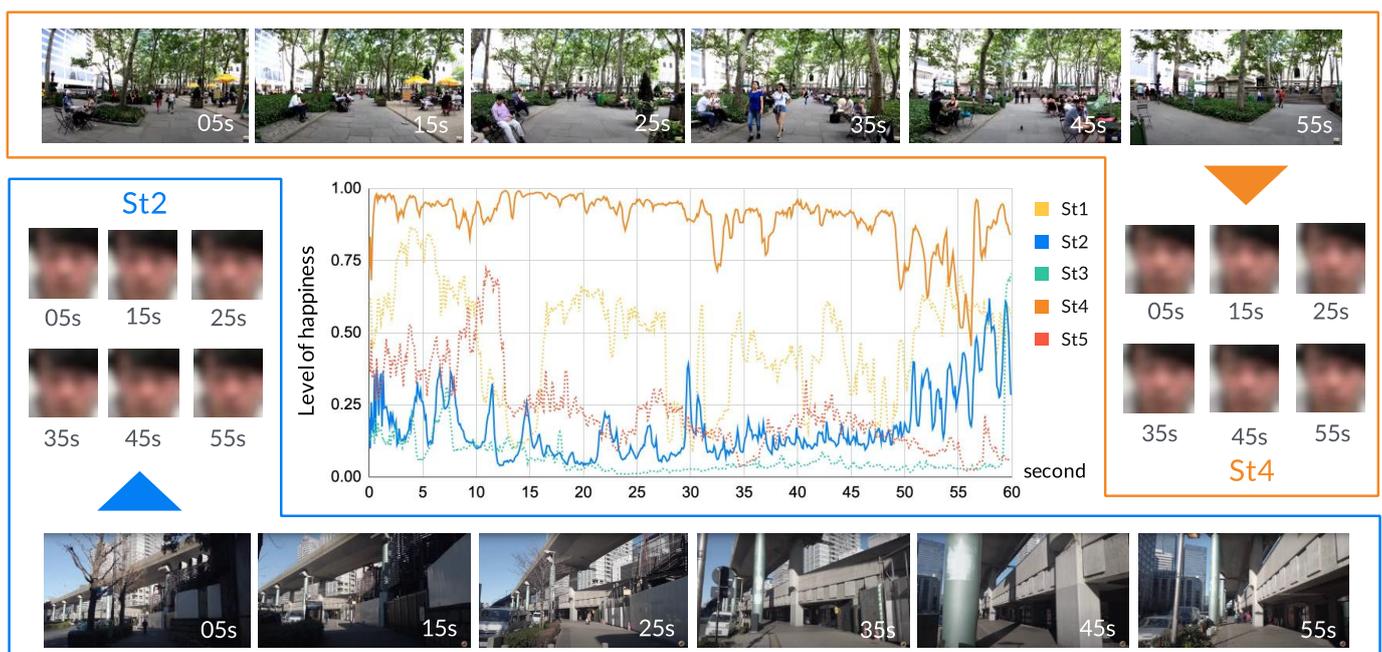


Figure 13. Spatio-temporal variation of estimated facial expressions/happiness, St2: (Link: <https://youtu.be/1xQMolBXgsE>) (accessed on 28 July 2021), St4: (Link: <https://youtu.be/C6QaVf2oY3A>) (accessed on 28 July 2021).

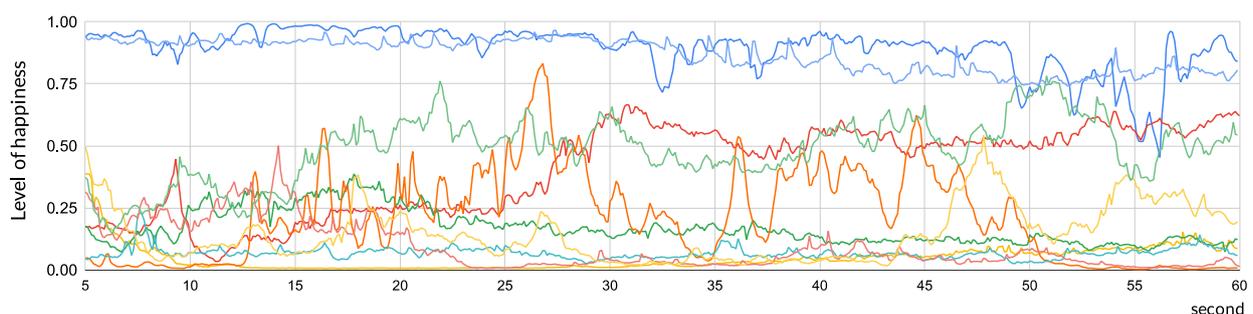


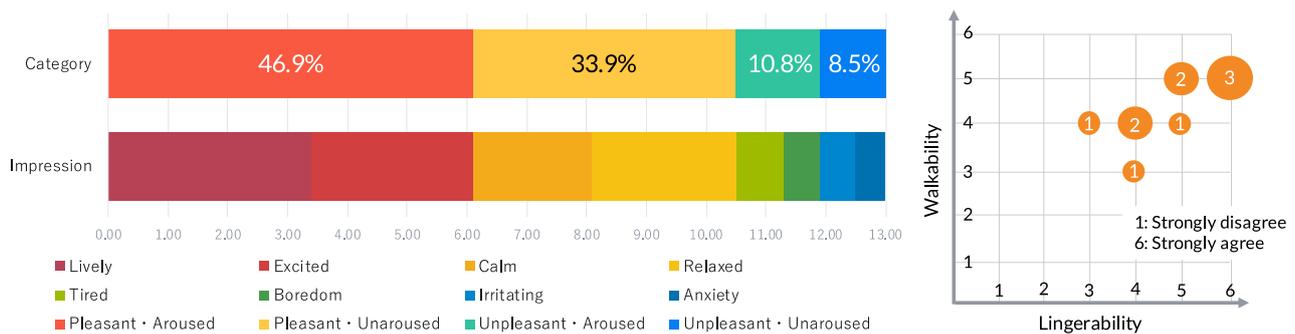
Figure 14. Street evaluation focusing on all respondents for St4.

Table 1. Transition pattern of happiness level of facial expression for St4.

Level of Happiness	5–10 s	10–15 s	15–20 s	20–25 s	25–30 s	30–35 s	35–40 s	40–45 s	45–50 s	50–55 s	55–60 s
0.75–1.00	22.2%	22.2%	22.2%	22.4%	23.6%	20.9%	22.2%	22.2%	20.0%	16.9%	18.9%
0.50–0.75	0.0%	0.2%	7.1%	9.6%	13.6%	16.0%	6.4%	20.0%	19.8%	24.7%	19.8%
0.25–0.50	15.6%	22.4%	22.2%	15.6%	17.3%	9.6%	22.7%	11.6%	18.0%	5.8%	12.4%
0.00–0.25	73.3%	66.2%	59.6%	63.6%	56.7%	64.7%	59.8%	57.3%	53.3%	63.8%	60.0%

The survey results obtained from the questionnaires are shown in Figure 15. The left-hand side bar chart in Figure 15 represents the results of the questionnaire survey on respondents' feelings immediately after watching the video. The respondents chose the closest statement to their feelings among the eight kinds of feelings defined by Russell [35]. The right-hand side of Figure 15 shows the results of the questionnaire survey on street impressions. The respondents answered questions about the degree of walkability and lingerability of each street image they watched in the video. The numbers in the circles denote the number of respondents corresponding to each pair of feelings.

For the evaluation results of St1, the transition pattern of the level of happiness became more scattered. In addition, the satisfaction of lingerability and walkability was relatively lower than that of the other streets. For the emotions stated, the arousal was about 67%, while pleasant and arousal accounted for 40%. It is suggested that a higher level of happiness is caused by arousal, especially for pleasant and arousal emotions.

**Figure 15.** Results of subjective evaluation of street space for St4.

For St2, the transition pattern of the happiness level gradually changed towards zero as the video progresses. Compared to other streets, the overall performance of happiness was relatively low. The survey results showed that walkability was relatively higher and lingerability was lower than that of other streets. In addition, the emotions of “unpleasant and unaroused” and “pleasant and unaroused” accounted for 35% and 30%, respectively. The high proportion of these two responses shows that depression (low arousal) emotion plays an important role in influencing respondents. Moreover, unpleasant feelings accounted for more than pleasant feelings. The results of St3 had a similar tendency as St2, while the transition pattern of happiness level of St3 decreased more slowly. The survey results of St3 showed high performance of both walkability and lingerability. The “pleasant and aroused” and “pleasant and unaroused” comprised the most (about 38% for both emotions), and unpleasant and arousal comprised the least. Comparing the estimated results of facial expression recognition and emotions surveyed, the gradual decrease in the level of happiness may be because of the high tendency shown in both “pleasant and aroused” and “pleasant and unaroused”.

St4 had a higher level of happiness than the other streets. The survey results demonstrated a lower level of walkability and better lingerability performance. “Pleasant and aroused” accounted for the majority of emotions (46.9%), followed by “pleasant and unaroused” (33.9%). “Unpleasant and unaroused” was the least common. In Table 1, 60% of the respondents represented a low level of happiness less than 0.25, while only 18% of the estimated emotions were “unpleasant”. The results seem contradictory; therefore, the

evaluation results of FERM with respect to the respondents who felt “unpleasant” were checked. To analyze the inconsistency between the FERM results and the questionnaire replies, we conducted a correlation analysis between the time-averaged happiness values by FERM and the difference in agreement levels to “pleasant” and “unpleasant” and found a relatively high correlation coefficient of 0.701. This result indicates that the FERM evaluation is consistent to some extent with the subjective evaluation of the respondents.

St5 had a relatively poor performance in the level of happiness. Compared to St2, the level of happiness decreased more sharply as the video time passes. With high performance of lingerability, “pleasant and unaroused” accounted for the most emotion (42.7%), followed by pleasant and arousal (41.0%). The pleasant was the most expressed emotion in St5 compared to the other streets. Combining the results obtained from the questionnaire survey and facial expression recognition, the higher level of pleasant emotions, especially “pleasant and unaroused”, resulted in a lower level of happiness.

In conclusion, the evaluation results of street spaces can be classified into the following three categories: First, for street spaces, such as St1, respondents’ arousal and unpleasant emotions result in great fluctuations in the transition pattern of happiness levels. Without a relaxing atmosphere, the coziness in such a street space is considered disappointing. Second, owing to the arousal and pleasant emotion of respondents in spaces, such as St4, the happiness level keeps a relatively stable and high state. Street spaces are considered lively. Because respondents state the unaroused and pleasant emotions more than their arousal and pleasant feelings, such street spaces are positioned in the middle of all street spaces in terms of lingerability. Finally, for the last category, including St2, St3, and St5, respondents expressed their deposition (low arousal) towards these street spaces. These streets show a similar tendency in that the happiness level decreases gradually as time passes. It is found that, if the arousal and pleasant emotions are as high as their deposition, the decrease in respondents’ happiness may be mitigated. St2 and St5 are similar in their rapid decrease in happiness levels over time. To understand the reasons for the rapid reduction in happiness, for example, boredom or relaxing feelings toward street spaces, it is necessary to develop an evaluation method that combines the FERM.

4.3. Toward the Integrated Evaluation of SIEM and FERM

This section explores the consistency and differences of SIEM and FERM evaluations in an attempt to propose an integrated design process using them. Figure 16 shows the average evaluation results of the 10 respondents by FERM, while Figure 17 shows the results of the respective streets by SIEM. Regarding coziness, the model reflecting the questionnaire results in Figure 9 showed more stable results, while walkability showed a large temporal variance. Based on this result, this section focuses on the level of coziness for each street. The time-averaged evaluation values from 5 s to 60 s are shown in Table 2, which also includes the average values of lingerability, pleasant, unpleasant, pleasant/unaroused, and pleasant/aroused of all the respondents who answered the questionnaire.

Table 2. Comparison of model evaluation results and questionnaire survey results for each street.

Evaluation Method	Indicator	St1	St2	St3	St4	St5
Models	Level of coziness by SIEM	0.16	0.13	0.58	0.36	0.18
	Level of happiness by FERM	0.41	0.17	0.29	0.34	0.11
Questionnaire (Degree of agreement to stated emotions)	“Lingerable”	2.8	3.3	5.0	4.8	4.8
	“Pleasant”	2.6	1.8	3.5	3.9	4.1
	“Unpleasant”	2.9	2.1	1.1	0.9	0.8
	“Pleasant and unaroused”	0.8	2.3	3.5	3.3	4.2
	“Pleasant and aroused”	4.3	1.3	3.5	4.6	4.0

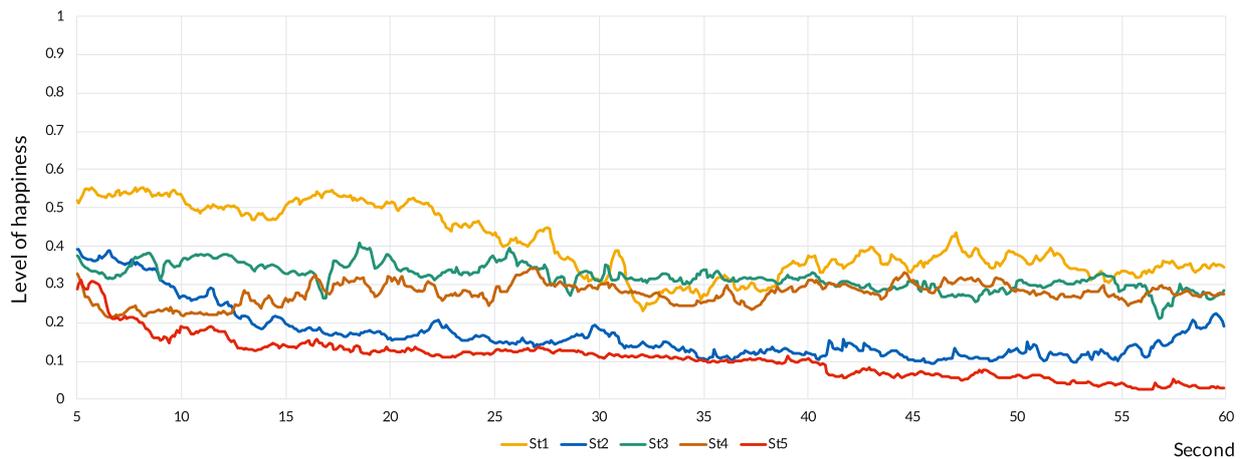


Figure 16. Average value of happiness level of respective streets by FERM.

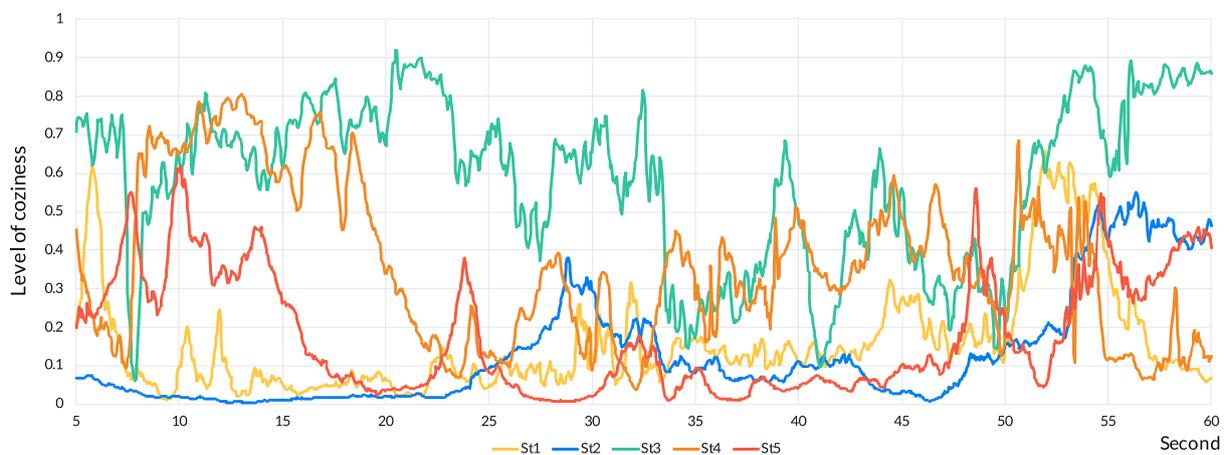


Figure 17. Level of coziness of respective streets by SIEM.

When the level of coziness was evaluated for each street by SIEM, the time-averaged values of coziness for all sections of St1, St2, and St5 were generally low, and the variance was small. The results suggest that these three streets are less cozy and less volatile in the SIEM evaluation. In contrast, the coziness of St3 and St4 showed both a higher time-average and greater variance in all sections than other streets. Therefore, while these two streets are considered relatively comfortable, it should be noted that the evaluation results may vary considerably across sections.

From the comparison of Figures 16 and 17, it was observed that the transition of the scores by FERM was more stable or had smaller variation. As shown in Table 2, the coziness value by SIEM was small in St1, whereas the happiness value by FERM was rather large. For St3, the happiness value was larger than the coziness value. For St4, the values of both SIEM and FERM represented almost the same values, and both were in a large value group among the streets. St2 and St5 showed comparatively low levels of happiness and coziness. When comparing the time-averaged values of happiness and coziness for each street, the difference between the two values is larger in St1 and St3 than in St2, St4, and St5.

Furthermore, we integrated the collating model evaluation results and questionnaire results, and then we evaluated each street. St1 showed a low level of coziness and a low degree of agreement to “lingerable” and “pleasant and unaroused” because St1 is crowded. However, St1’s degree of agreement with “pleasant” is almost the same as that to “unpleasant.” The level of agreement with “pleasant and aroused” is much higher than that to “pleasant and unaroused”, since the pedestrian crowd could impress the examinees to be lively and excited. The influence of the aroused situation on pleasantness varies from person to person, whereas that on happiness was indicated to be positive by the evaluation

result of FERM. These results show that St1 enhances users' hedonic well-being to some extent, although it is not cozy. St2 has low levels of coziness and happiness, and shows a low degree of agreement to "lingerable," "pleasant", "pleasant and unaroused", and "pleasant and unaroused".

Since St3 and St4 have high levels of coziness and happiness and high degrees of the agreement to "lingerable", "pleasant", and "pleasant and unaroused", they are considered the streets that enhance the well-being of users. Meanwhile, St5 is seen as a "lingerable" and "pleasant" street based on the questionnaire results, while it was evaluated to be low in terms of coziness and happiness by the models. Regarding coziness, St1 and St2 showed low values due to the lack of street depth visibility because of the pedestrian crowd in St1 and the overpass in St2. St5 seems to be clear enough to show street depth; however, its video image includes a nearby roadside environment that sometimes influences the depth visibility. Fluctuations in the examinees' perspectives between the distant and near views likely lowered the level of coziness.

In addition, while the level of happiness is low, the degree to agreement to "pleasant" is high. Thus, there is a contradiction between the evaluation results of the AIHCE (SIEM and FERM) and the results of the questionnaire for St5.

For St4, where happiness and coziness were comparatively higher than those of other streets and showed less variance, the differences in the temporal variation patterns of happiness and coziness are as shown in Figure 18. As shown in the figure, the level of coziness is higher than that of happiness in the time period from 8 s to 20 s. The factors affecting the level of coziness were identified and highlighted in red and yellow colors by Grad-CAM in the upper images in Figure 19, while the points respondents focused on or looked at were visualized using the eye tracker in the lower images in Figure 19.

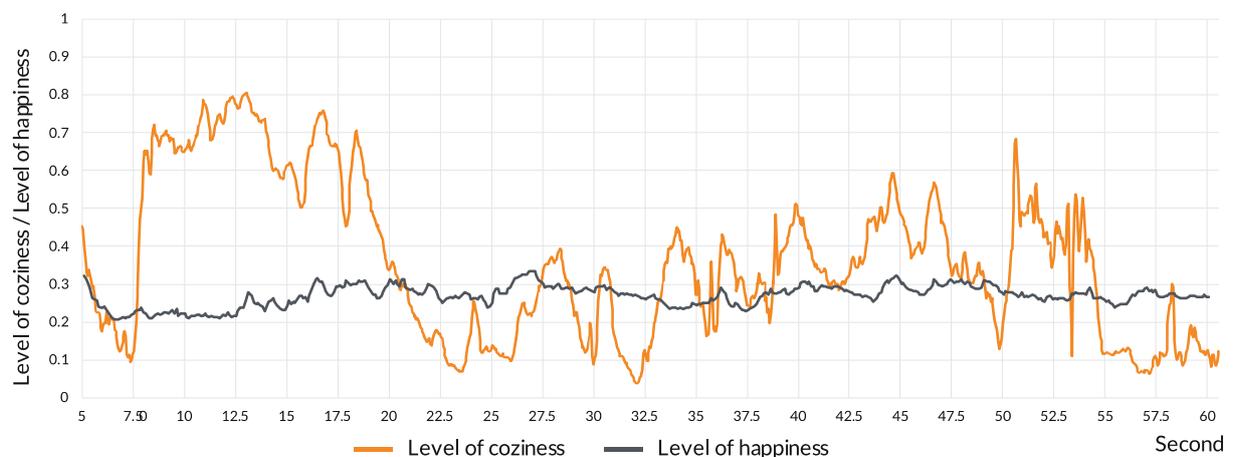


Figure 18. Comparison of time variability between SIEM and FERM for St4.

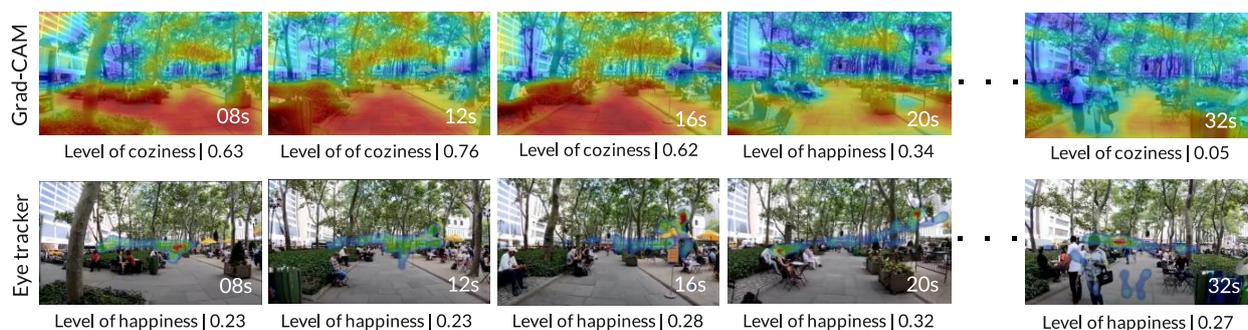


Figure 19. Comparison of AI and human viewpoints using Grad-CAM and eye tracker from 8 s to 20 s with the larger value of coziness than that of happiness and 32 s with the lower coziness value, (Link: <https://youtu.be/C6QaVf2oY3A>) (accessed on 28 July 2021).

The visualized results by Grad-CAM suggest that the SIEM seems to focus on the road surface, benches along the road, resting people sitting on the bench, and trees covering the road, which may increase the level of coziness on St4. However, at 32 s, two pedestrians standing on the road hid the resting people and the benches, which might have decreased the level of coziness. Meanwhile, the results of the eye-tracker showed that the respondents mainly looked far ahead up the street rather than nearby scenery, both in 8 to 20 s and at 32 s, resulting in a relatively high level of happiness.

Additionally, the viewpoints of the AI and the respondents were completely different in this experiment on St4. Therefore, it is desirable to use both SIEM and FERM in a complementary and integrated manner, rather than using them alone. Their evaluation results enabled us to develop alternative street designs. For example, on St4, it is effective to improve the visibility of street depths and build a space for the rest far ahead of the street. Another idea is to attract attention to the surrounding resting space by widening the road and not hiding the space.

4.4. Validation under Limited Data Conditions

At present, the validity of the AIHCE evaluation results was examined by collating SIEM with the analysis results by Grad-CAM and FERM with the eye-tracking results (by Eye-Tracker). As a result, it was confirmed that the results of both models constructed based on a relatively small amount of training dataset were generally in agreement with human perception. For the SIEM training data, only a simple screening was performed to remove noise images from web scraping images. In FERM, we conducted a video viewing experiment in a controlled room so that the facial expressions of the subjects were not affected by the external environment other than the visual and auditory information of the target street. Consequently, the training datasets of SIEM were limited to 80 images (40 images for both walkability and lingerability, respectively) through the questionnaire survey, and those of FERM were 100 images. In the future, in addition to using existing datasets, we would like to add data that is expected to be acquired along with the operation of AIHCE, as well as to quantitatively enhance the construction of our own datasets.

5. Conclusions

Within the meta-design for the pursuit of happiness and hedonistic sustainability in urban space in the new normal era, this study proposed an AIHCE framework for street space evaluation focusing on happiness and well-being in the context of a new local. The AIHCE consists of a FERM and SIEM. The data preparation method of the SIEM includes two phases: (a) deploying web scraping to acquire training data and (b) narrowing down the data sample based on the results of the questionnaire survey. We obtained stable results of the evaluation of the street's coziness performed by this two-phase process of data collection. It is revealed that training data containing images of warm colors, such as yellow or red, result in better scores for the coziness of street spaces, meaning that the better-evaluated results showed in testing data contained warm colors with the same tint.

For the FERM, the same web scraping method was used to acquire the training data. With the testing data collected from respondents' facial expressions when watching the videos, in which respondents were asked to imagine themselves walking in the streets with the real-world sceneries provided, a machine learning model was constructed to evaluate their happiness levels. It is confirmed that there is a certain degree of correspondence between the happiness level estimated by the facial expressions and the stated emotions learned from the questionnaire survey. This result suggests the effectiveness of a continuous evaluation of street coziness based on the subject evaluation of pedestrians. Furthermore, as shown in Figure 20, FERM and SIEM can be operated in a complementary and integrated manner, rather than using them alone. Since previous studies have often short-circuited the process of human perception and evaluation of urban space, and simply linked spatial images and user impressions, they have a limitation in their application to practical street design. AIHCE, by contrast, is an integrated framework of FERM and SIEM, and has a

novelty in that it considers both embodiment and verbalization of human emotions toward urban spaces. It is also advantageous in that it enables a wider range of street designs in a simple procedure. This will contribute to promoting communication design between practitioners and stakeholders.

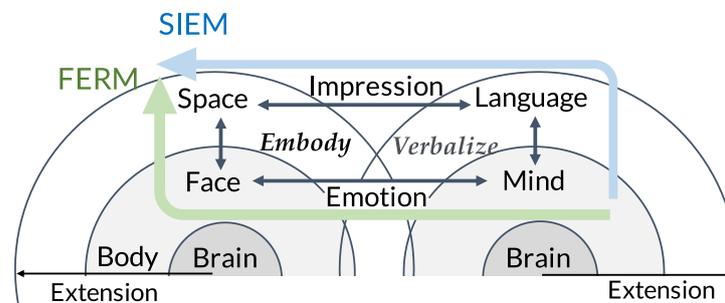


Figure 20. Compound viewpoints and evaluation process of AIHCE.

The methods developed in this study have several limitations. First, the SIEM training data of “cozy-street” images were collected on the web for a certain period, while the happiness level was determined by the FERM when respondents’ facial expressions were detected. Accordingly, the authors have carefully thought about the fact that, if the accumulation of momentary happiness leads to “coziness”, which becomes part of the well-being measured, then there remains a need for future research on this topic. Second, the dataset acquired by web scraping is limited in this study, especially for images of “cozy-street” and “walkable-street”. However, privacy protection issues can arise in the data collection of human facial expressions. Future studies can explore privacy-preserving methods to cope with these problems. Third, the results were based on a two-dimensional evaluation, taking visual and auditory information into account. Future research could further explore changes in pedestrians’ emotions owing to communication or odor in the environment. In addition, the precision and accuracy of well-being evaluated could be enhanced by combining the FERM with location information or personal vital data. Finally, it is essential to develop a design method that can take personal attributes into consideration to cater to the diverse needs of a society.

In the future, FERM can be utilized to recognize facial expressions with multiple attributes in real time using cameras. In addition, the SIEM can make the decision-making process more transparent and auditable in a workshop. As a consequence, the AIHCE enables residents’ engagement, promotes hedonistic sustainability and citizen well-being, and enriches collaboration among various stakeholders in street space design.

Author Contributions: Conceptualization, K.D.; supervision, K.D.; methodology, K.S. and H.S.; validation, K.Y.; data curation, K.S. and H.S.; writing—original draft preparation, K.S. writing—review and editing; K.D. and K.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Science and Technology Research Partnership for Sustainable Development (SATREPS), grant number JPMJSA1704 (co-leaders: Yoshitsugu Hayashi and Thanaruk Theeramunkong).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Acknowledgments: The authors wish to thank Ikuo Sugiyama, for his advice all through the study. Our gratitude also goes to Chun-Chen Chou, Graduate Student at Osaka University, for her linguistic support. This research is a part of Smart Transport Strategy for Thailand 4.0 project supported by the Japan Science and Technology Agency (JST) and the Japan International Co-operation Agency (JICA).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. “Igokoti ga Yoku Arukitakunaru Matinaka” Kara Hazimaru Tosinosaisei (Urban Regeneration Starting with Comfortable Town Centers that Make People Want to Walk around). Available online: <https://www.mlit.go.jp/common/001301647.pdf> (accessed on 7 April 2021).
2. Singata Korona Kiki wo Keiki Tosita Matidukuri no Hōkōsei (The Direction of Urban Development in the Wake of the New Corona Crisis). Available online: <https://www.mlit.go.jp/toshi/machi/content/001361466.pdf> (accessed on 7 April 2021).
3. Fischer, G.; Scharff, E. Meta-design: Design for designers. In Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, New York, NY, USA, 17–19 August 2000; pp. 396–405. [CrossRef]
4. Daijiro, M. Design Research Through Practice as an Interdisciplinary Field of Enquiry—Research into, for, through Design. *Keio SFC J.* **2014**, *14*, 62–80.
5. Christopher, D.; Julie, G. *Living Architecture, Living Cities: Soul-Nourishing Sustainability*; Routledge: London, UK, 2019.
6. FASTCOMPANY, Scientists Are Using Eye-Tracking To Discover How We See Design. 2017. Available online: <https://www.fastcompany.com/90153582/scientists-are-finally-discovering-how-our-eyes-really-see-space> (accessed on 3 June 2021).
7. Archinect News, Designing through Cognitive Architecture. 2017. Available online: <https://archinect.com/news/article/150040919/designing-through-cognitive-architecture> (accessed on 3 June 2021).
8. Fruin, J.J. *Hokōsya no Kūkan—Risō to Dezain—(Pedestrian Spaces—Ideals and Design)*; Kajima Institute Publishing: Tokyo, Japan, 1974.
9. Bill, H.; Juicne, H. *The Social Logic of Space*; Cambridge University Press: Cambridge, UK, 1984.
10. Davies, A.; Clark, S. Identifying and prioritizing walking investment through the PERS audit tool. In Proceedings of the Walk21, 10th International Conference for Walking, New York, NY, USA, 6 October 2009.
11. Christopher, A.; Kanna, H. *Patan Rangēzi-Kankyō no Tebiki-(Pattern Language—A Guide to Environmental Design-)*; Kajima Institute Publishing: Tokyo, Japan, 1984.
12. Rafael, A.C.; Dorian, P.; Jun, W.; Dominique, C. *Uerubīngu Sekkeiron—Hito ga Yoriyoku Ikiru Tameno Zyōhōgizyutu—(Well-Being Design Theory—Information Technology for Better Living)*; BNN: Tokyo, Japan, 2017; pp. 28–47, 143–153, 233–245.
13. Ryan, R.M.; Deci, E.L. On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annu. Rev. Psychol.* **2001**, *52*, 141–166. [CrossRef]
14. Diener, E.; Suh, E.M.; Lucas, R.E.; Smith, H.L. Subjective Well-Being: Three Decades of Progress. *Psychol. Bull.* **1999**, *125*, 276–302. [CrossRef]
15. Compton, W.C.; Smith, M.L.; Cornish, K.A.; Qualls, D.L. Factor structure of mental health measures. *J. Personal. Soc. Psychol.* **1996**, *71*, 406–413. [CrossRef]
16. McGregor, I.; Little, B.R. Personal projects, happiness, and meaning: On doing well and being yourself. *J. Personal. Soc. Psychol.* **1998**, *74*, 494–512. [CrossRef]
17. Jan, G.; Tshio, K. *Ningen no Mati: Kōkyōkūkan No Dezain (Human City: Design of Public Space)*; Kajima Institute Publishing: Tokyo, Japan, 2014; pp. 28–32.
18. Kaoru, O.; Yuji, H.; Kota, M. An approach for understanding remembered experiences in urban city. *J. Jpn. Soc. Civ. Eng. Ser. D1 (Archit. Infrastruct. Environ.)* **2015**, *71*, 133–150. [CrossRef]
19. Alex, K.; Ilya, S.; Geoffrey, E.H. ImageNet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems—Volume 1 (NIPS’12), Lake Tahoe, NV, USA, 3–6 December 2012; Curran Associates Inc.: Red Hook, NY, USA, 2012; Volume 1, pp. 1097–1105.
20. Ye, Y.; Zeng, W.; Shen, Q.; Zhang, X.; Lu, Y. The visual quality of streets: A human-centred continuous measurement based on machine learning algorithms and street view images. *Environ. Plan. B Urban Anal. City Sci.* **2019**, *46*, 1439–1457. [CrossRef]
21. Yin, L.; Wang, Z. Measuring visual enclosure for street walkability: Using machine learning algorithms and Google Street View imagery. *Appl. Geogr.* **2016**, *76*, 147–153. [CrossRef]
22. Li, Y.; Yabuki, N.; Fukuda, T.; Zhang, J. A big data evaluation of urban street walkability using deep learning and environmental sensors—A case study around Osaka University Suita campus. In Proceedings of the 38th eCAADe Conference, TU Berlin, Berlin, Germany, 16–18 September 2020; Volume 2, pp. 319–328.
23. Lun, L.; Elisabete, A.S.; Chunyang, W.; Hui, W. A machine learning-based method for the largescale evaluation of the qualities of the urban environment. *Comput. Environ. Urban Syst.* **2017**, *65*, 113–125. [CrossRef]
24. Dubey, A.; Naik, N.; Parikh, D.; Raskar, R.; Hidalgo, C.A. Deep Learning the City: Quantifying Urban Perception at a Global Scale. In Proceedings of the European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 8–16 October 2016.
25. Zhang, F.; Zhou, B.; Liu, L.; Liu, Y.; Fung, H.H.; Lin, H.; Ratti, C. Measuring human perceptions of a large urban region using machine learning. *Landsc. Urban Plan.* **2018**, *180*, 148–160. [CrossRef]
26. Ji, H.; Qing, L.; Han, L.; Wang, Z.; Cheng, Y.; Peng, Y. A New Data-Enabled Intelligence Framework for Evaluating Urban Space Perception. *ISPRS Int. J. Geo Inf.* **2021**, *10*, 400. [CrossRef]
27. Seresinhe, I.C.; Preis, T.; Moat, H.S. Using deep learning to quantify the beauty of outdoor places. *R. Soc. Open Sci.* **2017**, *4*, 170170. [CrossRef] [PubMed]
28. Satoshi, Y.; Kotaro, O. Development and verification of the impression deduction model for city landscape with deep learning. *J. Archit. Plan. (Trans. AII)* **2019**, *84*, 1323–1331. [CrossRef]

29. Taichi, F.; Aya, K.; Hisashi, K. Study on evaluation method for street space focusing on pedestrian behavior. *J. Jpn. Soc. Civ. Eng. Ser. D3 (Infrastruct. Plan. Manag.)* **2011**, *67*, 919–927. [[CrossRef](#)]
30. Andre, T.L.; Edilson, D.A.; Albeesrto, F.D.S.; Thiago, O.S. Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order. *Pattern Recognit.* **2017**, *61*, 610–628. [[CrossRef](#)]
31. Deepak, K.J.; Zhang, Z.; Kaiqi, H. Multi angle optimal pattern-based deep learning for automatic facial expression recognition. *Pattern Recognit. Lett.* **2020**, *139*, 157–165. [[CrossRef](#)]
32. Shungo, N.; Tatsuya, K. Analysis of Age, Gender ratio, Emotion, Density of Pedestrians in Urban Street Space with Image Recognition. *J. City Plan. Inst. Jpn.* **2020**, *55*, 451–458. [[CrossRef](#)]
33. Min, S. How Emotional AI Can Benefit Businesses, Campaign. Available online: <https://www.campaignasia.com/article/how-emotional-ai-can-benefit-businesses/455912> (accessed on 27 July 2021).
34. Ramprasaath, R.S.; Micheal, C.; Abhishek, D.; Ramakrishna, V.; Devi, P.; Dhruv, B. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 618–626. [[CrossRef](#)]
35. Russell, J.A. A circumplex model of affect. *J. Personal. Soc. Psychol.* **1980**, *39*, 1161–1178. [[CrossRef](#)]
36. Christie, A.M.; Atkins, P.W.B.; Donald, J.N. The Meaning and Doing of Mindfulness: The Role of Values in the Link Between Mindfulness and Well-Being. *Mindfulness* **2017**, *8*, 368–378. [[CrossRef](#)]
37. Fredrickson, B.L. What good are positive emotions? *Rev. Gen. Psychol.* **1998**, *2*, 300–319. [[CrossRef](#)]
38. Watson, D.; Clark, L.A.; Auke, T. Development and validation of brief measures of positive and negative affect: The PANAS scales. *J. Personal. Soc. Psychol.* **1998**, *54*, 1063–1070. [[CrossRef](#)]
39. Rowland, Z.; Wenzel, M.; Kubiak, T. A mind full of happiness: How mindfulness shapes affect dynamics in daily life. *Emotion* **2020**, *20*, 436–451. [[CrossRef](#)] [[PubMed](#)]
40. Verywellmind, The Color Psychology of Yellow. 2020. Available online: <https://www.verywellmind.com/the-color-psychology-of-yellow-2795823> (accessed on 7 July 2021).
41. 99designs by Vistaprint, Colors and Emotions: How Colors Make You Feel. Available online: <https://en.99designs.jp/blog/tips/how-color-impacts-emotions-and-behaviors/> (accessed on 16 July 2021).