RESIDENT'S SATISFACTION IN STREET LANDSCAPE USING THE IMMERSIVE VIRTUAL ENVIRONMENT-BASED EYE-TRACKING TECHNIQUE AND DEEP LEARNING MODEL

J. Han, S. Lee*

Dept. of Urban Planning & Engineering, Hanyang University, Seoul, Republic of Korea - (tommorello,sugielee@hanmail.net)

Commission IV, WG IV/9

KEY WORDS: Street Landscape, Visual Attention, Immersive Virtual Reality, Eye-tracking, Deep-learning.

ABSTRACT:

Virtual reality technology provides a significant clue to understanding the human visual perception process by enabling the interaction between humans and computers. In addition, deep learning techniques in the visual field provide analysis methods for image classification, processing, and segmentation. This study reviewed the applicability of gaze movement and deep learning-based satisfaction evaluation on the landscape using an immersive virtual reality-based eye-tracking device. To this end, the following research procedures were established and analysed. First, the gaze movement of the test taker is measured using an immersive virtual environment-based eye tracker. The relationship between the gaze movement pattern of the test taker and the satisfaction evaluation result for the landscape image is analysed. Second, using the Convolutional Neural Networks (CNN)-based **Class Activation Map (CAM)** technique, a model for estimating the satisfaction evaluation result is constructed, and the gaze pattern of the test taker is derived. Third, we compare and analyse the similarity between the gaze heat map derived through the immersive virtual environment-based gaze tracker and the heat map generated by **CAM**. This study suggests the applicability of urban environment technology and deep learning methods to understand landscape planning factors that affect urban landscape satisfaction, resulting from the three-dimensional and immediate visual cognitive activity.

1. BACKGROUND

1.1 Research Background

Researchers define landscape as the visual interaction between humans and nature (Westerink et al., 2017; Tress and Tress, 2001; Lynch, 1960). In other words, this means that looking at the urban landscape means recognizing and recognizing the information provided by the urban landscape through the visual organ. However, most of the landscapes discussed in previous studies strongly implied the meaning of looking at a specific object far away from the observer. Therefore, it is somewhat far from the sense of the landscape that we unconsciously experience in our daily life (Im, 1988).

On the other hand, people's perception of the urban environment affects how they use and navigate it. Therefore, understanding these people's perceptions will be helpful for planning, designing, and managing urban spaces (Rossetti et al., 2019). Also, considering that most of our perception of the external environment occurs through the visual organ, understanding the process of people's perception of the urban landscape will provide insight into creating a better urban environment.

Improving people's visual experience of urban landscapes has become very important in urban planning and design. Therefore, the visual effect of the urban landscape has become an essential consideration in urban planning (Tveit, 2009). Modelling and predicting an individual's subjective preference for urban landscape can be a handy tool to increase the level of the urban landscape.

Knowing the urban landscape elements related to people's visual preference for the urban landscape and the interrelationship between each aspect can be a ground-breaking proposal for urban design (Im, 2009). In other words, modelling and predicting an individual's subjective preference for urban landscape can be a handy tool to increase the level of the urban landscape. The existing landscape image preference model paid attention to the correlation between the visual perception behavior of landscape elements and landscape preference. However, although the traditional survey method can intuitively and easily know an individual's subjective opinion, there is a concern that the researcher's intention may be included or interpreted during the survey and performance process. Therefore, to overcome the limitations of the existing landscape preference model, it is necessary to study the landscape elements that people see (what do they see?) and the time they stay in the landscape elements (how long do they stare?).

Meanwhile, computer vision technology has recently achieved many possibilities and achievements in visual recognition research. The computer vision method attracts attention in many landscapes preference studies because it can control researchers' intervention in the traditional landscape image analysis process to the maximum and minimize the input of time and human resources through automation. However, there are criticisms that research using such computer vision technology cannot ultimately bridge the gap between computers and humans, and there are questions about whether the results and interpretations are reliable.

1.2 Research Aims

The purpose of this study is to examine the applicability of computer vision technology to the analysis of the relationship between landscape image preference and eye movement patterns on pedestrian paths in residential areas where most of the landscape experiences in the city occur. Human visual

* Corresponding author

cognitive behavior is not simply acquiring external information but a series of psychological processes that comprehensively judge and determine visible details (Balcetis and Dunning, 2006). Therefore, if we understand people's visual perception process, we can objectively understand and explain people's best landscape images. In this case, gaze movement provides a helpful basis for establishing a landscape image preference model. However, measuring and analysing the gaze movement requires time and human resources. This is because the movement of the gaze is a primary signal from the human body, and there is no method other than attaching a measuring device to the human body to measure it in real-time. The application of computer vision technology can be considered as an alternative to overcome these limitations. Existing computer vision technology allows the artificial intelligence model created by the researcher to classify images and search for objects automatically. However, traditional computer vision technology is complicated to understand how the model works and what it means, and there is a limit to understanding and applying the process of the computer vision model. However, CAM (Class Activation Model) has recently been proposed as a way to solve this problem. It helps to understand the AI model by schematically showing which part of the image the AI model can focus on. Therefore, in this study, the applicability of computer vision technology that can be interpreted as a means of overcoming the limitations of visual cognition-related research by using the existing gaze movement patterns is comparatively analysed and applied.

2. RELATED RESEARCH

2.1 Landscape image preference analysis model

The landscape change affects the public's perception of the changing landscape and people's well-being. The unity, complexity, order, and aesthetics of the landscape image are major influencing factors affecting the level and perception of the urban landscape (Zakerhaghighi). 2022). In addition, naturalness has a positive effect on landscape perception (Schroeder, 1987), and the background of landscape images (urban, suburban, and rural) and proximity to the city have a negative effect on landscape preference (Baldwin et al., 1996).

In general, since most of the visual landscape experience experienced by city dwellers occurs in the street, the landscape image people perceive from the streetscape is significant. Accordingly, many studies have measured the level of street space (Li et al., 2018; Harvey et al., 2015; Salesses et al., 2013). As the perceived level of streetscape increases, the attachment of residents increases (Low and Altman, 1992), and devotion to a place increases people's stay time (Hidalgo and Hernández, 2001). In addition, greenery is a critical component of urban landscapes. Furthermore, the higher the proximity to nature, the higher the landscape preference (It was preferred over the existing urban landscapes, including grass and mowed grass (Burgess et al., 1988; Kaplan, 2001).

On the other hand, in the landscape image, individual elements composing the landscape affect people's perception, but each aspect collectively forms a landscape image or atmosphere and affects people's perception. The enclosure forms a space, which is an essential landscape element in determining the spatial scale of the human scale. In particular, visually sealed streets provide more opportunities for outdoor activities by providing pedestrians with a sense of safety (Rapoport, 2016; Owens, 1993; Alexander et al., 1977). In addition, the spatial composition and physical characteristics of landscape features can significantly affect the perception of fear and danger, and the spatial enclosure formed by surrounding plants can considerably affect the safety perception of park users (Baran et al., 2018). The continuity of the street is the ratio of the edge of the street that intersects with the building, and the complete and continuous facade of the building forms the order of the street and a lively landscape atmosphere (Harvey, 2014; Gehl and Svarre, 2013; Ewing and Handy, 2009). In addition, the perceived tree canopy density in the street affects people's stress, and the higher the density of green areas, the safer people perceive it to be (Li et al., 2015; Jiang et al., 2015).

2.2 Immersive virtual environments as tool of presentation

Various methods have been used to study the correlation between image recognition and personal preference (Sreetheran and Van Den Bosch, 2014). Most surveys and interviews were used to investigate people's perceptions of the urban environment (Madge, 1997; Jorgensen et al., 2007), indicating preferred regions on a map (Nasar et al., 1993). The survey method is the most used in social science research. It enables in-depth reasoning and analysis of the factors affecting residents' environmental perception and the relationship between each element. However, these survey methods are relatively complicated and expensive (Maruthaveeran and van den Bosch, 2014). An online-based survey method has recently been used as an alternative to the traditional paper-based survey method. The online survey does not require a separate space and has the advantage that anyone can get the questions the researcher wants regardless of location. However, compared to the face-to-face survey method, the online-based survey method may give an unintended response if the respondent has little understanding of the survey question. In addition, analysis of the collected data and interpretation of the results may be limited because there is room for response or non-response (Han and Lee, 2021).

Visualization and VR emphasize the visual sense as a communication tool in common (Portman et al., 2015), and simulating reality can be the core of VR technology. In addition, if VR is used for design purposes, it can be extended to a greater extent based on "real" simulations or cloning (Paes et al., 2017). In particular, immersive VR has received attention in many studies because it provides a high sense of reality (Bystrom et al., 1999, Schroeder, 2008) and provides presentations by visualizing experiences in dangerous and impossible situations direct. In addition, the heightened reality reproducibility of this immersive virtual environment technology makes immersive VR a means to treat anxiety disorders such as phobias in fields such as behavioural therapy. In addition, safety-related studies analysed the behaviour of people in emergencies such as fire (Gamberini et al., 2003) and aviation safety (Buttussi and Chittaro, 2017) in an immersive VR environment and recently at construction sites (Sacks et al., 2013), floods (Zaalberg and Midden, 2013), and tunnel accidents, such as accidents and large-scale disasters or large-scale disasters, measured and responded to individual behaviours were studied (Kinateder et al. ., 2015, Mühlberger et al., 2015)

2.3 Eye-tracking based landscape preference survey

Gaze is a primary means of communication because it can express emotions, feelings, and intentions (Lund, 2007) and provides insight into problem-solving, reasoning, mental imagery, and exploration strategies (Jacob and Karn, 2003). In particular, since the human gaze is closely related to human psychology, attention was paid to the human gaze in various fields such as medical/clinical evaluation, driver alert monitoring, consumer market analysis, and behavioural pattern analysis of disability or disease (Cazzato et al. 2020). Eye-tracking technology is a set of methods and techniques that measure where a person sees, what he sees, and how long his gaze stays at a specific point by measuring eye movement (Mele and Federici, 2012).

Researchers often use the terms eye tracking and eye tracking as synonyms, but there are some differences (Lund, 2007). Specifically, eye-tracking means measuring eye movement and activity, and gaze tracking means analysis of eye-tracking data related to a head or a visual scene (Lund, K., 2007). Eye-tracking technology detects the presence and position of the human eye in the image and measures activities such as gaze movement, eye blinking, fixation, and intermittent movement according to the time landscape. Through this, it is possible to infer a person's psychological state. As a helpful tool, it is a valuable technique to obtain public feedback for better landscape planning and management through people's perception of the landscape and the observation process (Zhou et al., 2022). In addition, tracking eye movement provides quantified information on gaze distribution that can understand human perception and cognition (Holmqvist et al., 2011; Duchowski, 2017). Therefore, many studies focused on the relationship between landscape elements and landscape recognition according to visual stimuli by introducing eye-tracking technology for landscape recognition and evaluation (Ebenberger and Arnberger, 2019). Most of the urban landscape consists of green areas, including buildings and trees, and an individual's level of Nature Relatedness (NR) is closely related to eye movement in outdoor spaces. Additionally, individual variances exist for NR because people with lower NR scores spend more time looking at buildings than trees (Chen et al., 2022). In addition, the lack of green space due to urban development activities paradoxically emphasized the importance of the forest landscape. As a result of studying the relationship between the forest landscape and preference through gaze movement, the satisfaction preference was relatively lower in the place where eye movements were more frequent. It was low.

In addition, spatial perception factors affect visual behavior, and satisfaction differs according to many indicators of forest recreation space (Zhang et al., 2022). On the other hand, spatial familiarity is related to human landscape perception, and eye movements change during a map-based search in familiar or unfamiliar urban environments. Spatial familiarity was measured through eye-tracking technology, and there was a significant relationship between gaze movement and spatial closeness through eye movement analysis. Eye-tracking technology helps detect spatial closeness (Liao et al., 2022). In addition, it was found that the social and spatial factors of the street space had a significant effect on the pedestrian's visit, and the gaze of the pedestrian stayed a lot at the edge of the street through the eye tracker. This means that improvement of the roadside environment can contribute to pedestrian activation and spatiality (Zhou et al., 2022).

2.4 Computer Vision and Visual Explanations from deep learning

Previous studies on human cognition have mainly focused on individuals because human understanding is highly dependent on individual experiences. Likewise, personal experiences play a significant role in landscape perception, and the spatial properties of landscapes can limit the co-production of knowledge operating across scales and sectors (Wei et al., 2022). However, advanced computer vision and deep learning methods enable easy mapping and analysis of the human perception of the urban landscape, helping to understand the relationship between human perception and perception. Computer vision research using deep learning deals with techniques such as object detection, semantic segmentation, and image classification and helps to explore the correlation between landscape and human perception.

However, suppose a deep learning method is used to distinguish people's preferences and dislikes for images. In that case, it will be possible to classify images and infer the overall picture that people prefer based on the classified images. However, the existing deep learning method can approximate human-like or desired results, but it was difficult to explain to the model because the basis for the results was unknown. However, Class Activation Mapping (CAM) enables visualization to know the basis for the result by adding location information in the image. Specifically, CAM applies Global Average Pooling (GAP) of Convolution Neural Network (CNN) so that CNN can learn object localization so that you can see the heat map of specific classes. It is a visualization technique that allows the researcher to check which part of the image the network saw and judged through this (Zhou et al., 2016). In addition, since it visualizes the heat map without supervised learning, the researcher does not need to write separate learning materials, thereby preventing time wastage.

2.5 Research Differences

The eye-tracking technique is a helpful tool for understanding human perception of the external environment. In particular, the high-resolution Head Mount Display (HMD) helps to understand the advanced human recognition system by enhancing the sense of immersion in virtual reality and enabling the capture of human perception according to experience. However, research on human perception based on eye trackers has several difficulties. First, many experiment participants are required to understand the relationship between human perception and the external environment through the data obtained through the eye-tracking experiment, which means time and cost consumption. Second, the eye tracker experiment must comply with the established protocol and strict bioethics because it collects information from the human body. Finally, since there is a difference between human cognitive ability and the resulting environment perception, making a predictive model takes a lot of effort.

On the other hand, the deep learning method is attracting attention as a valuable method of understanding the mechanism of human recognition. In particular, computer vision technology classifies images and searches for objects to understand the correlation between environment and perception. In addition, the automated algorithm can simultaneously process a large amount of information to produce new recognition information. Therefore, we want to verify whether an interpretable deep learning method can replace the existing eye-tracking process. Specifically, the purpose of this study is to investigate whether the limitations of the eye tracker preferred in the existing human recognition research can be applied as an efficient and objective analysis tool through the advanced deep learning method. The question of this study is, “Can the deep learning method replace the experience-based human recognition mechanism of the immersive virtual environment-based eye tracker?” And at the same time, it becomes the differentiation of the study.

3. METHODOLOGY

3.1 Research scope and material

This study analyses landscape images observed from pedestrian paths in residential areas in Seoul. The pedestrian path in the residential area is chosen because we routinely experience the urban landscape in the street space around the residential area. The landscape evaluation image is a Google Street View (GSV) image provided by the Google API service. Traditionally, in studies related to landscape preference, photos were taken by the researcher have been used. But the image taken by the researcher may reflect the subjective intention of the individual researcher in analysis and interpretation. Therefore, we selected GSV that was taken mechanically and did not reflect the personal preference of the researcher.

Meanwhile, the GSV image is taken on the vehicle's roof, slightly higher than the human line of sight. However, when we experience a city landscape, our gaze is fixed at a certain level or higher rather than the floor, so it is assumed that a person's eye and the eye of the GSV image coincide to some extent. In addition, we randomly sampled and selected the acquired GSV images for use in landscape preference studies (Figure 1).

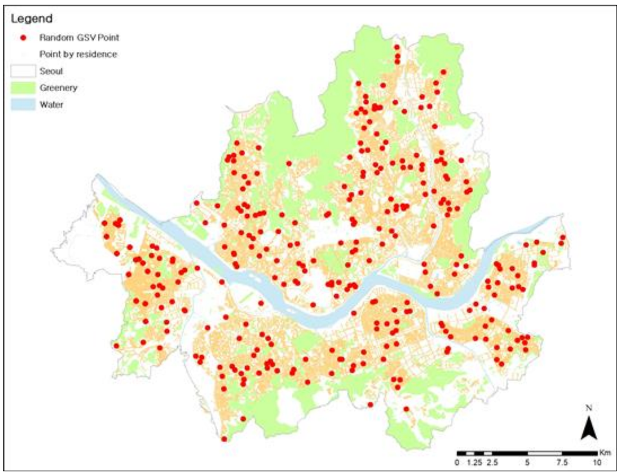

Figure 1. Random sampling GSV point

3.2 Immersive virtual environment evaluation process

Finally, we recruited 30 students majoring in urban engineering at Hanyang University's Graduate School of Urbanism. We used SNS to recruit participants in an immersive virtual environment-based landscape preference survey. Finally, we recruited 30 students majoring in urban engineering at Hanyang University's Graduate School of Urbanism. Since landscape evaluation is an evaluation that targets visual images, individual ability to see is essential. Therefore, we excluded students with corrected visual acuity of 0.6 or less from among the recruits. In addition, students who complained of dizziness during the experiment, which may harm the survey results, were excluded from the survey evaluation results. At the same time, we excluded students who complained of heterogeneity or dizziness of the HMD device in the immersive virtual environment, judging that they could harm the landscape preference questionnaire. Finally, we used the results of the landscape preference evaluation of 13 males and 12 females, a total of 26 (Table 1).

Age	Male	Female	Total
20~29	7 (28.0%)	8 (32.0%)	15 (60.0%)
30~39	5 (20.0%)	4 (16.0%)	9 (36.0%)
40~49	1 (4.0%)	0 (0.0%)	1 (4.0%)
Total	7 (52.0%)	7 (48.0%)	7 (100.0%)

Table 1. IVR experiment participant characteristics

Humans obtain information from the external environment through the five senses: sight, hearing, smell, hearing, and touch. Therefore, because people perceive the external environment through various sensory organs, people can react sensitively to the external environment under certain circumstances. Thus, the experimental environment was conducted in a noise-blocked laboratory so that the subject could focus on the immersive virtual environment screen, and the issue was carefully observed during the experiment.

We used HTC's VIVE-PRO head mount as a device to create an immersive virtual environment. It provides a resolution that offers immersive goggles that real people visually experience. In addition, the video displayed on the VR HDM is a panoramic video, which was confirmed as a distorted video on the monitor screen. However, in VR HDM, it is projected as an image on a curved surface and is visually recognized in the same way as a natural human gaze. In addition, the HTC device used in this study is equipped with a gaze tracking function to record the frequency and movement of people's gazes in real-time and support mapping. This allows you to monitor people's gazes, analyze gaze movement patterns, and generate visual attention heat maps.

In the protocol of this experiment, a landscape preference survey was conducted by randomly selecting gender and age, and subjects were allowed to experience panoramic images of residential areas through a VR screen. Afterward, the researcher directly investigated the issue for 11 evaluation items and collected data simultaneously. Finally, for the sake of the accuracy of the experiment, the experiment was conducted after 2 hours of waking up.

Figure 2. IVR based landscape preference experiment

3.3 Measurement of visual attention using eye-tracker

A sensor that tracks the movement of a person's pupil is attached around the Head Mount Display (HMD) of the HTC-VIVE device, an immersive virtual environment device. This sensor tracks a person's gaze and the pupil's size and movement to measure the eye's length and frequency. When a person's look stays for a long time, it means that interest or attention is high in the image area, which is caused by the unconscious reaction of the person. Therefore, since the eye-tracking data connect the relationship between the external environment and perception, it can be used as an objective and quantitative basis to understand the mutual relationship between people's perception of the domain.

In the experiment, the evaluator (experiment participant) experiences 15 different landscape images projected on the HMD, an immersive virtual environment presentation medium. At this time, the researcher orally conducts a questionnaire about each landscape image to the experimental participants and records the responses to the questions. At this time, the sensor attached to the HMD measures the movement of the human eye, and the measured data is automatically recorded on the computer. In addition, the recorded information automatically records the location information based on the coordinates on the mounted image. After the experiment, the researcher sets the Area of Interest (AOI) in the evaluation image. AOI work is a semantic segmentation work that selects individual elements existing on an image. When the AOI work is completed, it is combined with the information measured by the existing eye-tracking sensor. The time and frequency of the person's gaze on the image element and the order of eye movement are obtained as result values.

3.4 Deep learning approaches for image classification

Deep learning is a concept included in machine learning and is a learning method that operates through a multi-layer neural network. Deep learning learns by stacking a separately trained hidden layer between neural networks, and this structural characteristic enables effective learning even with a small amount of data. These structural characteristics enable effective learning even with a small amount of data. However, even if deep learning performance is good, it is challenging to interpret deep-learning results (Chen et al., 2014).

In general, a predictive model based on a built model must be able to explain the reason for the prediction to determine the authenticity of the model and its reliability of the model. In particular, image classification models used in computer vision have been challenging to explain why AI makes predictions. Suppose you can define the image classification model by AI. In that case, you can not only increase the reliability of the model built by the researcher but also interpret the results, which helps to understand the predictive model. To this end, in this study, Class Activation Mapping (CAM), a concept of an interpretable deep learning visualization technique, is applied to the model to visualize the landscape preference classification model built in the study. The CAM indicates which image parts will lead the CNN to the final classification decision. Additionally, it is possible to compare and understand the attention span recognized by humans and computers by comparing the result of displaying the attention region for image classification in the deep learning-based image classification model with the visual attention region based on the immersive virtual environment.

4. ANALYSIS

4.1 Multilevel Ordinal Logistic Regression Results

The relationship between individual landscape evaluation items suggested in previous studies and the overall landscape evaluation score was compared and analysed. Specifically, 15 randomly selected urban landscape images for each individual were evaluated for landscape evaluation. Therefore, a multi-level analysis should be applied, and ordinal logistic regression analysis should be used since questionnaire evaluation was rated from 1 to 5 points.

First, ICC (Intra-class correlation) value, explanatory power by level, was found to be 0.0000. That is, it is a reference value for judging the influence on the overall model at the individual level. In the case of the VR-based evaluation questionnaire, it is 0.00%, meaning that there is no difference between individuals. In other words, individual preferences do not affect the landscape preference in the immersive virtual environment. Next, the AIC and BIC values were checked to compare the model's overall fit. The AIC and BIC values are 522.2801 and 585.1110, respectively, indicating the model has a good fit.

The effect of detailed landscape evaluation items on overall landscape satisfaction was analysed. Individual age characteristics, gender, comfort, friendliness, naturalness, beauty, regularity, individuality (characteristics), cleanliness, artificiality, and safety evaluation items were found to significantly affect landscape satisfaction and overall satisfaction with the residential street landscape. Specifically, it was said that as the age increased, they were generally satisfied with the residential landscape. Furthermore, men gave a more positive evaluation of overall landscape satisfaction than women. In addition, it was found that as the detailed evaluation indexes of comfort, friendliness, naturalness, beauty, regularity, individuality, cleanliness, and safety increased, it positively affected the overall landscape evaluation. On the other hand, as the artificiality, which is a negative evaluation index, decreases, the comprehensive landscape evaluation changes positively.

variable		VR Survey Assessment		
Personal characteristics	Gender	-0.593	**	-2.20
	Age	0.026		0.80
Landscape preference items	Comfort	0.385	**	2.34
	Friendliness	0.151		1.04
	Harmony	0.458	***	2.68
	Natural	0.591	***	3.67
	Beauty	0.992	***	5.06
	Regularity	0.001		0.01
	Personality	0.268	**	2.07
	Cleanliness	0.763	***	4.82
	Artificial	-0.042		-0.37
	Safety	1.329	***	7.33
Obs.		375		
Wald chi^2 (12)		205.03	***	
/cut 1		6.8895		
/cut 2		11.2784		
/cut 3		16.2604		
/cut 4		217458		
ICC		0.0000		
AIC		522.2801		
BIC		585.1110		

*** p<0.01, ** p<0.05, * p<0.1

Table 2. Multilevel Ordinal Logistic Regression

4.2 Deep learning-based image classifier

Image classification is to classify image preferences based on the unique characteristics of the image, and it is to organize landscape images based on the results evaluated in the immersive virtual environment below. As a result of multi-step analysis, it was found that there was almost no individual difference in the images. This means that there is no difference between the 15 street images evaluated by 25 participants and the results of 1 photo by 375 people each. Therefore, it is possible to build a deep learning-based image classification model based on the evaluation results of a small number of photos. A landscape image classification model was constructed using the image model learning technique.

First, a pre-processing task was performed to compose the data with the correct answer, training, and evaluation sets. At this time, each image was labelled with five preference values. In addition, in the case of a multi-classification model, the correct answer data needs to be encoded. We encode using the function provided by TensorFlow. Keras. After data pre-processing was completed, a full-scale model was designed. The data was made into two-dimensional data of 28 pixels and 28 pixels, spread out in one dimension using the Flatten function in the input layer. The hidden layer was arbitrarily stacked. In the output layer, the activation function is suitable for multiple classifications, and the softmax function is used. Next, the model's performance was improved by adjusting hyperparameters such as the optimizer and loss, and the model was trained in earnest. As a result of evaluating the training data on the model, the loss value steadily decreased, and it converged to 27.5% from about 1 to 2 epochs, indicating that the model's performance was not excellent(Figure 3, Figure 4).

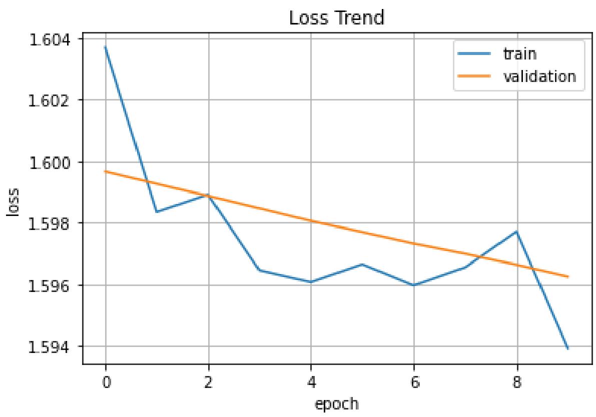

Figure 4. Pre-train data set loss rend

The model's performance can be addressed by increasing the training data and improving the hyperparameter values. In future research, we decided to build an enhanced image classification model and performed image classification visualization that can be interpreted with the already created classification model.

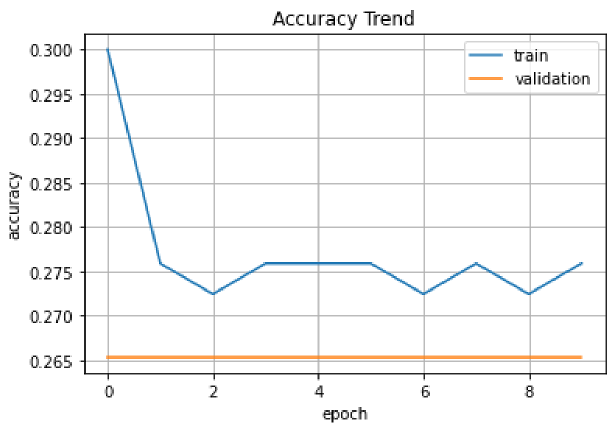

Figure 5. Fitness of fit of the learning model

4.3 CNN based Class Activation Map

CAM is a visualization technique that can determine where the CNN model focuses on the input image. The working principle is that it is not a map with the exact resolution as the input. Hence, it sees the upscaled information and is generally calculated using the information of the last layer (gap, global average pooling). As a result, we can extract the activation map for each class.

First, a picture that matches the preference predicted by the image classification model and the preference evaluated by the evaluator in the immersive virtual environment was selected, and the activation map was visualized. The following figure shows the landscape image the evaluator responded with satisfaction 1 and the image that the evaluator does not prefer. At the same time, the image classification model has the same preference as the value evaluated by the evaluator as satisfaction 2 for the image.

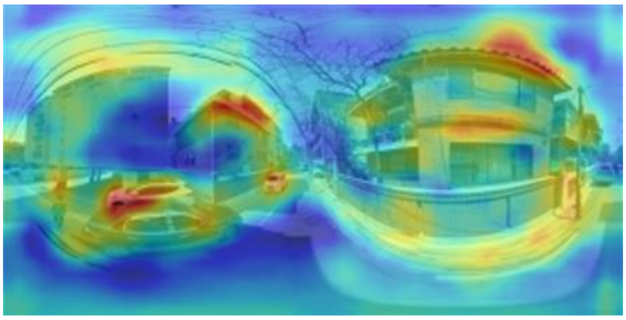

Figure 6. Class Activation Map

4.4 Eye-tracking Multilevel Ordered Logistic Regression Results

The results of gaze movement patterns measured in the immersive virtual environment are as follows. First, for comparison with the class activation map, the gaze movement of the evaluator was confirmed with the same picture.

Figure 7. Area of interest

As a result of eye tracking, it was found that the subject was looking at the front part of the road. This is because a protocol similar to actual walking was presented in the landscape preference experiment based on the immersive virtual environment. However, what is unique is that the heat map is visualized on the vehicle and building on the left. Compared with the result of the class activation map, it was consistent in some areas. Still, it can be seen that the eye-tracking result in the immersive virtual environment and the attendance area of the image classification model do not match in general.

5. CONCLUSION

Knowing the visual factors people prefer to see in the city, and our cities can expect a better environment. However, many researchers have focused on the objective quantification of factors influencing people's preferences. On the other hand, humans obtain most of the external environmental information through our visual organs, which is the basis for our perception of the urban environment. It was discussed in a study dealing with visual preference because it quantifies and presents what to see and how-to see-through gaze movement. This study surveyed and analysed the preference for horizontal landscapes in an immersive virtual environment. At the same time, the image area that people mainly viewed was analysed by measuring the gaze movement of a person using the eye tracker technology. However, although the eye-tracking method is beneficial for understanding the mechanism of recognition of the urban environment, it requires a limited experimental environment and sophisticated experimental protocol, which consumes time and money.

On the other hand, the developed deep learning technique is very effective in classifying images with a small amount of data, so many researchers have paid attention. Therefore, in this study, we create an image classification model through deep learning techniques, visually confirm the basis for classifying images by AI using class activation map visualization and compare them with human eye movement patterns in an actual immersive virtual environment. Furthermore, the applicability of the deep learning image classification model was verified.

In the case of the deep learning-based analysis model, the performance of the image classification model was poor due to the lack of training images and inappropriate parameter values. Next, the prediction result of the image classification model and the visual attention area of the same image as the result of the immersive virtual environment-based questionnaire were compared and analysed. As a result of comparative analysis, it was confirmed that, although the reflection of gaze attention and the activated area were consistent in some areas, they did not match overall.

This study verified the possibility of an image classification prediction model of deep learning techniques that can present a prediction model with only a small amount of data. However, the appropriateness of the actual deep learning image classification model was not identified through the small number of images and the optimized input value. Therefore, further research is needed based on future performance improvement of the image classification model.

REFERENCES

Alexander, C., 1977. *A pattern language: towns, buildings, construction*. Oxford university press, New York.

Balcetis, E., and Dunning, D., 2006. “See what you want to see: motivational influences on visual perception”, *Journal of Personality and Social Psychology*, 91(4): 612-625.

Baran, P. K., Tabrizian, P., Zhai, Y., Smith, J. W., and Floyd, M. F., 2018. “An exploratory study of perceived safety in a neighborhood park using immersive virtual environments”, *Urban Forestry & Urban Greening*, 35: 72-81.

Burgess, J., Harrison, C. M., and Limb, M., 1988. “People, parks and the urban green: a study of popular meanings and values for open spaces in the city”, *Urban Studies*, 25(6): 455-473.

Buttussi, F., and Chittaro, L., 2017. “Effects of different types of virtual reality display on presence and learning in a safety training scenario”, IEEE Transactions on Visualization and Computer Graphics, 24(2): 1063-1076.

Bystrom, K. E., Barfield, W., and Hendrix, C., 1999. “A conceptual model of the sense of presence in virtual environments”, Presence: Teleoperators & Virtual Environments, 8(2): 241-244.

Cazzato, D., Leo, M., Distante, C., and Voos, H., 2020. “When i look into your eyes: A survey on computer vision contributions for human gaze estimation and tracking”, *Sensors*, 20(13): 1-42.

Chen, B., Gong, C., and Li, S., 2022. “Looking at buildings or trees? Association of human nature relatedness with eye movements in outdoor space”, *Journal of Environmental Psychology*, 80:1-9.

Chen, Y., Lin, Z., Zhao, X., Wang, G., and Gu, Y., 2014. “Deep learning-based classification of hyperspectral data”, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6): 2094-2107.

Ebenberger, M., & Arnberger, A. (2019). Exploring visual preferences for structural attributes of urban forest stands for restoration and heat relief. *Urban Forestry & Urban Greening*, 41, 272-282.

Ewing, R., and Handy, S., 2009. “Measuring the unmeasurable: Urban design qualities related to walkability”, *Journal of Urban Design*, 14(1): 65-84.

Gamberini, L., Cottone, P., Spagnolli, A., Varotto, D., and Mantovani, G., 2003. “Responding to a fire emergency in a virtual environment: different patterns of action for different situations”, *Ergonomics*, 46(8): 842-858.

Gehl, J., and Svarre, B., 2013. *How to study public life (Vol. 2)*. Washington, DC: Island press, Washington, DC.

Guo, Y., Liu, Y., Lu, S., Chan, O. F., Chui, C. H. K., and Lum, T. Y. S., 2021. “Objective and perceived built environment, sense of community, and mental wellbeing in older adults in Hong Kong: A multilevel structural equation study”, *Landscape and Urban Planning*, 209: 1-12.

Han, J. and Lee, S. 2021. “Application and Verification of Streetscape Evaluation Methods Based on Immersive Virtual Reality Environment: Focused on Streetscape in Residential Areas of Seoul, Korea”, pp. 198-212. Jeju Halla Convention Center, Jeju-do, Oct. 14.

Harvey, C., 2014. *Measuring streetscape design for livability using spatial data and methods*. The University of Vermont and State Agricultural College.

Harvey, C., Aultman-Hall, L., Hurley, S. E., and Troy, A., 2015. “Effects of skeletal streetscape design on perceived safety”, *Landscape and Urban Planning*, 142, 18-28.

Heath, G. W., Brownson, R. C., Kruger, J., Miles, R., Powell, K. E., and Ramsey, L. T., 2006. “The effectiveness of urban design and land use and transport policies and practices to increase physical activity: a systematic review”, *Journal of Physical Activity and Health*, 3(s1): S55-S76.

Hidalgo, M. C., and Hernandez, B., 2001. “Place attachment: Conceptual and empirical questions”, *Journal of Environmental Psychology*, 21(3): 273-281.

Im, S., 1988. “A study on the landscape analysis and evaluation method: A phenomenological Approach”, *Journal of the Korean Institute of Landscape Architecture*,16(1): 43-53.

Jiang, B., Larsen, L., Deal, B., and Sullivan, W. C., 2015. “A dose–response curve describing the relationship between tree cover density and landscape preference”, *Landscape and Urban Planning*, 139: 16-25.

Jorgensen, A., Hitchmough, J., and Dunnett, N., 2007. “Woodland as a setting for housing-appreciation and fear and the contribution to residential satisfaction and place identity in Warrington New Town, UK”, *Landscape and Urban Planning*, 79(3-4): 273-287.

Kaplan, R., 2001. “The nature of the view from home: Psychological benefits”, *Environment and Behavior*, 33(4): 507-542.

Kinateder, M., Gromer, D., Gast, P., Buld, S., Müller, M., Jost, M., Nehfischer, M., Mühlberger, A., and Pauli, P., 2015. “The effect of dangerous goods transporters on hazard perception and evacuation behavior–A virtual reality experiment on tunnel emergencies", *Fire Safety Journal*, 78: 24-30.

Li, J., Zhang, Z., Jing, F., Gao, J., Ma, J., Shao, G., and Noel, S., 2020. "An evaluation of urban green space in Shanghai, China, using eye tracking", *Urban Forestry & Urban Greening*, 56, 1-11.

Li, X., Zhang, C., and Li, W., 2015. "Does the visibility of greenery increase perceived safety in urban areas? Evidence from the place pulse 1.0 dataset", *ISPRS International Journal of Geo-Information*, 4(3): 1166-1183.

Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q., and Zhang, W., 2015. "Assessing street-level urban greenery using Google Street View and a modified green view index", *Urban Forestry & Urban Greening*, 14(3): 675-685.

Liao, H., Zhao, W., Zhang, C., Dong, W., and Huang, H., 2022. "Detecting individuals' spatial familiarity with urban environments using eye movement data", *Computers, Environment and Urban Systems*, 93, 1-12.

Low, S. M., and Altman, I., 1992. "Place attachment: A conceptual inquiry", *Human Behavior & Environment: Advances in Theory & Research*, 12: 1–12.

Lund, K., 2007. "The importance of gaze and gesture in interactive multimodal explanation", *Language Resources and Evaluation*, 41(3): 289-303.

Lynch, K., 1964. *The image of the city*, MIT press.

Madge, C., 1997. "Public parks and the geography of fear", *Tijdschrift voor economische en sociale geografie*, 88(3): 237-250.

Mele, M. L., and Federici, S., 2012. "Gaze and eye-tracking solutions for psychological research", *Cognitive Processing*, 13(1): 261-265.

Mühlberger, A., Kinateder, M., Brütting, J., Eder, S., Müller, M., Gromer, D., and Pauli, P., 2015. "Influence of information and instructions on human behavior in tunnel accidents: a virtual reality study", *Journal of Virtual Reality and Broadcasting*, 12(3): 1-13.

Nasar, J. L., Fisher, B., and Grannis, M., 1993. "Proximate physical cues to fear of crime", *Landscape and Urban Planning*, 26(1-4): 161-178.

Owens, P. M., 1993. "Neighborhood form and pedestrian life: Taking a closer look", *Landscape and Urban Planning*, 26(1-4): 115-135.

Peperkorn, H. M., Diemer, J., and Mühlberger, A., 2015. "Temporal dynamics in the relation between presence and fear in virtual reality", *Computers in Human Behavior*, 48: 542-547.

Portman, M. E., Natapov, A., and Fisher-Gewirtzman, D., 2015. "To go where no man has gone before: Virtual reality in architecture, landscape architecture and environmental planning. Computers, *Environment and Urban Systems*, 54: 376-384.

Paes, D., Arantes, E., and Irizarry, J., 2017. "Immersive environment for improving the understanding of architectural 3D models: Comparing user spatial perception between immersive and traditional virtual reality systems", *Automation in Construction*, 84: 292-303.

Rapoport, A., 2016. *Human aspects of urban form: towards a man—environment approach to urban form and design*, Elsevier.

Rossetti, T., Lobel, H., Rocco, V., and Hurtubia, R., 2019. "Explaining subjective perceptions of public spaces as a function of the built environment: A massive data approach", *Landscape and Urban Planning*, 181: 169-178.

Salesses, P., Schechtner, K., and Hidalgo, C. A., 2013. "The collaborative image of the city: mapping the inequality of urban perception", *PloS One*, 8(7): 1-12.

Schroeder, R., 2008. "Defining virtual worlds and virtual environments", *Journal for Virtual Worlds Research*, 1(1):1-3.

Sklenicka, P. and Molnarova, K., 2010. "Visual perception of habitats adopted for post-mining landscape rehabilitation", *Environmental Management*, 46(3): 424-435.

Sreetheran, M., and Van Den Bosch, C. C. K., 2014. "A socio-ecological exploration of fear of crime in urban green spaces–A systematic review", *Urban Forestry & Urban Greening*, 13(1): 1-18.

Tress, B., and Tress, G., 2001. "Capitalising on multiplicity: a transdisciplinary systems approach to landscape research", *Landscape and Urban Planning*, 57(3-4): 143-157.

Tveit, M. S., 2009. "Indicators of visual scale as predictors of landscape preference; a comparison between groups", *Journal of Environmental Management*, 90(9): 2882-2888.

Wei, J., Yue, W., Li, M., and Gao, J., 2022. "Mapping human perception of urban landscape from street-view images: A deep-learning approach", *International Journal of Applied Earth Observation and Geoinformation*, 112: 1-11.

Westerink, J., Opdam, P., Van Rooij, S., and Steingröver, E., 2017. "Landscape services as boundary concept in landscape governance: Building social capital in collaboration and adapting the landscape", *Land Use Policy*, 60, 408-418.

Xi, C., Guo, Y., He, R., Mu, B., Zhang, P., and Li, Y., 2022. "The use of remote sensing to quantitatively assess the visual effect of urban landscape-A case study of zhengzhou, China", *Remote Sensing*, 14(1): 203.

Zhang, Z., Gao, Y., Zhou, S., Zhang, T., Zhang, W., and Meng, H., 2022. "Psychological cognitive factors affecting visual behavior and satisfaction preference for forest recreation space", *Forests*, 13(2): 1-24.

Zaalberg, R., and Midden, C. J., 2013. "Living behind dikes: mimicking flooding experiences", *Risk Analysis*, 33(5): 866-876.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., and Torralba, A., 2016. "Learning deep features for discriminative localization", *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2921-2929.

Zhou, S., Gao, Y., Zhang, Z., Zhang, W., Meng, H., and Zhang, T., 2022. "Visual behaviour and cognitive preferences of users for constituent elements in forest landscape spaces", *Forests*, 13(1): 1-22.