Development of a Method for Wide-area Analysis of Streetscapes Using Deep Learning

Yuji Sano*1 Kuniaki Ando*2

Summary

We have developed a neural network capable of analyzing streetscapes across wide urban areas at the town scale. To construct training data for scenic impression evaluation, approximately 100 architectural designers and urban planners annotated a database of about 1,000 landscape images. Through streetscape analysis of Shibuya and Hisaya-ōdōri Park, we confirmed the validity of changes in visual impressions across different areas and times of day.

Keywords: streetscape, impression evaluation, deep learning, urban planning

1 Introduction

Many cities have been concerned about the decline in regional vitality caused by population decline and the aging of society in recent years, and a common need across cities has been the improvement in their attractiveness and the creation of lively environments. Effective evaluation methods and design techniques that contribute to improving the attractiveness of pedestrian spaces are needed for creating walkable and appealing urban spaces that people want to visit.

The Ministry of Land, Infrastructure, Transport, and Tourism has set guidelines on streetscapes as an aspect of spatial design necessary for ground-level design in its policy to create “comfortable and walkable” urban areas[1]. Hence, the importance of urban planning based on streetscapes is high. Therefore, we organized the specifications of the evaluation method and design technology required for streetscapes. The result is the streetscape evaluation tool that we have devised, shown in Fig. 1, where we have targeted the following two points considered important:

1, An evaluation method and design technology that are easy for planners to use in practice.

2, Planners should be able to analyze the correlation between an impression evaluation and its associated elements and features.

Creating quantitative evaluation criteria for impressions, which are subjective evaluations, is difficult; however, this has become easier to achieve with the recent development of deep learning technology. Research on neural networks targeting street names and visit motivation has already been conducted[2]; however, it is not clear what streetscape components contribute to impressions. We can use this tool to extract the elements and features that are considered the basis of impression evaluation from streetscape images, and we can evaluate the impression using an artificial intelligence model (AI) that has learned human impression judgment. We can then compare the extracted elements and features with the impression evaluation results to examine how the design contributes to the impression given. Moreover, we can apply this tool to a wide area rather than just a single building to refine development plans based on the town's attractions and features.

We interviewed planners about their functional needs prior to the development of this tool. The results indicated two strong needs for achieving urban planning based on streetscapes, namely, the ability to perform impression evaluations for each visitor attribute and the ability to compare data that evaluates the attractiveness of the city other than streetscapes. We divided the overall

Fig. 1 Overview of the proposed streetscape analysis tool

＊1 Senior Researcher, Research & Development Institute, Dr. Hum. Info.
＊2 Chief Researcher, Research & Development Institute

research concept for the development of the tool that responds to these needs into three steps, as shown in Fig. 2. In Step 1, "Development of a streetscape analysis tool using street images by deep learning," we provide an overview of the development, create training data for impression evaluation, and conduct streetscape analysis, AI learning, and visualization.

2 Overview of the development

2.1 The streetscape evaluation tool and selection of the streetscape image collection area

This tool extracts "elements and features" and evaluates "impressions" by inputting streetscape images or perspectives created by designers. The streetscape images containing location information were visualized on a map. The targets in this study were areas around Shibuya Station, Takeshita Street, Omotesandō, and Oku-Shibuya in Shibuya Ward. Shibuya Ward has a compact residential area and a downtown area, and we believe that effective comparisons can be made when visualized on a map.

Table 1 lists the evaluation items for streetscape images. The evaluation items consisted of major items set according to Caption Evaluation Method[3] and medium and small items set based on interviews with our company's architectural designers and urban planners. Previous research[4] used "regionality" as an evaluation item, so we added "Shibuya-like" to the impression items corresponding to the Shibuya area targeted in this study.

2.2 Image collection method

Fig. 3 shows the image capture device. The photography area comprised the streets around Shibuya Station, which is considered important by our company's urban design planners, and 360° images were manually captured using a dolly-type photography device. We set a photography period from July 21st to September 13th, 2021. During this period, we took the photographs during the day for a total of five days, including four sunny days in the morning and afternoon and one cloudy day. The photography days were as follows: four sunny days on July 21st (maximum temperature: 33.6 °C), August 3rd (32.9 °C), August 4th (34.5 °C), and August 5th (34.7 °C); and one cloudy day on September 13th (30.7 °C). The total number of photography locations was approximately 11,000. The 360° images were cropped at 90° horizontally and 74° vertically based on previous research[5], and the photographs taken in the forward direction were used.

3. Construction of training data for impression evaluation

Training data are necessary for learning in neural networks. We built training data for impression evaluation by conducting impression labeling ("annotation") of images using a questionnaire survey, relying on previous research[5]. We constructed a questionnaire response system, shown in Fig. 4, and requested about 100 architects and urban planners to evaluate the impressions

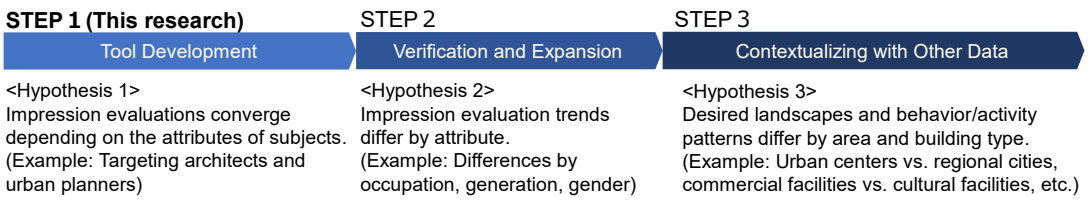

Fig. 2 Overall research concept

Fig. 3 Street Scene Imaging Equipment

Fig. 4 Questionnaire response system for impression evaluation annotation

of approximately 4,000 streetscape images. We asked the participants to compare two randomly displayed streetscape images and select the one with the higher impression scale. They answered 100 times for each impression evaluation item and 800 times for the eight items, which yielded a total of approximately 80,000 responses.

We created a histogram of the number of annotations to confirm the degree to which the images in the dataset were annotated. Fig. 5 shows the histogram for the annotation of "stylish/tasteful" as an example.

This dataset has more than 4,000 images, and annotating all image combinations would require over 8 million annotations. However, the annotation can be considered sufficient if the image ranking based on the annotation results shows a clear tendency that can be quantified. Fig. 6 shows the top images selected as "stylish/tasteful." These images were selected

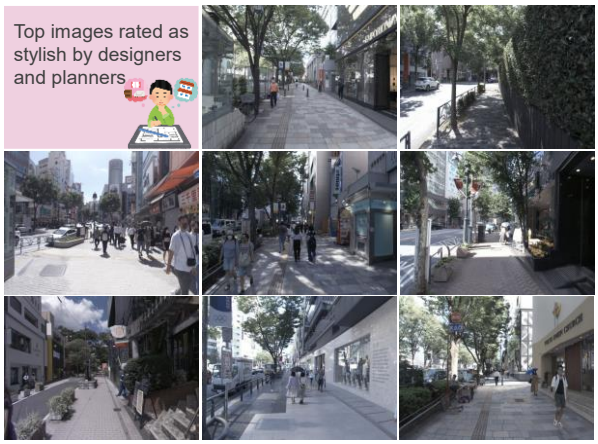

Fig. 6 Example of impression-annotated image

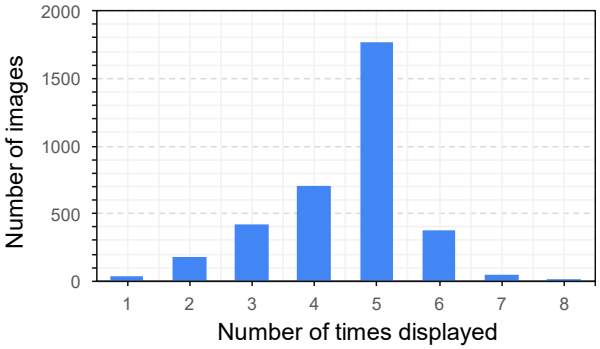

Fig. 5 Distribution of the number of times an image is displayed in the annotation

Table 1 List of landscape image analysis and evaluation items

Main category: element		Main category: feature	
Subcategory	**Sub-subcategory**	**Subcategory**	**Sub-subcategory**
Person	Person	Person	Activity
Architecture	Window	Architecture	Window Ratio
	Outer wall		Building Unevenness
	Street-level store		Edge
	Tent / Awning		Character
	Shop board		Text (amount, languages)
	Store display items		Color various
	Sign		Material composition
	Condenser unit	Infrastructure	Pavement damage
	Shed		Graffiti
	Vending machine	Nature, terrain	Green view index
Infrastructure	Roadway / Sidewalk		Sky view index
	Utility pole		
	Bicycle	Main category: impression	
	Streetlight	Basic impression	
	Garbage collection point / Garbage	**Subcategory**	**Sub-subcategory**
	Guardrail / Bollard	Individuality	Stylish / Fashionable
	Street furniture	Routine	Calm / Serene
Nature, terrain	Plant	Orderliness	Messy / Chaotic
	Sky	Familiarity	Friendly / Approachable
	Water feature	Comprehensive impression	
		Subcategory	**Sub-subcategory**
		Openness	Lively / Bustling
			Anxiety-inducing / Unsettling
		Vivacity	Cozy
		Other	Shibuya-like

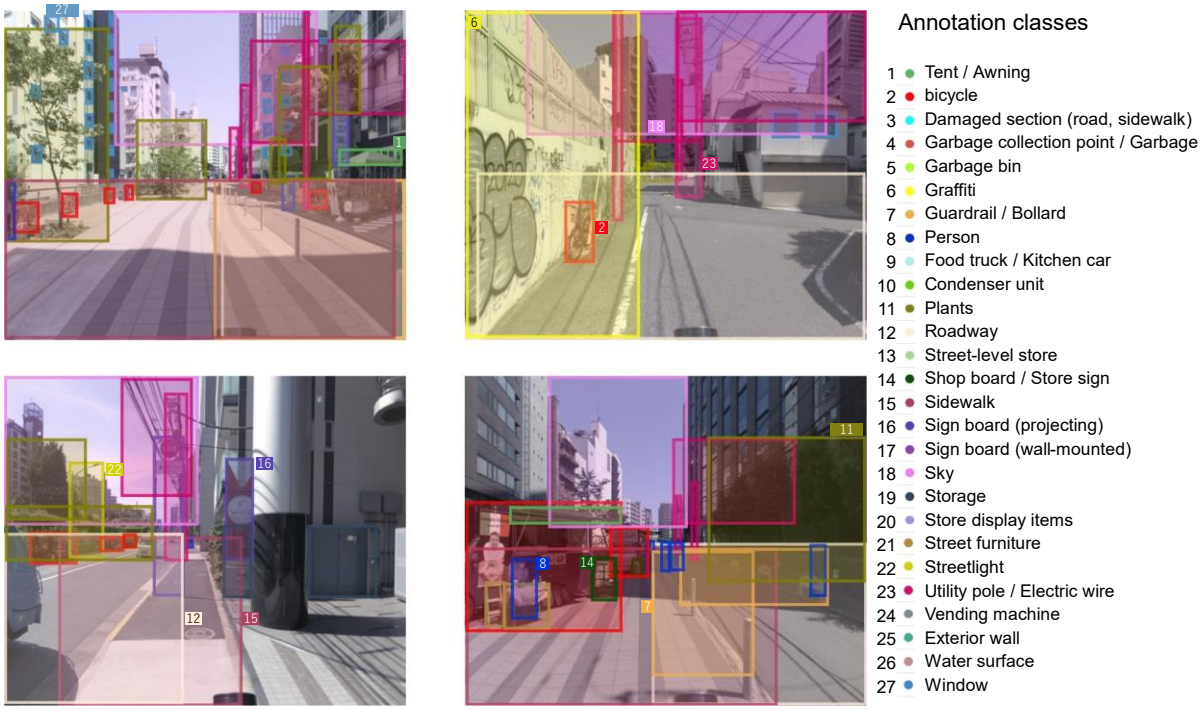

Fig. 7 Example of element annotation for landscape images

because the sunlight filtering through the trees and the building facades were in harmony, they included a moderate number of pedestrians, and the designers and planners who annotated them provided feedback stating that the results were satisfactory. We judged from these results that the training data were sufficiently annotated and that there was a clear tendency in impression evaluation that could be quantified by deep learning.

4. Learning of streetscape analysis AI

We describe the training of the streetscape analysis AI with the output presented in Table 1. The element items in Table 1 were annotated as shown in Fig. 7 for the approximately 11,000 images collected, as described in Section 2.2, and additional training was performed on the pre-trained EfficientNet[6]. Items with fewer than 100 training data were supplemented with existing public image datasets.

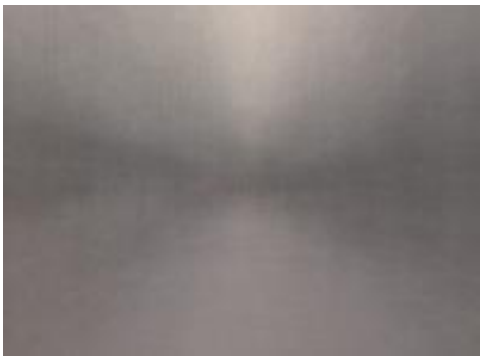

Fig. 8 Reference image averaged from 500 superimposed street images

Among the feature items, color diversity was expressed using Simpson's diversity index[7] of a hue histogram with a resolution of 16. The window opening ratio was defined as the ratio of the window detection rectangle area to the exterior wall detection rectangle area. The green view ratio, sky ratio, and building unevenness were colored using the DeepLab v3+ network[8], which colors the objects detected in an image, and were expressed as the plant area ratio, sky area ratio, and standard deviation of the outline of the building area, respectively. The edges were extracted using a Sobel filter[9], and the entropy of the extracted images was calculated.

For the impression items, we trained a model that could compare and evaluate images so that the output would be the winning rate with respect to the reference image when each impression item was compared, as in the annotations in Section 3. The reference image was the average of 500 images randomly extracted from the training data (Fig. 8).

5. Streetscape analysis AI output analysis

5.1 Extraction of elements

Average precision (AP) and other methods are generally used to verify the accuracy of object detection. However, in design and planning, the number of elements is more important than the detected position. Therefore, we verified the number of extracted elements based on reproducibility, which indicates the percentage of images in which the number of annotations in the training data matched. The reproducibility of each element item presented in Table 2 indicates that bikes and awnings had high reproducibility, whereas garbage, storage sheds, and water surfaces had a reproducibility of 0. This occurred mainly because the collected images contained extremely few of the above elements, ranging from a few to several dozen, and the training data lacked the elements contained in the streetscape images, even though over 100 images had been supplemented and learned from other datasets. The lack of annotation for garbage owing to ambiguity in the definition was also thought to have had an impact. It is expected that improvements could be made not only by providing further annotation of streetscape images containing the missing elements but also by using semantic segmentation for planar elements such as water surfaces.

Fig. 9 shows the results of element extraction. The numbers in the boxes in the images indicate the reliability of object recognition (AI confidence). A trend in element extraction was that object recognition was biased from the center to the upper left area of the image, even though the extraction results had a certain degree of validity. The model should be improved so that the entire image could be evenly extracted.

Table 2 Element extraction accuracy (Recall ratio)

Elements	Recall ratio	Elements	Recall ratio	Elements	Recall ratio	Elements	Recall ratio
Motorcycle	0.95	Garbage	0.00	Graffiti	0.82	Store display items	0.77
Awning	0.90	Shopboard	0.72	Guardrail	0.60	Streetlight	0.87
Bicycle	1.00	Sidewalk	0.96	Person	0.45	Street furniture	1.00
Car	0.80	Projecting sign	0.52	Kitchen car	1.00	Utility pole and wire	0.58
Damaged pavement	1.00	Wall sign	0.60	Condenser unit	0.97	Vending machine	1.00
Garbage collection point	0.84	Sky	0.90	Plant	0.88	Outer wall	0.89
Trash box	1.00	Shed	0.00	Street-level store	0.84	Water feature	0.00
Roadway	0.96	Window	0.51				

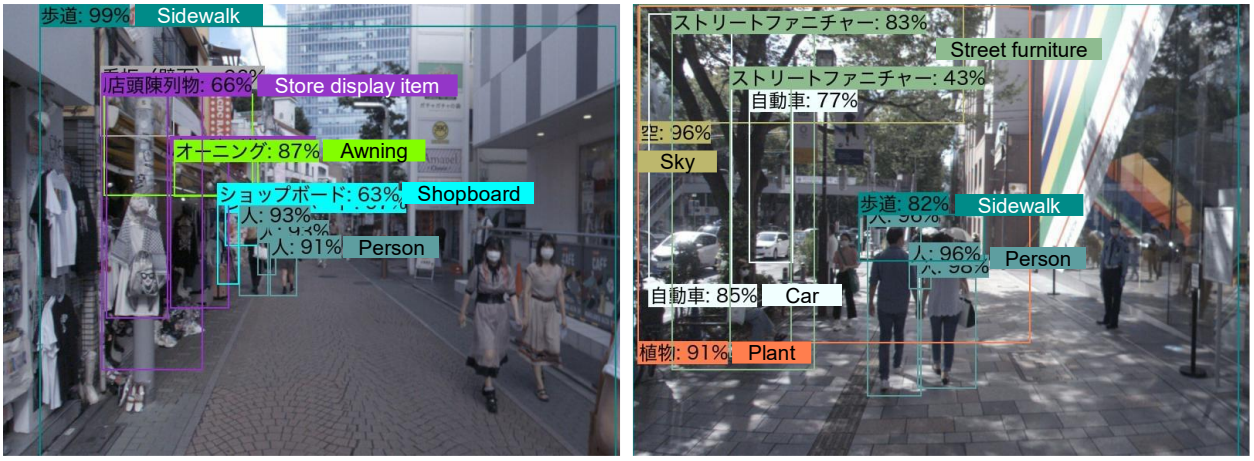

Fig. 9 Example of detection of streetscape elements by element extraction AI

5.2 Results of impression evaluation

We analyzed the output of the AI learned in Section 4 by inputting approximately 11,000 streetscape images of streets in Shibuya Ward. Fig. 10 shows examples of the top and bottom impression evaluations when a ranking table was created from the winning and losing tables of all images input to the AI. This AI is not a regression model that directly quantifies impression values, but we can obtain similar results by defining the winning rate with respect to the reference image in Fig. 8 as the impression value.

Whether the reference image is a truly standard streetscape image is unknown. Thus, we verified the validity of the reference image by checking whether the behavior as a classification model and the behavior as a pseudo regression model matched in all combinations of the input image group. The agreement rate between the training data and inference result was defined as the comparative correct answer rate and was evaluated for each impression item (Table 3). The accuracy rate was approximately 70%. The comparative accuracy rate was also verified for images with almost the same impression evaluation, and the rate was higher for images with a sufficient difference in impression evaluation.

Fig. 10 Examples of top and bottom impression ratings by AI

Table 3 Accuracy of comparison for Impression Rating

Impression item	Accuracy of comparison
Stylish / Fashionable	67%
Calm / Serene	66%
Messy / Chaotic	67%
Friendly / Approachable	65%
Lively / Bustling	74%
Anxiety-inducing / Unsettling	72%
Cozy	64%
Shibuya-like	65%

5.3 Relationship between impression evaluation and elements/features

Fig. 11 shows the correlation coefficients of the main 25 items from the output of Section 2. The results in Fig. 11 showed that the elements/features that exhibited a correlation ($|R| \geq 0.5$) with the impression items were people, plants, green view ratio, and edge (entropy). The “people” element exhibited a positive correlation with liveliness and Shibuya-like, and a negative correlation with degree of anxiety, which corresponded to the impression given by the large number of people.

The “plant” element exhibited a correlation with stylishness and friendliness. The green view ratio was affected by the area of detected plants, so a correlation was also observed between stylishness and friendliness. However, a similar correlation was also observed for the edge (entropy), as well as a correlation with calmness. Well-maintained plants were assumed to give a positive impression, and sunlight filtering through the trees increased the edge component. We also observed a correlation among the impression items, with a particular correlation between anxiety and Shibuya-like/liveliness and between stylish/tasteful and friendliness.

Fig. 11 shows the correlation coefficients of the main 25 items from the output of Section 2. The results in Fig. 11 showed that the elements/features that exhibited a correlation ($|R| \geq 0.5$) with the impression items were people, plants, green view ratio, and edge (entropy). The "people" element exhibited a positive correlation with liveliness and Shibuya-like, and a negative correlation with degree of anxiety, which corresponded to the impression given by the large number of people.

The "plant" element exhibited a correlation with stylishness and friendliness. The green view ratio was affected by the area of detected plants, so a correlation was also observed between stylishness and friendliness. However, a similar correlation was also observed for the edge (entropy), as well as a correlation with calmness. Well-maintained plants were assumed to give a positive impression, and sunlight filtering through the trees increased the edge component. We also observed a correlation among the impression items, with a particular correlation between anxiety and Shibuya-like/liveliness and between stylish/tasteful and friendliness.

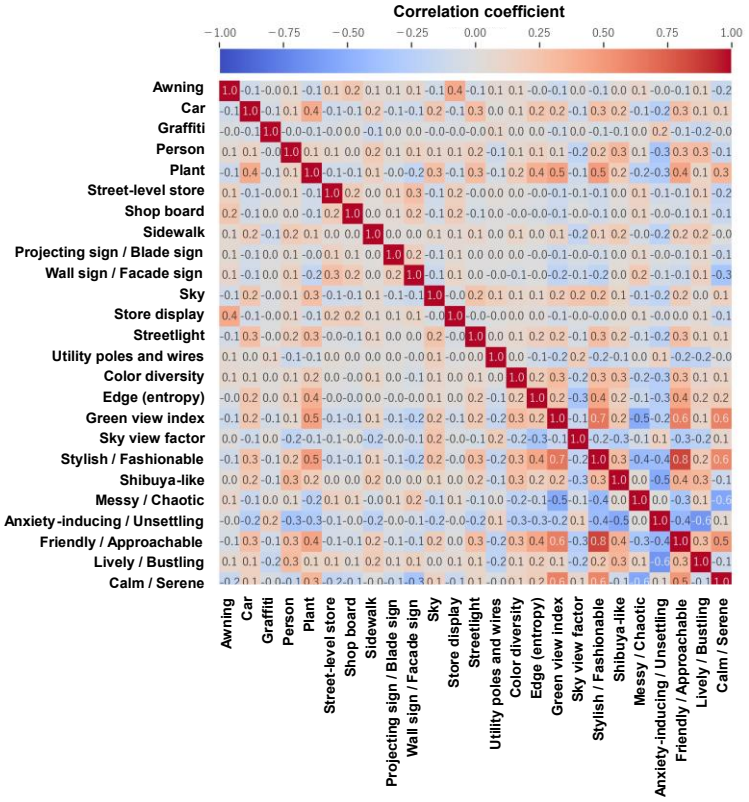

Fig. 11 Correlation coefficient grid among elements, characteristics, and impressions

6. Visualization using streetscape analysis map

An aspect that is considered important in the development of an "evaluation method and design technology that are easy for planners to use in practice" is the determination of the characteristics of the streetscape distributed in towns and streets. Therefore, the results obtained in Section 5 were visualized using an online application. Fig. 12 shows the visualized streetscape analysis map. In this application, the evaluation items can be selected and filtered on the left side, and a radar chart can be used to compare locations in the right window.

Among the streets studied, Omotesandō was particularly distinctive, ranking either top or bottom in many impression items, as

Fig. 12 Web application for visualization of landscape analysis results

shown in Fig. 10. As an example of the elements and feature items that are correlated with the impression of Omotesandō, Fig. 11 shows a correlation between edges and "stylish/tasteful." The visualization of the edge features on a map showed that Omotesandō certainly exhibited high values. The photographs of Omotesandō exhibited a large amount of sunlight filtering through the trees. This increased entropy, which represents edge features.

Planners can use this tool in this manner to determine the impressions of the streetscape of a city or street while analyzing the correlated elements and features.

7. On-site verification for performance evaluation of the streetscape analysis AI

We evaluated the performance of the developed streetscape analysis AI based on an event held in Nagoya City in October 2022. First, we used the AI to compare the development project before and after the event and then evaluated the change in impression during the event period in a square. We measured and evaluated the streetscape impression of this event by taking photographs with two methods, namely, fixed-point and mobile photography.

For fixed-point photography, we used a Brinno camera TLC200 to take photographs at three locations (Points A, B, and C), shown in Fig. 13, at intervals of 10 min per photograph. Point A was at one of the entrances to the event space, Point B was in the space where the main events, such as live music concerts, took place, and Point C was in a space where people could eat and drink. We selected two weekdays and two weekends from the photography period, and we used images taken from sunrise to lights out to evaluate whether the changes in streetscape impression over time were captured depending on the purpose of the shooting location (traffic, events, and staying).

For the mobile photography, we recorded videos of the event space and its surroundings in 360° video mode using an Insta360 ONE RS camera manufactured by Insta360 three times on Saturdays during the event period at 10:00, 13:00, and 16:00 (Fig. 13). The front direction of the shot video was trimmed to a horizontal angle of 80° and a vertical angle of 60° for analysis.

In this study, we used two methods of analysis. For analysis method 1, we loaded images from the morning (6:00–11:00), afternoon (11:00–17:30), and evening (17:30–22:00) into the AI and analyzed them by time period in the square (images taken with fixed-point photography). For analysis method 2, we loaded images into the AI and analyzed the changes due to the route followed (images acquired with mobile photography). We then used these two methods as a basis to evaluate the square in seven categories: stylish/fashionable, calmness, messy/chaotic, friendly, liveliness, anxious and cozy.

The impression evaluation value of the group of images acquired with fixed-point photography was calculated as a standard deviation using the images as the population. Fig. 14 shows the "liveliness" that was emphasized at the event.

At Points A and B, the standard deviation was biased to 51–60 during the daytime hours of the holiday (red frame), and the holiday daytime was evaluated as a higher value than the same time period on weekdays. Point B also showed an area with lights, suggesting that the "liveliness" evaluation value was higher than that at Point A because the participants of the evening event gathered there.

In Fig. 14, the standard deviation was biased at Point C, which was attributed to the fact that the park with many

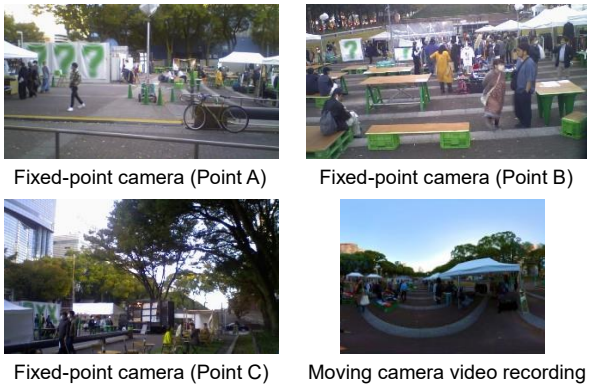

Fig. 13 Fixed-point and mobile photography for verification of landscape assessment

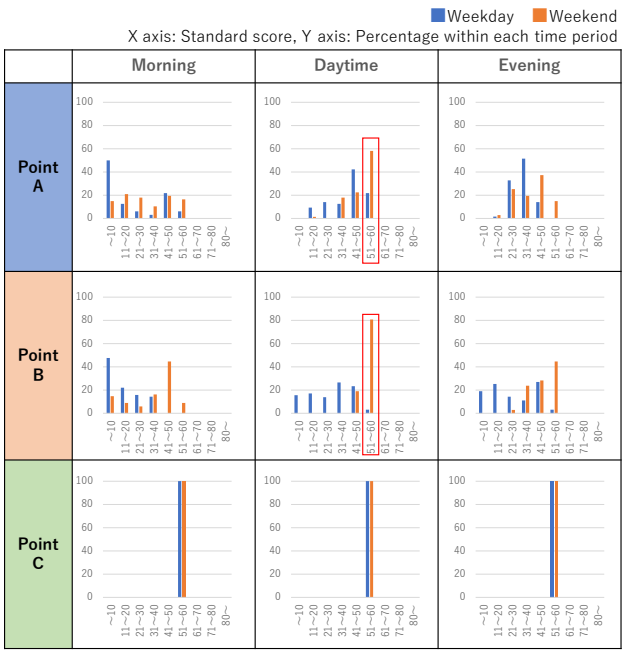

Fig. 14 Results of impression analysis of the "liveliness" of the event space

temporary structures shown in Fig. 13 was not included in the AI learning data, and the AI was unable to evaluate the change in the streetscape.

The comparison of impression evaluations between weekdays and holidays presented in Fig. 15 showed that Point B, which had more elements such as fixtures and stalls than other points, consistently had a high evaluation value for the impression of being “cluttered.”

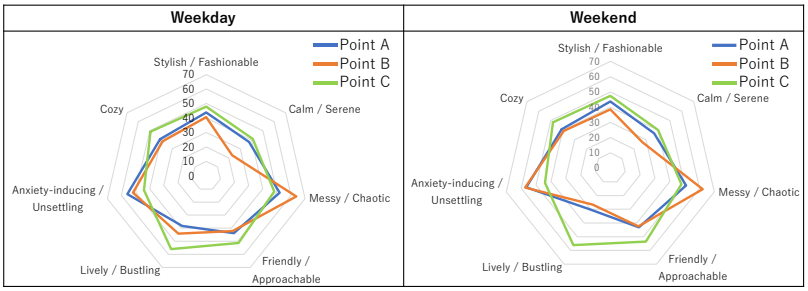

Fig. 15 Evaluation results of fixed-point photography of event space

The results of analysis method (2) were also calculated with the standard deviation in the same way as analysis method (1). Here, we focused on the “liveliness” of the photographs taken at 13:00, that is, the session with the most users around the square among the three photography sessions. Fig. 16 presents photographs that were highly rated in the event space and its surrounding areas, and Fig. 17 shows the overall sequence evaluation. Fig. 16(a) exhibits an image of the event's central location, with many participants and products from the stalls lined up, and Fig. 16(b) shows an image in front of a commercial facility with benches and parasols. Both images matched the subjective impression of “Lively / Bustling.” Fig. 17 shows that the evaluation of “Lively / Bustling” had a large distribution on the lower side, so we could not see any fluctuations in the standard deviation above 60, but low standard deviations were observed in places such as behind buildings and in quiet squares.

As can be observed in Figs. 15 and 16, we performed evaluations on weekdays and holidays and by time of day and were able to capture and compare the changes in impressions by evaluating consecutive streetscape images, as shown in Fig. 17. Taking photographs of consecutive locations also enabled us to express the characteristics of impressions at specific locations, as shown in Fig. 18. Meanwhile, Fig. 14 exhibits an area at Point C where the AI evaluation results were unnaturally biased.

We can improve the bias in the AI evaluation by incorporating streetscape images with diverse compositions into the current limited learning data. Then, we can quantitatively evaluate streetscape impressions even in applications such as before-and-after comparisons of development projects.

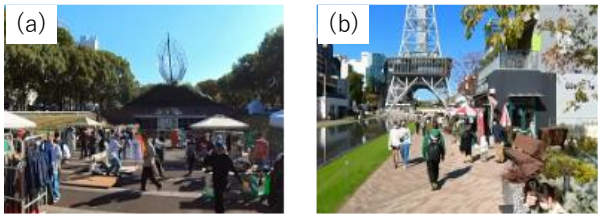

Fig. 16 Landscape images highly rated for “liveliness”

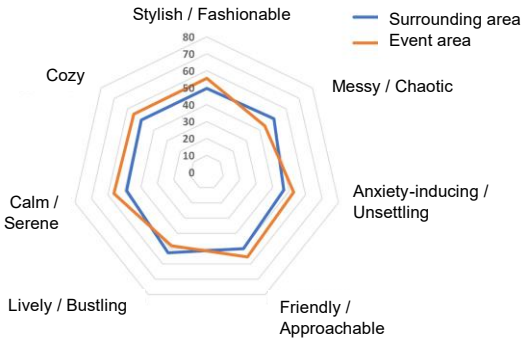

Fig. 18 Results of evaluation of event space and surrounding area

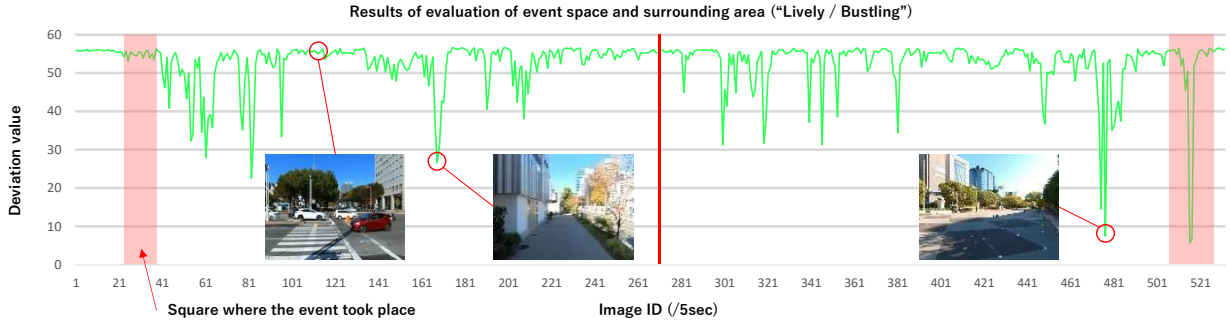

Fig. 17 Results of the evaluation of the event space and surrounding area for “Lively / Bustling”

8. Conclusion

In this study, we proposed an evaluation tool that extracts elements and features from streetscape images and infers impressions of the images. It is an effective evaluation method for streetscapes that contribute to improving the attractiveness of pedestrian spaces. For this tool, we specifically targeted the following important goals: (1) an evaluation method and design technology that are easy for planners to use in practice, and (2) planners should be able to analyze the correlation between impression evaluation and its associated elements and features. We conducted streetscape analysis based on deep learning by building an AI that learned the streetscape impression evaluations performed by approximately 100 architects and urban planners and output a total of 44 elements, features, and impressions. The results showed that the elements and features that show correlation ($|R| \geq 0.5$) with the eight streetscape impression items in the target area were people, plants, green view ratio, and edges. These results indicated that we achieved goal (2). We also visualized the obtained results on a streetscape analysis map, where we confirmed that the characteristics of the street were captured. These results indicated that we achieved goal (1).

We evaluated the performance of the developed streetscape analysis AI by using it to compare the state before and after a development project. We evaluated the change in impression during the period of an event held in Hisaya- ōdōri Park, Nagoya City, as a demonstration test. The results showed that we were able to capture the characteristics of impressions depending on the time of day and the sequential change in impressions from consecutive streetscape images. Some failures were found in the impression evaluation, which were attributed to a lack of training data. We may be able to improve these failures by learning additional diverse streetscapes that include temporary structures and parks.

The development of this tool will enable planners to capture the characteristics of the streetscape of a city over a wide area, allowing the planning of a streetscape that blends with the city and enhances its appeal. We also expect that this tool will enable planners to consider the elements and features that are necessary to obtain the impression of the streetscape that they have in mind.

References

1) Ministry of Land, Infrastructure, Transport and Tourism: Guidelines for the utilization of road space for creating a “comfortable and walkable” urban center, April 2022
2) Yamada, S., Ono, K.: Development and verification of the impression deduction model for city landscape with deep learning: Street names city landscapes and desire / no desire or degree of desire to visit, Journal of Architecture and Planning (Transactions of AIJ), Vol. 84, No. 759, pp. 1323-1331, May 2019
3) Kojima, T., Koga, T., Munakata, J., Hirate, K.: Multivariate analysis on verbal data of “caption evaluation method”: Studies of the cognition and evaluation of townscape Part 2, Journal of Architecture and Planning (Transactions of AIJ), Vol. 67, No. 560, pp. 51-68, December 2002
4) Hyakuri, M.: Systematization of impression evaluation indicators in streetscapes–Discussions from nighttime streetscapes, Master’s thesis, Graduate School of Frontier Sciences, University of Tokyo, March 2006
5) P. Salesses, K. Schechtner, and C. A. Hidalgo. "The collaborative image of the city: mapping the inequality of urban perception," *PloS one*, Vol. 8, No. 7, pp. e68400, 2013.
6) M. Tan, and Q. V. Le. "EfficientNet: Rethinking model scaling for convolutional neural networks." International conference on machine learning. PMLR, 2019.
7) E. H. Simpson. "Measurement of diversity." Nature, No. 163, Vol. 4148, pp.688, 1949.
8) R. Azad, et al. "Attention deeplabv3+: Multi-level context attention mechanism for skin lesion segmentation." European conference on computer vision. Springer, Cham, 2020.
9) N. Kanopoulos, N. Vasanthavada, and R. L. Baker. "Design of an image edge detection filter using the Sobel operator." IEEE Journal of solid-state circuits, No. 23, Vol. 2, pp. 358-367, 1988.