ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Spatial Perception and Optimization Strategy of Environmental Art Design Driven by Deep Learning

Bin Chen[1,*] *and Liang Tao*[1]

[1] Jinling Institute of Technology, Nanjing, Jiangsu, 211169, China

SUMMARY: *This study constructs a street spatial imagery elements dataset as the research object. A visual hyperbolic spatial model based on deep learning and supervised learning is established, as well as a street architectural landscape classification model ResNet and a street element analysis model DeepLabV3+. The models are used for feature extraction and semantic segmentation of street spatial imagery to improve the classification accuracy while preserving the detailed features, and to lay a data foundation for the subsequent calculation of Pearson correlation coefficients between street spatial imagery elements and visual perceptual experience. According to the calculation results, a total of 19 street spatial imagery elements in 6 categories are adjusted and optimized. The correlation of the elements in the positive perception dimension is increased to more than 0.9, that of the neutral perception dimension is increased to more than 0.8, and that of the negative perception dimension is decreased to less than 0.7, and the optimized spatial perceptual experience is biased in the positive direction.*

KEYWORDS: *visual hyperbolic space; ResNet; DeepLabV3+; Pearson correlation coefficient; spatial perception*

1 Introduction

Environmental art design is a modern design discipline with “space” and “interface” as the main research object, creating a more comfortable and beautiful living environment for human beings as the main design purpose. In the environment, “entity” and “space” are two important components, figurative and abstract complement each other, is an indispensable environmental art design of environmental elements and design objects [1, 2]. At the same time, environmental art design is the main design of people's living space, creating a “second nature” for human beings to “shrink and simulate” is the core purpose of environmental art design. The influence of environmental art design on space is manifested in three aspects: space level, space performance and space performance. From the perspective of space level, environmental art space has the function of separating space and creating multi-dimensional space. From the perspective of spatial performance, environmental art design can play a role in increasing the function of space use and guiding people's lifestyle [3]. From the perspective of spatial expression, environmental art design has the obligation to combine artistic concepts with local folklore, continue the regional culture, display the humanistic atmosphere and artistic atmosphere, and better decorate the space [4, 5]. Therefore, fully exploring the spatial perception of environmental art design can further promote the expression of spatial design and the formation of new design concepts.

Although traditional environmental art design takes into account aesthetics and life needs,

*jinlingkeji211169@163.com
https://doi.org/10.65102/is2026004
Published: 30 April, 2026

with the diversification of human needs, traditional design methods expose defects, such as low space utilization and spatial sense, lack of consideration of the environment, low design efficiency, and lack of personalization due to the trend of homogenization [6-8]. With the wide application of artificial intelligence, deep learning (DL), as a machine learning method based on neural network architecture, has demonstrated its powerful ability in many fields. In the field of art creation and design, deep learning is also gradually playing an important role. Yu et al [9] (2025) used DL and spatial grammar to assess the visual perception of street space, and perceived the spatial quality of urban streets by extracting visual elements from streetscape images and analyzing them in association with the street guide elements, identifying beauty and bustle as key elements. Kang et al [10] (2020) introduced a DL-based 3D reconstruction technique for indoor environments, which, based on the advantages of the DL-based hierarchical pyramid structure and learnable parameters, is able to seamlessly integrate indoor and outdoor spaces. Wang et al [11] (2025) proposed a DL-based multi-view stereo framework for improving indoor 3D reconstruction techniques, DL by capturing the features of indoor spatial environments under untextured surfaces and dynamic lighting conditions. Wei et al [12] (2022) used the DL algorithm to analyze human perceptions of urban landscapes based on street view images, which can clearly capture spatial features of the landscape and map urban landscape perceptions, capable of spanning spatial scales and domains. Liu and Deng [13] (2025) used a convolutional neural network model to distinguish between AI-rendered and real interior spaces from interior design images, and could obtain up to 97% classification accuracy, which helps to strengthen interior space perception.

In addition, Sun [14] (2022) integrated various neural networks such as 3D spatial convolutional neural network, fuzzy neural network, and adversarial neural network, and proposed a DL algorithm for spatial layout optimization of interior design. Chen [15] (2023) developed an environmental landscape design and planning system based on computer vision and DL, where the system monitored carbon dioxide indoors and around buildings, combined with the ability of plants to absorb toxins, and realized landscape design in a virtual environment that reduced environmental pollution levels and optimized design solutions. Xu [16] (2025) proposed a DL-based optimization method for landscape design and ecological balance, in which the ecological needs and aesthetic values were balanced to promote the growth of landscape plants.

In this study, we construct a visual hyperbolic spatial model that imitates binocular physiological features, and quantify the relationship between spatial imagery element placement and visual perception experience through the computational conversion of visual perception distance and actual physical distance. Combining the street view picture dataset with the street building style classification model ResNet and the street element analysis model DeepLabV3+, deep learning and semantic segmentation of picture features are realized to categorize different street spatial imagery elements. Meanwhile, residual network and multiplicative up-sampling are introduced to ensure the feature extraction accuracy and segmentation effect of the model. After obtaining the street spatial imagery elements and visual perception dimensions, the Pearson correlation coefficient calculation method is used to study the correlation between the two and provide reference for spatial perception optimization.

2 Design optimization based on spatial perception and correlation of spatial elements

2.1 Spatial quantification of visual perception

2.1.1 Visual hyperbolic spatial modeling

Lüneburg's theory of binocular spatial perception is the basis for all subsequent models of visual hyperbolic space, which is a model of visual space with geometric properties of hyperbolic space derived from the experimental data of the “Alley Experiment” in conjunction with the physiological characteristics of binocular vision.

The research on visual space adopts a comprehensive geometric approach, and thus establishes the following geometric axioms about visual space:

1) Visual space is characterized as a metric space (calculable distance);

2) Visual space is a convex space, i.e., there exist any two points $P1$ and $P3$ on visual space, and the existence of a third point $P2$ on the line connecting the two of them is able to satisfy:

$$D(P_1, P_2) + D(P_2, P_3) = D(P_1, P_3) \tag{1}$$

3) Visual space is compact, and the axiom allows the assumption that visual space is continuous. This means that for any point $P1$ and any number ε, there exists another point $P2$ that satisfies:

$$0 < D(P_1, P_2) < \varepsilon \tag{2}$$

Lüneburg's model is simplified and improved by replacing Lüneburg's calculus-based metric of non-Euclidean geometry with a metric of Euclidean geometry to establish a basic model formulation of visually perceived hyperbolic space:

$$ds^2 = \overline{d(d)^2} + M^2\left(d\varphi^2 + \cos^2\varphi\right) \tag{3}$$

where d represents the visual perception distance (VD), s represents the visual perception size (VS), ϕ represents the horizontal angle of view (HA), θ represents the vertical angle of view (VA), and M represents the linear size factor $ds/d\varphi$. The formulas are transformed and simplified using geometric operations to obtain a general formula for the hyperbolic spatial visual perception distance d and the physical spatial distance D:

$$\frac{d}{D} = \frac{R_A}{R_A + D} \tag{4}$$

According to the “size invariance assumption”, the change in the perceived size of an object in visual space can be defined in relation to the distance from the observer in physical space:

$$\begin{aligned} s &= \frac{BS}{R_A + D} \\ B &= R_A + \delta \end{aligned} \tag{5}$$

where s is the visually perceived size, S is the true size of the object perceived at a distance of δ (which can also be interpreted as the true size of the object), D is the true distance of the object from the observer in the physical space, and R_A serves as the limiting distance in the visual space of the bounded hyperbolic space, which is often based on the experimental data fitting.

2.1.2 Environmental representations of visual space

If we take the perceived distance d as the independent variable and the perceived size s as the dependent variable, and bring in the "size-distance invariance assumption" formula of the physical space, it can be transformed into the visual spatial size derivation formula that is:

$$s = d\tan\phi = d\frac{S}{D} = dk \tag{6}$$

Equation (6) associates the size and distance metrics of visual space with those of physical space, from which some basic characterization information about visual space can be obtained. For example, if there is a reference building of size (height) S at a distance D from the observation point, and if one wants to build (or find) a building that looks (with a visual size s) half the size of the reference building at the same observation point, the simplest way to do this would be to reduce the size to half of what it was, or to double the distance from the observation point, i.e., to reduce the ratio of S to D to one-half of what it was. These ways are consistent with the visual size change characterized by equation (6), but at the same time the equation implies another approach - reducing the distance of its visual perception to half of its original size.

2.1.3 General representation of visuospatial features

Further, Eq. (6) can also reflect the interconversion relationship between visual space and physical space, where the perceived size of visual space is equal to the physical size if the perceived distance is equal to the physical distance, i.e., if the geometrical properties of visual space and physical space are the same (or at least in terms of the perceived distances), otherwise the ratio of the two is equal to the ratio of the distances in the corresponding space. This formula corresponds exactly to the phenomenon described by Emmert's Law (EL).

Emmert's law, also known as the law of size of the visual afterimage (where "afterimage" refers to the visual perceptual image). Emmert's law states that objects of the same size on the retina will appear different in physical size (magnitude) if they appear to be located at different distances. That is, objects with the same viewing angle are perceived to be larger in size the farther away they are from (d). A more obvious example of this is in the size comparison process: Equation (7) shows the comparison of the perceived sizes of different objects derived from Equation (6):

$$\frac{s_1}{s_2} = \frac{d_1 k_1}{d_2 k_2} \tag{7}$$

According to formula (7), two identical objects $(k1 = k2)$ are located at different distances from the observation point one near and one far, the distance of the distant object from the observation point is twice the distance of the near object from the observation point, then if the visual space and the physical space is the same, then $d1/d2 = D1/D2 = 1/2$, then at this time,

the size size of the visually perceived by the distant object is half the size of the visually perceived size of the near object. This is based on the visual space is the same as the physical space of Euclidean geometry (or Riemannian space with curvature of 1.0), if the visual space is non-Euclidean geometry, Lüneburg visual hyperbolic spatial model, for example, in the process of transforming from the Euclidean physical space, the further the distance of the greater the proportion of compression (i.e., the same physical distance from the point of observation in the visual space, the further the visual distance is smaller), then $d1/d2 > D1/D2$, at this time the size of the visual perception of distant objects will be greater than half of the size of the visual perception of near objects, the visual perception of the object compared to the results calculated in the Euclidean space appeared in the phenomenon of "enlargement". This phenomenon can be regarded as a basic characterization of the influence of the geometric properties of visual space on visual perception under the condition of visual hyperbolic space.

2.2 Supervised Learning Based Street Architectural Landscape Classification Model

2.2.1 Street building style classification model construction

The research in this chapter is based on the street view image dataset with specified labels, and the building functional style recognition of street view images to study the compliance of street view building style construction. Therefore, the techniques used in this paper are closely related to the supervised image classification task.

The image classification network model used in this paper is ResNet, which enables deeper network layers to ensure the accuracy of model training and extract deeper features of the image.

Residual learning is inspired by the fact that when a deep network is built by continuously piling up new network layers, model training will not be degraded if, in extreme cases, these added layers do not undergo the learning process and simply replicate the features of the shallow network. This solves the problem of training degradation when the number of network layers is constantly increased. Therefore through the idea of constant mapping (Im), residual learning is proposed to solve the degradation problem during network training.

Assuming that the input to the model is x and the features learned after passing through the network are notated as $H(x)$, the original mapping is represented as:

$$H(x) := F(x) + x \tag{8}$$

Then the residual can be expressed as:

$$F(x) = H(x) - x \tag{9}$$

The purpose of using residuals is because residual learning is relatively easy compared to direct learning of raw features, firstly the residual unit can be represented as:

$$y_l = h(x_l) + F(x_l, W_l) \tag{10}$$

$$x_{l+1} = f(y_l) \tag{11}$$

where x_l and x_{l+1} denote the inputs and outputs of the lth residual cell, respectively, noting that each residual cell typically contains a multilayer structure. F is the residual function,

which represents the learned residuals.

$$h_{x_l} = x_l \tag{12}$$

The constant mapping is represented using Eq. (13), where f is the RELU activation function. Based on the above introduction, the learning feature from shallow l to deep L can be derived as:

$$x_L = x_l + \sum_{i=l}^{L-1} F(x_i, W_i) \tag{13}$$

According to the chain rule, the gradient of the backpropagation process can be found to be:

$$\frac{\partial loss}{\partial x_l} = \frac{\partial loss}{\partial x_L} \cdot \frac{\partial x_L}{\partial x_l} = \frac{\partial loss}{\partial x_L} \cdot \left(1 + \frac{\partial}{\partial x_L} \sum_{i=l}^{L-1} \left(F(x_i, W_i)\right)\right) \tag{14}$$

The first factor $\frac{\partial loss}{\partial x_L}$ of Eq. represents the gradient of the loss function to reach L, and the 1 in parentheses indicates that the short-circuiting mechanism can propagate the gradient losslessly, whereas the other item of the residual gradient needs to go through the layer with the weights, and it is not passed directly. Therefore even though the residual gradient will be small, it will not be easy for the gradient to disappear and residual learning will be easier. When the residual value is 0, the stacked network layers are equivalent to realize a constant mapping, so that the performance of the network will not be degraded. It also makes the residual module learn new features based on the input features, thus having a better performance. Fig. 1 shows the residual learning module, which solves the degeneracy problem with a kind of short-circuit connection.

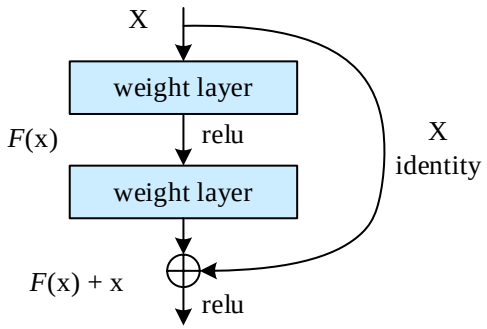


Figure 1: Residual Learning Unit

The two residual learning units of ResNet design Figure 2, which are structured for ResNet34 (Figure 2, left) and ResNet50/101/162 (Figure 2, right), are generally referred to as a “building block”. The right figure is also known as the “bottomleneck design”. Since ResNet allows the network to reach up to 162 layers, it is a good solution to the problem of network degradation caused by the deepening of the network layers. Therefore, ResNet can extract deeper image features, which improves the accuracy of the classification task. The main model used in this chapter for classification and recognition of street view building features is ResNet162.

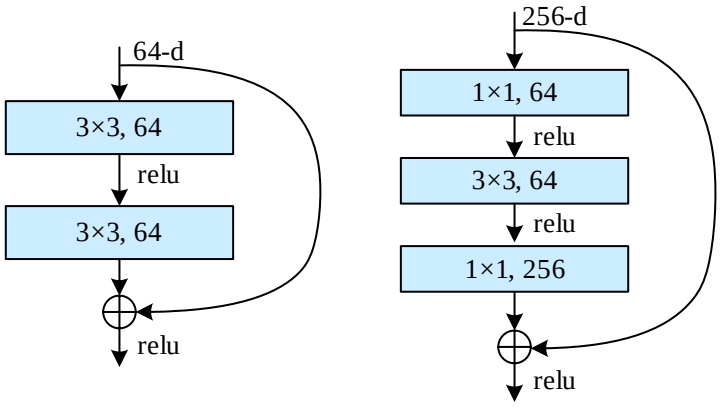


Figure 2: Two ResNet residual unit structures

2.2.2 Modeling of street element analysis

Another research component of this chapter is the analysis of images with the help of semantic segmentation tools for features such as elemental occupancy. Thus a segmentation technique like MaskRCNN where detection is the main tool, the main goal is to segment the instances. If the segmentation element is a person, MaskRCNN will segment different people as well. But for street view element analysis there is no need to segment out different people. Therefore, in this paper, DeepLabV3+, a codec-based semantic segmentation model, is chosen for segmentation and element analysis of street scene images.

DeepLabV3+ model is a typical semantic segmentation network and its segmentation accuracy is one of the highest models available. It combines the hollow convolutional pyramid pooling (ASPP) technique with the codec technique, using the encoder to extract the depth features of the image to achieve effective feature extraction, and the decoder to up-sample the depth feature map and fuse it with the shallow features, and use the shallow features to optimize the positional information that can not be recovered by the up-sampling, and ultimately get the result of the semantic segmentation. Figure 3 is an intuitive - representation of the null convolution technique. Mathematically, assuming that the input is x, the convolution kernel is w, the sampling rate (dr) is r, and the output of its cavity convolution is y, the convolution output is:

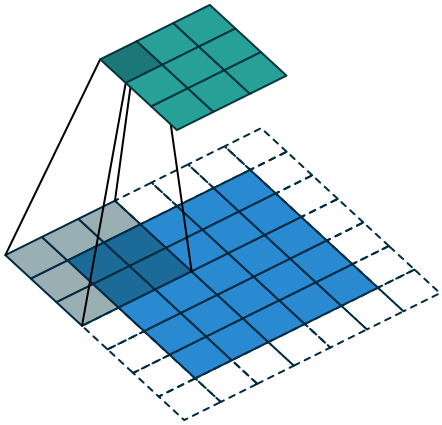
$$y[i] = \sum_k x[i + rk]w[k] \tag{15}$$

where x is the input feature map, y is the output feature map after cavity convolution, and k is the size of the convolution kernel. Compared to the general convolution kernel, the receptive field of the cavity convolution is extended to k_e:

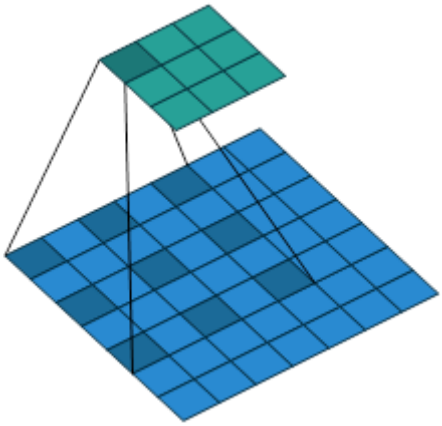
$$k_e = k + (k-1)(r-1) \tag{16}$$

When the sampling rate is 1, the convolution kernel of the null convolution is the ordinary convolution kernel. Because the ordinary deep convolutional neural network through the pooling layer will reduce the size of the image to increase the sensory field, but in the recovery of the semantic segmentation map will be used to upsampling to expand the size of the image, this process of reducing and then increasing the size of the process, there will be some information loss, and at the same time for the information of the small objects can not be

reconstructed, which have become the main problem of the semantic segmentation accuracy can not be further improved. Null convolution does not carry out pooling operation, which is somewhat similar to the form of VGG, by modifying the form of convolution kernel to make the feeling field increase, so as to make the feature information contained in the convolution calculation in a larger range, and then improve the semantic segmentation accuracy.



(a)Standard convolution (convolution kernel size 3×3)



(b)Empty convolution (convolution kernel size 3×3, sampling rate 2)

Figure 3: Empty Convolutional Architecture Structure

The deeplabV3+ feature extraction network model used in this paper is the Xception network, and ASPP utilizes a 1×1 convolution kernel to transform the output feature maps of the Xception network from 2049 dimensions to 264 dimensions, and then performs null convolution of the resulting feature maps using three 3×3 convolution kernels with different sampling rates. As the sampling rate increases, the role of the convolution kernel decreases, and when the size of the null convolution kernel is the same as the size of the output feature maps, the 3×3 convolution kernel plays the same role as the 1×1 convolution kernel. The feature extraction layer containing the ASPP is the encoder module of the DeepLabV3+ model.

For semantic segmentation, after the deep convolutional neural network extracts the features, the size of the feature map is very different from the size of the original image, so semantic segmentation generally uses up-sampling and fusion of shallow network features to restore the image size and improve the segmentation accuracy.

DeepLabV3+ network does not choose layer-by-layer up-sampling, it carries out segmentation by two 5-fold up-sampling, and ASPP is the merging of feature maps according

to channels. The deep feature maps are reduced to the same size as the shallow feature maps by 5-fold upsampling. In order to prevent the semantic segmentation resolution from being too low due to the large feature weights of the shallow feature maps, DeepLabV3+ does not choose to directly fuse the deep feature maps with the shallow feature maps, but instead uses a 1×1 convolution kernel to downscale the shallow feature maps, at which time the weights of the deep feature maps are larger than those of the shallow feature maps, and then performs the feature fusion, which avoids the problem caused by the large weights of the shallow feature maps, and then performs the feature fusion, which results in a lower semantic segmentation resolution. Then feature fusion is performed, thus avoiding the phenomenon of increasing network training difficulty due to the large weight of the shallow feature map. Finally, the output dimension is recovered using 3×3 convolution, and then 5 times upsampling is performed, the final output image is the same size as the original image, the dimension of the output is the type of segmentation, and the position of the highest output value of each pixel point is the predicted category of that pixel. Figure 4 shows the network architecture of DeepLabV3+ model:

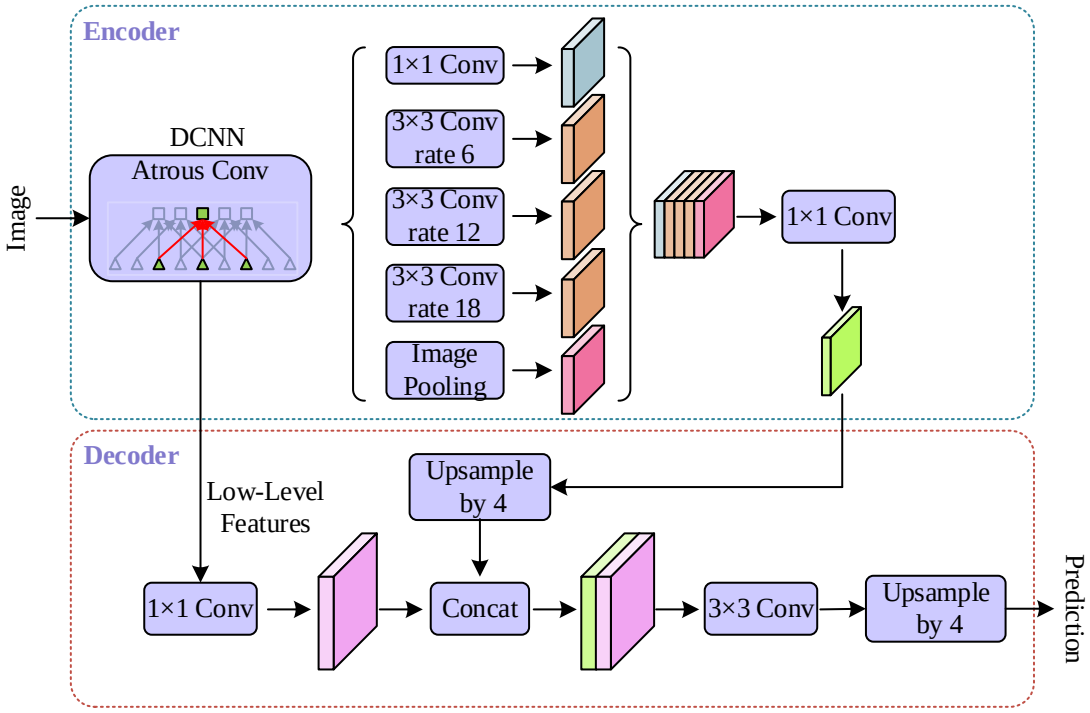


Figure 4: DeepLabV3+ Network Architecture

In this paper, DeepLabV3+ is used as the main model for the analysis of street scene elements, by training the DeepLabV3+ model, and then semantically segmenting the street scene pictures, calibrating different elements with different colors, and calculating the percentage of each element. The rationality of the design is judged according to the planning index of urban construction, so as to give reasonable suggestions.

2.3 Pearson's correlation coefficient

To optimize the perception level of environmental art design space, it is necessary to find the correlation between each street element and the perception dimension to achieve targeted optimization. Pearson's correlation coefficient (PCC), which is used to measure the strength and direction of the linear relationship between two vectors, is an important metric in statistics, and has been widely used in data analysis, regression analysis and other fields. The value of

Pearson's correlation coefficient is between [-1.0, 1.0], which indicates the strength of the linear correlation, and the closer the absolute value is to 1.0, the stronger the linear correlation is, and the closer the absolute value is to 0..0, the weaker the linear correlation is. Positive and negative coefficients indicate the direction of correlation, with positive values indicating positive linear correlation and negative values the opposite. The specific definitions are as follows:

Definition 1: The Pearson correlation coefficient of the vectors $\vec{x} = (x_1, x_2, \ldots, x_n)$ and $\vec{y} = (y_1, y_2, \ldots, y_n)$ is defined as:

$$\begin{aligned} cor(\vec{x}, \vec{y}) &= \frac{\operatorname{cov}(\vec{x}, \vec{y})}{\sigma_{\vec{x}} \sigma_{\vec{y}}} = \frac{E\left[(\vec{x} - E(\vec{x}))(\vec{y} - E(\vec{y}))\right]}{\sigma_{\vec{x}} \sigma_{\vec{y}}} \\ &= \frac{\sum_{k=1}^{n} (x_k - \bar{x})(y_k - \bar{y})}{\sqrt{\sum_{k=1}^{n} (x_k - \bar{x})^2} \sqrt{\sum_{k=1}^{n} (y_k - \bar{y})^2}} \end{aligned} \tag{17}$$

where $\bar{x} = \frac{1}{n} \sum_{k=1}^{n} x_k$, $\bar{y} = \frac{1}{n} \sum_{k=1}^{n} y_k$, and $\operatorname{cov}(\vec{x}, \vec{y})$ denotes the covariance of the vectors $\vec{x}$ and $\vec{y}$. $\sigma_{\vec{x}}$ and $\sigma_{\vec{y}}$ are the standard deviations of the vectors $\vec{x}$ and $\vec{y}$, respectively. The $E(\cdot)$ represents the expected value.

Also, since $\operatorname{cov}(\vec{x}, \vec{y}) = E\left[(x - E(x))(y - E(y))\right] = E(\vec{x}\vec{y}) - E(\vec{x})E(\vec{y})$, Eq. (17) can also be expressed as:

$$\begin{aligned} cor(\vec{x}, \vec{y}) &= \frac{E(\vec{x}\vec{y}) - E(\vec{x})E(\vec{y})}{\sqrt{E(\vec{x}^2) - (E(\vec{x}))^2} \sqrt{E(\vec{y}^2) - (E(\vec{y}))^2}} \\ &= \frac{n \sum_{k=1}^{n} x_k y_k - \left(\sum_{k=1}^{n} x_k\right)\left(\sum_{k=1}^{n} y_k\right)}{\sqrt{n \sum_{k=1}^{n} x_k^2 - \left(\sum_{k=1}^{n} x_k\right)^2} \sqrt{n \sum_{k=1}^{n} y_k^2 - \left(\sum_{k=1}^{n} y_k\right)^2}} \end{aligned} \tag{18}$$

It can be seen that the Pearson's correlation coefficient is a standardized measure, which means that it is not affected by the scale or units. Further give the properties of Pearson's correlation coefficient:

1) For $\forall \vec{x}, \vec{y} \in \mathbb{R}^n$ there is $-1.0 \le cor(\vec{x}, \vec{y}) \le 1.0$.

2) For $\forall \vec{x} \in \mathbb{R}^n$ there is $cor(\vec{x}, \vec{x}) = 1.0$.

3) For $\forall \vec{x}, \vec{y} \in \mathbb{R}^n$ there is $\operatorname{cor}(\vec{x}, \vec{y}) = \operatorname{cor}(\vec{y}, \vec{x})$.

4) There is $\operatorname{cor}(\vec{x}, \vec{y}) = 1.0$ when and only when $\vec{x}$ and $\vec{y}$ are perfectly positively and linearly correlated. That is, $\exists a, b \in \mathbb{R}$ such that $\vec{y} = a \cdot \vec{1} + b \cdot \vec{x}$, where $\vec{1} = (1, \ldots, 1)$ is an n -dimensional vector. Moreover, there is $cor(\vec{x}, \vec{y}) = -1.0$ if and only if $\vec{x}$ and $\vec{y}$ are perfectly negatively linearly correlated.

5) If there exists $a, b, c, d \in \mathbb{R}$ and b and d are nonzero positive real numbers such that $\vec{x}' = a \cdot \vec{1} + b \cdot \vec{x}$ and $\vec{y}' = c \cdot \vec{1} + d \cdot \vec{y}$ hold, then there is $cor(\vec{x}', \vec{y}') = cor(\vec{x}, \vec{y})$.

3 Deep learning-driven spatial perception experience optimization practices

3.1 Evaluation data on the dimensions of visual perception of the regional environment

3.1.1 Statistics on regional environment perception dimension scores

Taking the environmental perception data of each area space within N Street as the research object, the visual spatial perception level of each area in this street is quantitatively evaluated from five dimensions: "beautiful", "lively", "safe", "boring" and "depressing". Figure 5 shows the statistical results of the visual perception dimension scores of the environment in each area of N Street. As can be seen from Figure 5, the evaluation curves of the five dimensions of "beauty", "liveliness", "safety", "boredom" and "depression" all have multiple peaks. Overall, they all take 5 as the dividing line, showing a fluctuating upward trend in the 0-5 range and a fluctuating downward trend in the 5-10 range. The visual landscape environment of each area on N Street gives an overall impression of being uncoordinated. There may be situations such as chaotic layout of spatial elements, which need to be optimized to enhance the visual spatial perception experience.

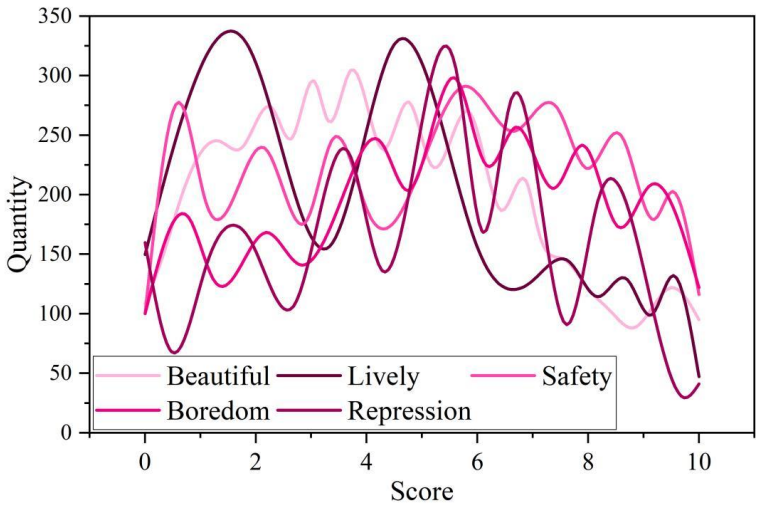


Figure 5: Score of the visual perception dimension of the regional environment

3.1.2 Spatial distribution of regional environmental perception dimensions

Spatial visualization of the environmental visual perception dimension data of the five subjects in each area of N street is shown in Figure 6. The thicker the line, the higher the degree of perception, and the five dimensions of environmental perception have a clear pattern in their spatial distribution. The environmental perceptions of “beautiful” and “lively” dimensions are mainly distributed in A scenic zone, C road, E north road, G building, and I scenic zone. The environmental perceptions of the “safe” and “boring” and “depressing” dimensions are mainly distributed in Middle B, Road D, East F and Building H. The positive dimensions (beautiful and lively) are more spatially dispersed, while the neutral and negative dimensions (safe vs. boring and depressing) are more concentrated. The positive dimensions (beauty and liveliness) are more spatially dispersed, while the neutral and negative dimensions (safety and boredom, depression) are more centrally distributed, and the positive dimensions and the neutral and negative dimensions are generally mutually exclusive in spatial distribution.

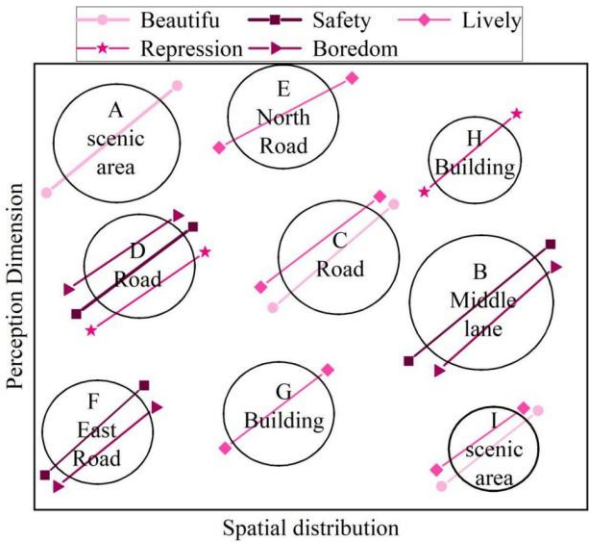


Figure 6: Spatial distribution of the regional environmental perception dimension

3.1.3 Spatial distribution of quantitative indicators of landscape composition

Different landscape composition elements (including green visibility, sky openness, building space share, motorization degree, road space share, and people space share, etc.) in each area within Street N determine the spatial distribution of visual perception dimensions of the regional environment. Figure 7 shows the spatial distribution of quantitative indicators of landscape composition in each area of N Street. High green visibility, high sky openness, and high human space ratio are mainly distributed in the perception space corresponding to "beautiful" and "lively", such as the I Scenic Zone, E North Road, and other areas. High building space ratio, high motorization degree, and high road space ratio are mainly distributed in the perceived space corresponding to "safe", "boring", and "depressing", such as Road D, Building H, and so on. D roads, H buildings, and so on. In general, the visual perception is more positive in areas with a high proportion of natural scenery, and more neutral or negative in areas with a high proportion of humanistic scenery.

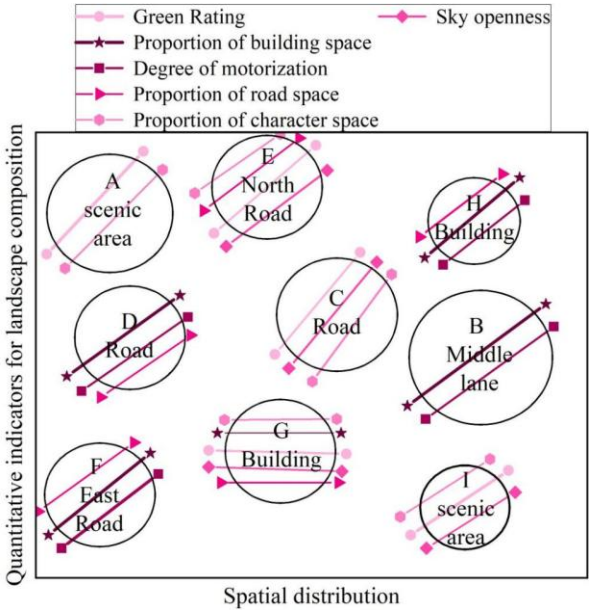


Figure 7: Quantitative indicators for spatial distribution of landscape composition

3.2 Judgment of spatial design quality based on elemental perception

3.2.1 IPA Analysis of Perceived Spatial Imagery Elements of Streets

Construct the street view image dataset of each area of N street. The spatial imagery elements related to the six major quantitative indicators of the landscape composition of the street are segmented and extracted using the supervised learning-based classification model of the street architectural landscape and the analysis model of the street elements. At the same time, the importance-performance analysis method (IPA) is used to jointly analyze the perceived intensity and favorability of each imagery element, which can uncover the perception of different imagery element types and more purposefully guide the optimization of the construction of spatial perception for environmental art design.

Figure 8 shows the results of IPA analysis of the spatial imagery elements related to the 6 major landscape composition quantitative indicators in each area of N. There are 19 spatial imagery elements related to the 6 major landscape composition quantitative indicators in each area of N. The highest degree of perception is the urban green space, which reaches 56,951.41, and belongs to the indicator of the character space ratio; and the highest degree of favorable comments is the trees, which reaches 0.945, and belongs to the indicator of the green visibility rate. It indicates that the positive influence of imagery elements related to natural landscape is greater when perceiving the spatial environment.

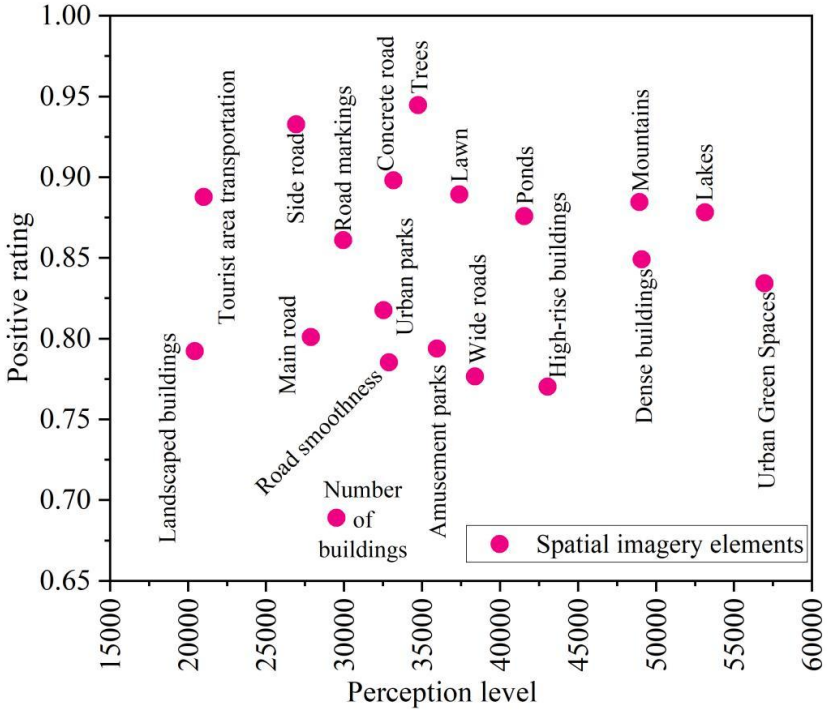


Figure 8: IPA analysis results of spatial imagery elements

3.2.2 Network structure of street space imagery elements

Figure 9 shows the network structure of spatial imagery elements in each area of N Street. From the network structure, it is obvious to see the perceptibility and favorability of spatial imagery elements of each landscape composition quantitative indicator. The perceptions (30,000-50,000) and favorable ratings (0.85-0.95) of the relevant spatial imagery elements of the landscape composition quantitative indicators of green visibility are high. Comparatively, the perceptions (20000-35000) and favorable ratings (0.65-0.80) of the relevant spatial imagery elements of the

landscape composition quantitative indicator of building space occupancy ratio are lower.

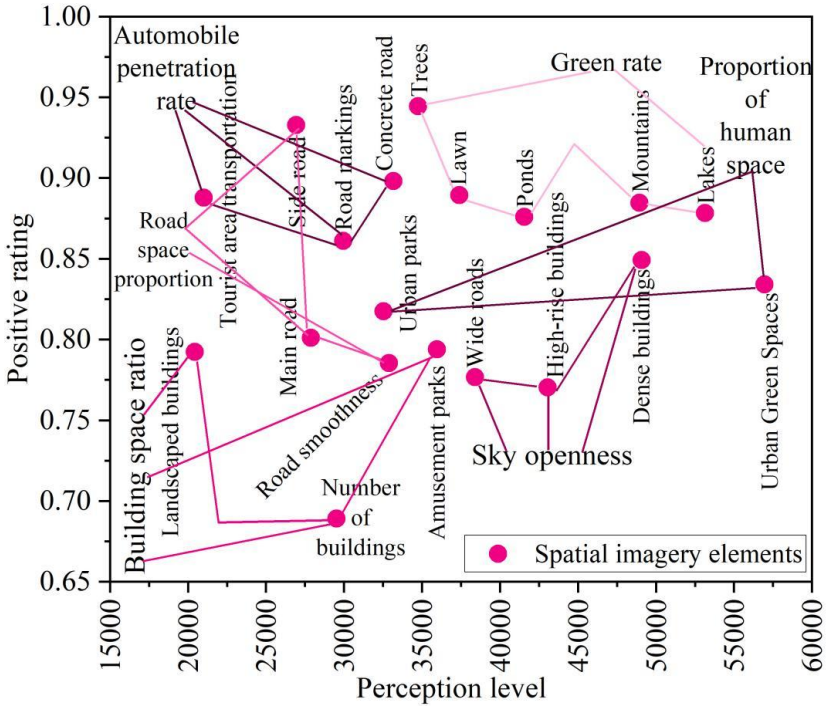


Figure 9: Network structure of spatial imagery elements in each area of Street N

3.3 Pearson correlation coefficient based spatial perception optimization

3.3.1 Correlation analysis of street space imagery elements and perception

By calculating the Pearson correlation coefficients between the 19 street spatial imagery elements and the five perceptual dimensions, the relevance can be judged. Figure 10 shows the results of calculating the Pearson correlation coefficients between the street spatial imagery elements and the perception dimensions. The visual perception of the beautiful dimension requires more mountains (0.85), lakes (0.83), lawns (0.71), urban green spaces (0.86), and urban parks (0.87). The visual perception of the lively dimension requires more ponds (0.81), trees (0.69), lawns (0.51), wide roads (0.84), playgrounds (0.79), urban green spaces (0.82), and urban parks (0.85). The visual perception of safety dimension requires trees (0.52), landscape buildings (0.75), concrete roads (0.73), road markings (0.79), large roads (0.82), and road smoothness (0.72). The number of buildings (0.81), minor roads (0.68), and road leveling (0.64) also lead to a more boring visual perception. High-rise buildings (0.79), dense buildings (0.83), scenic traffic (0.65), and small roads (0.77) can make visual perception more depressing.

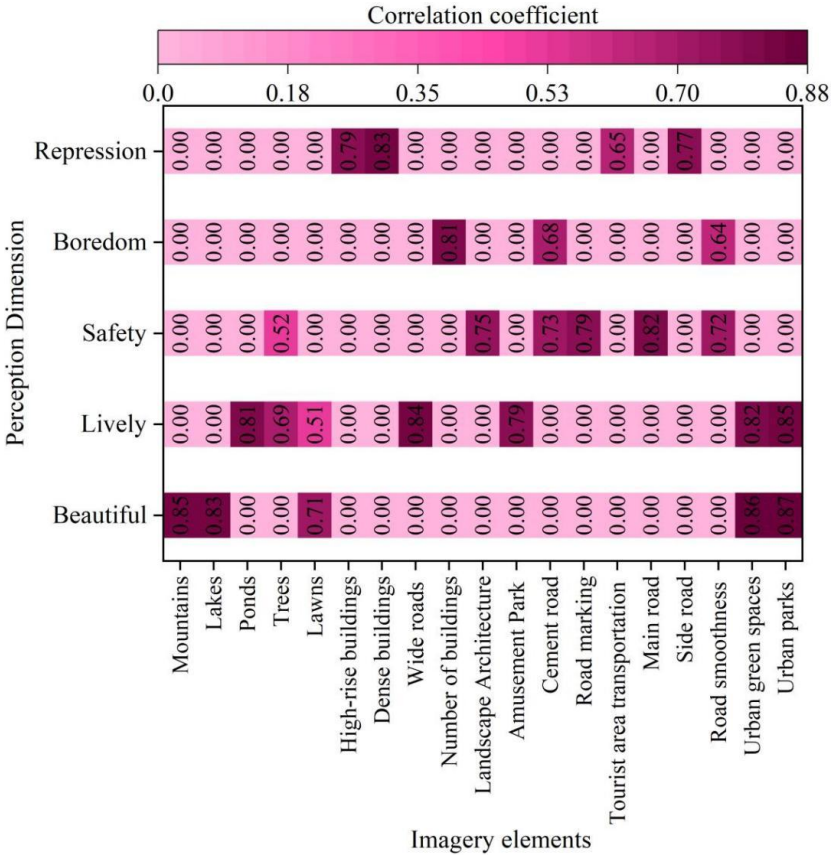


Figure 10: Calculation result of the Pearson correlation coefficient

3.3.2 Differences in street imagery perception under different perceptual conditions

There are also differences in the visual perception of different people for the street spatial imagery elements related to different landscape composition quantitative indicators, for which the differences in the perception of landscape imagery under different perception conditions are calculated by combining individual attributes. Table 1 shows the results of Pearson correlation coefficient calculation of the elemental perceptual differences of different individuals. the perceptual differences of people under 4 different individual attributes for the street spatial imagery elements related to the 6 quantitative indicators of landscape composition are significantly correlated at the 0.05 level, indicating that there are specific types of needs for different people for the design elements and perceptual experience of the street environmental space, which should be taken into account in the subsequent targeted optimization.

Table 1: PCC for differences in individual attribute perception

Landscape imagery	Individual attributes	Beautiful	Lively	Safety	Boredom	Depression
Green rate	Male	0.216*	0.237*	0.245*	0.223*	0.252*
	Female	0.453*	0.476*	0.494*	0.472*	0.479*
	With background in street planning	0.388*	0.411*	0.429*	0.401*	0.418*
	Without background in street planning	0.292*	0.313*	0.331*	0.309*	0.336*
Sky openness	Male	0.193*	0.216*	0.234*	0.215*	0.242*
	Female	0.461*	0.484*	0.502*	0.428*	0.415*
	With background in street planning	0.378*	0.401*	0.419*	0.397*	0.423*
	Without background in street planning	0.259*	0.282*	0.327*	0.305*	0.332*
Building space proportion	Male	0.304*	0.325*	0.343*	0.321*	0.348*
	Female	0.375*	0.398*	0.416*	0.394*	0.421*
	With background in street planning	0.422*	0.445*	0.463*	0.441*	0.468*
	Without background in street planning	0.256*	0.279*	0.297*	0.275*	0.302*
Automobile penetration rate	Male	0.341*	0.364*	0.382*	0.362*	0.389*
	Female	0.305*	0.328*	0.346*	0.324*	0.351*
	With background in street planning	0.467*	0.424*	0.452*	0.423*	0.405*
	Without background in street planning	0.272*	0.295*	0.313*	0.291*	0.318*
Road space proportion	Male	0.316*	0.339*	0.357*	0.335*	0.362*
	Female	0.400*	0.423*	0.421*	0.399*	0.426*
	With background in street planning	0.375*	0.398*	0.416*	0.394*	0.421*
	Without background in street planning	0.216*	0.239*	0.257*	0.235*	0.262*
Proportion of human space	Male	0.413*	0.436*	0.474*	0.458*	0.485*
	Female	0.527*	0.505*	0.523*	0.501*	0.521*
	With background in street planning	0.476*	0.499*	0.517*	0.495*	0.422*
	Without background in street planning	0.308*	0.332*	0.305*	0.284*	0.311*

3.3.3 Comparison of perceived differences in street imagery before and after optimization

Based on the effects of different street spatial imagery elements on the visual perception experience and the results of Pearson's correlation coefficient calculations, the spatial imagery elements of street N were optimized. The Pearson correlation coefficient was further used to calculate the difference in visual perception of streets before and after optimization in Table 2. For ease of reading, the 19 street spatial imagery elements of green visibility-character space occupancy ratio are labeled 1-19 in order. The correlation coefficients between the optimized

street spatial imagery elements and the visual perception dimensions of “beauty” and “liveliness” are all above 0.90 and are significantly correlated at the 0.05 level. The correlation coefficient with the visual perception dimension of “safety” also increased to over 0.80, whereas the correlation coefficient with the visual perception dimension of “boredom” and “depression” decreased to below 0.70. Optimized spatial imagery of the street amplifies the experience of positive and neutral perceptions and reduces the experience of negative perceptions by removing negative imagery.

Table 2: Visual perception differences of the street before and after optimization

Elements of Street Imagery		Before optimization					After optimization				
		Beautiful	Lively	Safety	Boredom	Depression	Beautiful	Lively	Safety	Boredom	Depression
Green viewership rate	1	0.85	0.00	0.00	0.00	0.00	0.94	0.00	0.00	0.00	0.00
	2	0.83	0.00	0.00	0.00	0.00	0.95	0.00	0.00	0.00	0.00
	3	0.00	0.81	0.00	0.00	0.00	0.00	0.91	0.00	0.00	0.00
	4	0.00	0.69	0.52	0.00	0.00	0.00	0.93	0.87	0.00	0.00
	5	0.71	0.51	0.00	0.00	0.00	0.93	0.95	0.00	0.00	0.00
Sky openness	6	0.00	0.00	0.00	0.00	0.79	0.00	0.00	0.00	0.00	0.64
	7	0.00	0.00	0.00	0.00	0.83	0.00	0.00	0.00	0.00	0.59
	8	0.00	0.84	0.00	0.00	0.00	0.00	0.92	0.00	0.00	0.00
Building space proportion	9	0.00	0.00	0.00	0.81	0.00	0.00	0.00	0.00	0.64	0.00
	10	0.00	0.00	0.75	0.00	0.00	0.00	0.00	0.84	0.00	0.00
	11	0.00	0.79	0.00	0.00	0.00	0.00	0.92	0.00	0.00	0.00
Automobile penetration rate	12	0.00	0.00	0.73	0.68	0.00	0.00	0.00	0.83	0.60	0.00
	13	0.00	0.00	0.79	0.00	0.00	0.00	0.00	0.81	0.00	0.00
	14	0.00	0.00	0.00	0.00	0.65	0.00	0.00	0.00	0.00	0.62
Road space proportion	15	0.00	0.00	0.82	0.00	0.00	0.00	0.00	0.80	0.00	0.00
	16	0.00	0.00	0.00	0.00	0.77	0.00	0.00	0.00	0.00	0.57
	17	0.00	0.00	0.72	0.64	0.00	0.00	0.00	0.81	0.62	0.00
Proportion of human space	18	0.86	0.82	0.00	0.00	0.00	0.92	0.96	0.00	0.00	0.00
	19	0.87	0.85	0.00	0.00	0.00	0.91	0.93	0.00	0.00	0.00

4 Conclusion

This study uses deep learning models and Pearson correlation coefficients to investigate the correlation between street space imagery elements and visual perception dimensions, and to accomplish perceptual optimization of street space. The optimized street space enhances the perceptual experience of the three dimensions of “beautiful”, “lively” and “safe”, and weakens the perceptual experience of the two dimensions of “boring” and “depressing”.

By studying the correlation between spatial imagery elements and visual perception dimensions, the overall goal of highlighting the environmental art design effect of street space and enhancing the visual perception experience is achieved. This also lays the foundation for the application of deep learning technology in neighborhood space. Subsequently, the scope of application can be gradually expanded to maintain the wholeness and coherence of spatial optimization.

Funding

This work was supported by “the Fundamental Research Funds for the Central Universities”.

About the Author

Bin Chen was born in Nanjing, Jiangsu, P.R. China, in 1974. He graduated from Nanjing Normal University of Art Design with a master degree and currently works as an associate professor at Jinling Institute of Technology. His research focuses on environmental art and design.

Tao Liang was born in Nanjing, Jiangsu, P.R. China, in 1981. He graduated from Sichuang University with a master degree and currently works as a senior engineer at Jinling Institute of Technology. He also works at the Key Laboratory of New Generation Artificial Intelligence Technology and Its Interdisciplinary Applications (Southeast University), Ministry of Education. His research focuses on electronic information technology.

References

[1] Wang, Y. (2022). The interaction between public environmental art sculpture and environment based on the analysis of spatial environment characteristics. Scientific Programming, 2022(1), 5168975.

[2] Li, P., Liu, K., & Chen, F. (2024). Cognitive characteristics of concept derivation phase in environmental art design: A epistemic network analysis of design scheme construction process. The Design Journal, 27(6), 1208-1228.

[3] Yang, Y., & Guerrini, L. (2020). Environmental (Art) Design VS Interior and Spatial Design: A dialogue between Chinese and Italian design disciplines. In E3S Web of Conferences (Vol. 179, p. 01008). E3S Web of Conferences.

[4] Zhang, L. (2019). The application of traditional folk art in modern environmental art design. Frontiers in Art Research, 1(5).

[5] Zhang, J., Liu, X., Feng, Z., & Feng, X. (2024). Research on the influencing factors of art intervention in the environmental graphics of rural cultural tourism space. Land, 13(10), 1680.

[6] Xu, Y., Guo, Y., Jumani, A. K., & Khatib, S. F. (2021). Application of ecological ideas in indoor environmental art design based on hybrid conformal prediction algorithm framework. Environmental impact assessment review, 86, 106494.

[7] Zhang, J. (2025). Analyzing the Application and Ecological Value of Green Roofs in Urban Environmental Art Design. In MATEC Web of Conferences (Vol. 410, p. 02009). EDP Sciences.

[8] Luo, X., & Gao, X. (2025). Implementation of Building Information Modeling Technology in Interior Space Environmental Art and Personalized Design. Journal of Circuits, Systems, and Computers.

[9] Yu, M., Chen, X., Zheng, X., Cui, W., Ji, Q., & Xing, H. (2025). Evaluation of spatial visual perception of streets based on deep learning and spatial syntax. Scientific Reports, 15(1), 18439.

[10] Kang, Z., Yang, J., Yang, Z., & Cheng, S. (2020). A review of techniques for 3d

reconstruction of indoor environments. ISPRS International Journal of Geo-Information, 9(5), 330.

[11] Wang, T., Li, M., Wang, H., Li, P., Xu, B., & Hu, D. (2025). Context-aware depth estimation for improved 3D reconstruction of homogeneous indoor environments. Automation in Construction, 177, 106343.

[12] Wei, J., Yue, W., Li, M., & 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, 102886.

[13] Liu, F., & Deng, K. (2025). AI knows Aesthetics: AI-Generated Interior design Identification using Deep Learning Algorithms. IEEE Access.

[14] Sun, Y. (2022). Design and optimization of indoor space layout based on deep learning. Mobile Information Systems, 2022(1), 2114884.

[15] Chen, X. (2023). Environmental landscape design and planning system based on computer vision and deep learning. Journal of Intelligent Systems, 32(1), 20220092.

[16] Xu, C. (2025). Deep learning-based landscape design and ecological balance optimization in gardens. Journal of Biotech Research, 20, 59-71.