J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 2737--2756

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Study on the Relevance of Jiangnan Garden Design Elements and Aesthetic Context Based on Visual Recognition

Wenjing Liang[1,✉], Yijing Chen[2], Nadia Binti Mohd Nasir[3]

[1] *School of Art, Shanghai Zhongqiao Vocational and Technical University, Shanghai, 201514, China*
[2] *School of Humanities & Art, Bengbu College of Technology and Business, Bengbu, Anhui, 233000, China*
[3] *Faculty of Creative Industry and Communication, City University Malaysia, Kuala Lumpur, 50000, Malaysia*

ABSTRACT

Jiangnan gardens have become a valuable cultural heritage of China with its elegant garden style. The article proposes a binocular visual recognition system by analyzing the composition of the garden spatial elements and performing feature fusion based on scene-driven coefficients. Ablation experiments are conducted on each part of the constructed data enhancement framework for generating the design of the Jiangnan garden plan, which is applied to generate a set of high-quality datasets and apply the data to image segmentation for generating the design of the Jiangnan garden. The algorithm training is carried out by applying the generated design plan dataset. On this basis, the data from the actual Jiangnan garden research and the spatially quantized feature data are used to do the correlation analysis between the design elements and the aesthetic mood. The data enhancement framework constructed in this paper improves the IOU of ST elements to 0.537, and the average intersection and merger ratio MIOU is 0.389. It shows that the data evaluation framework based on visual recognition is suitable for the study of plan generation of Jiangnan gardens. The correlation coefficients of connection value, spatial control value, average depth value, and integration degree regarding aesthetic context with the data of Jiangnan garden design elements are 0.173, 0.301, -0.278, and 0.325, respectively, which indicate that there is a significant correlation between all of them.

Keywords: jiangnan garden, visual recognition, binocular vision system, scene-driven, multiplexed feature fusion

1. Introduction

Jiangnan garden is the shining point of Chinese culture for thousands of years, and it is famous for its unique cultural connotation and unique style. Through the combination of literati painting style, Jiangnan gardens naturally utilize traditional techniques such as painting art in their design, and this

✉ Corresponding author.
E-mail address: 13361966375@163.com (W. Liang).
Received 10 June 2024; Revised 25 August 2024; Accepted 20 January 2025; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-154

technique of integrating landscape painting creation with garden construction creates unique and literati atmosphere of Jiangnan gardens [1-4]. Jiangnan garden art is the natural environment, architecture, poetry, painting, couplets, sculpture and other arts of integrated art, its far-reaching mood from the comprehensive artistic effect of the garden objects, mainly in the architecture, landscape, flowers and trees, space and time on the construction of these major elements [5-8]. The mood is the focus of Chinese garden design ink, which is also an indispensable element in classical art, without the mood of the garden has no value [9-11].

From the perspective of landscape painting, there may be a certain difference between the landscape depicted by the painter in the painting book and the real landscape, because part of it is the author's own thoughts, but also the product of the beauty of the mood [12-14]. For example, Chinese literati often use pine trees, plum blossoms and so on to express the spirit of robustness, defying winter and the spirit of difficulty, this self-aspiration is the author's emotional expression. Jiangnan's garden design also care about the scene on the blend, pay attention to the characteristics of God, but also pay attention to the expression of the idea of unity of man and heaven, so the work in the spiritual level, more to show the standing of all the elephants in the chest [15-18].

The article cuts into the visual recognition of Jiangnan space garden by analyzing the spatial constituents such as points, lines and surfaces of the garden. A binocular visual recognition system is proposed, and the localization principle of the visual system is analyzed in terms of binocular model, aberration model, and the transformation relationship of the four coordinate systems. Subsequently, the scene-driven parallel feature fusion pyramid structure (SDPMFPN) is investigated, with the help of which the small target feature information is recovered by up-sampling and down-sampling the intermediate feature map and adding parallel branches to fuse shallow features. And by calculating the scene-driven coefficients to adaptively fuse the middle feature layer, we solve the feature extraction conflict problem that occurs when large and small targets exist at the same time.

2. Visual identity of Jiangnan gardens

2.1. Analysis of garden space components

Compared with the architectural space, the garden space is more rich in change, through the points, lines, surfaces and other spatial components of the collocation, as well as stone walls, trees, water, buildings and other elements of the integration of the relationship between the landscaping elements, shaping the garden space is more natural. Garden space and through the combination of points, lines, surfaces and other elements of the layout, to create a memorable spatial mood. Points, lines and surfaces are interdependent and interconnected, so from the points, lines and surfaces and other elements, the spatial composition of Jiangnan gardens to analyze the elements, from the concrete to abstract and then to the concrete research mode to explore the basic combination of garden space elements.

2.1.1. The meaning of points and their spatial significance. In the spatial dimension, dots are often used to depict specific material objects and have fundamental properties associated with them. The composition of dots in both two and three dimensions can be shaped to create gradual, soothing, and undulating artistic effects. The same effect of the same set of points, such as alternation, repetition, and the regularity of the display of the first and then the second, can also present a delicate spatial order, and thus realize the transition from two-dimensional to three-dimensional space.

The composition of non-directional points often presents a highly disorganized free form, in the architectural space, if the same element or the application of the same design technique is too much, it will cause people's visual and psychological dullness, boredom, or a kind of inexplicable constraints on the heart. At this time it is necessary to use a free state of the elements of the composition of the point to disperse this embarrassing and boring picture, in order to alleviate people's visual fatigue and the heart of the feeling of depression. If it is used to decorate the space is too plain in a localized is also very good.

2.1.2. The meaning of lines and their spatial significance. Line is the trajectory of a point and the starting point of a surface. Divided by attributes, lines still have the basic attributes similar to points, but unlike them, lines have the binary enclosure, which can express the beauty and ugliness of things in two-dimensional space by means of two-dimensional composition and other means. In three-dimensional space, lines can not only describe the external contour of matter, but also show the framework structure of atoms and molecules, internal composition and other related properties.

Lines are not only outlines, but also movements, forces and relationships. Lines have the role of expression. Unlike points, lines can be traced with a sense of rhythm by bending in a certain direction and matching reality with reality.

Line is both energetic and full of binding elements in space, through certain proportionality, the spatial put and take, line through certain organizational relationships, rhythm and alternation, stretch and contraction, can achieve the beauty of things written, so that the viewer feels coordinated, comfortable, as if depicting a "silent symphony".

2.1.3. The meaning of surface and its spatial significance. Points and lines are organized, arranged, enclosed and disorganized to form a surface. Then the surface is endowed with morphological properties similar to those of points and lines. Unlike points and lines, in two-dimensional space, surfaces have unique area attributes, and are often used to limit and constrain the use of elemental properties and functions. In the spatial composition, it can be used to inspire modern art creation by showing the contradictory space through certain depiction techniques. According to human emotion, psychology through certain artistic treatment, to create a rich, pleasant use of space. In the three-dimensional space, the surface and the surface through the enclosure, superposition, separation and other means of restriction to shape the virtual and real functional space, when used in architectural design, often produce a sense of domain.

Surface also has positional attributes, in three-dimensional space through the position of the front and back relationship, supplemented by material, color and other material attributes of the elements of the rendering, the formation of a certain spatial hierarchy. In the interspersed and overlapping art techniques, it can shape a kind of spatial atmosphere between the real and the imaginary, when used in garden space, and walls, windows, verandas, water bodies are integrated to form a rich change, light and shadow interlaced with the subtle body suitable for the garden mood.

2.2. Binocular Vision System Analysis

2.2.1. Binocular model. In this paper, a parallel binocular model is used, i.e., two parallel cameras of the same model are used to form a binocular system. The parallel binocular model is shown in Fig. 1.

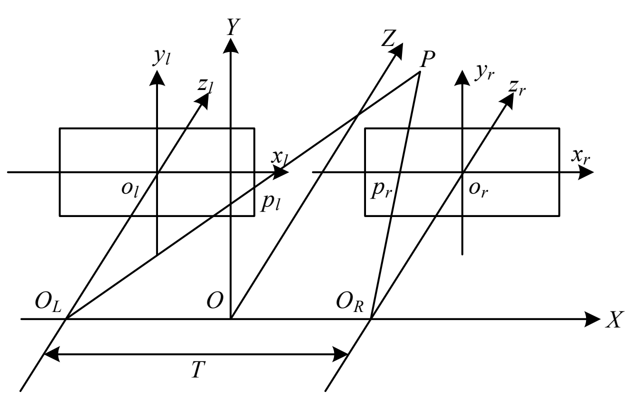

Fig. 1. Parallel binocular model

In the figure, O_L, O_R indicates the center of light of the left and right cameras placed in parallel, defining $f = O_R O_r$ as the focal length of the camera, the baseline length of the binocular camera is T, P_l, P_r are the imaging points of a point in space $P\left(X_w Y_w, Z_w\right)$ on the left and right imaging planes of the binocular camera, and the formula can be introduced due to the similarity of the triangle Px_1x_r and the triangle $PO_L O_R$:

$$\frac{x_r - x_l}{T} = \frac{Z_w - f}{Z_w} \Rightarrow Z_w = \frac{Tf}{T - \left(x_r - x_l\right)} \tag{1}$$

Let $x_l - x_r = d$, then the depth value of point P can be expressed as:

$$Z_w = \frac{Tf}{T - d} \tag{2}$$

where will d be referred to as the parallax value of point P at the left and right camera imaging points.

2.2.2. Aberration models. Camera aberrations can be roughly categorized into radial and tangential aberrations, and the aberration phenomenon is shown in Fig. 2, which needs to be corrected by building an aberration model for camera imaging.

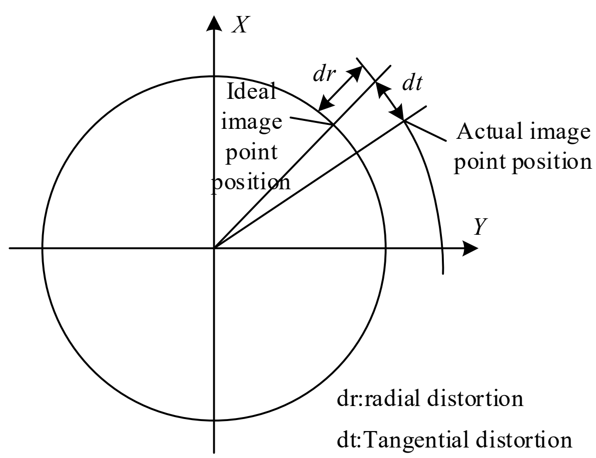

Fig. 2. Schematic diagram of lens distortion

The distortion model can be described as:

$$\begin{cases} x_d = x + \delta_x (x + y) \\ y_d = y + \delta_y (x + y) \end{cases} \tag{3}$$

where (x, y) is the coordinate of the imaging point in the ideal state, $\left(x_d, y_d\right)$ is the actual coordinate of the point after the occurrence of aberration, and $\delta_x(x, y)$ and $\delta_y(x, y)$ are the aberration parameters of the model.

In order to correct the aberrations in the imaging process and reduce the errors caused by the imaging process, it is necessary to establish the aberration correction relation equation to describe the aberration correction model. Let the distance between the imaging point and the imaging center be r, the coordinate changes before and after the radial aberration can be described by the Taylor expansion at $r = 0$, and the expression is shown in Equation (4):

$$\begin{cases} x_{corr} = x_{dis}\left(1+k_1r^2+k_2r^4+k_3r^6\right) \\ y_{corr} = y_{dis}\left(1+k_1r^2+k_2r^4+k_3r^6\right) \end{cases} \tag{4}$$

where $\left(x_{corr}, y_{corr}\right)$ represents the coordinates after the distortion correction, $\left(x_{dis}, y_{dis}\right)$ represents the coordinates of the imaging point in the case of distortion, and k_1, k_2, k_3 represents the coefficient of radial distortion.

The main reason for tangential aberrations is that the lens in the manufacturing and installation process errors, so that the lens and the camera imaging plane can not be kept completely parallel to each other, resulting in a change in the imaging position. The expression for correction of tangential aberration is shown in equation (5):

$$\begin{cases} x_{corr} = x_{dis} + \left(2p_1xy + p_2\left(r_2 + 2x^2\right)\right) \\ y_{corr} = y_{dis} + \left(p_1\left(r^2 + 2y^2\right) + 2p_2xy\right) \end{cases} \tag{5}$$

where $\left(x_{carr}, y_{carr}\right)$ denotes the coordinates after aberration correction, $\left(x_{dis}, y_{dis}\right)$ denotes the coordinates of the imaging point in the case of aberration, and p_1, p_2 denote the coefficients of tangential aberration. Finally, the distortion parameter of the camera is defined as $\left[k_1, k_2, k_3, p_1, p_2\right]$.

2.2.3. Conversion of the four basic coordinate systems. In the imaging process, the binocular camera involves four coordinate systems: the world coordinate system, the camera coordinate system, the physical coordinate system and the pixel plane coordinate system. Through the mutual conversion of these coordinate systems, the imaging plane can be completed on a point of two-dimensional pixel coordinates to the world coordinate system of three-dimensional mutual conversion, the conversion process shown in Figure 3:

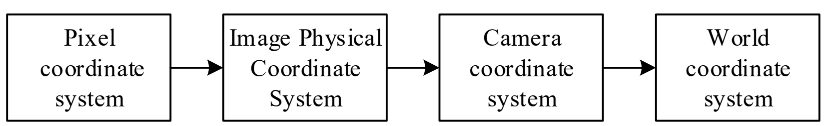

Fig. 3. Coordinate system transformation process

The four coordinate systems and their transformation relationships are described below:

1) Pixel coordinate system to image physical coordinate system. The relationship between the pixel plane coordinate system and the physical coordinate system of the image is shown in Fig. 4. Where o_{uv} represents the pixel plane coordinate system, the coordinate origin is the vertex of the upper left corner of the image, which is defined as $\left(u_0, v_0\right)$, and the unit of the coordinate is pixel.

o_{xy} represents the image physical coordinate system, the coordinate origin is the intersection of the center axis of the optical center of the camera and the imaging plane, which can be known from the imaging principle as the center of the pixel plane, defining the point as $\left(x_0, y_0\right)$.

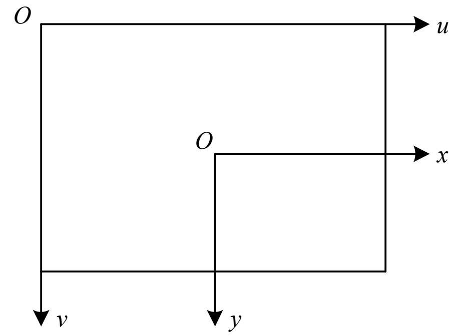

Fig. 4. The pixel plane coordinate system and the image physical coordinate system

The relationship between pixel coordinates and physical coordinates can be established by defining the actual physical lengths of a single pixel point corresponding to directions x and y in the physical coordinate system as d_x and d_y, respectively:

$$\begin{cases} u = \dfrac{x}{dx} + u_0 \\ v = \dfrac{y}{dy} + v_0 \end{cases} \tag{6}$$

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \dfrac{1}{d_x} & 0 & u_0 \\ 0 & \dfrac{1}{d_y} & v_o \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{7}$$

2) Image physical coordinate system to the camera coordinate system. Image physical coordinate system to the camera coordinate system transformation, actually from the two-dimensional plane to the three-dimensional space of the projection transformation, the relationship between the changes shown in Figure 5. The Z_c axis of the camera coordinate system passes through the origin of the image physical coordinate system, and the $XcOcYc$ plane is parallel to the xoy plane.

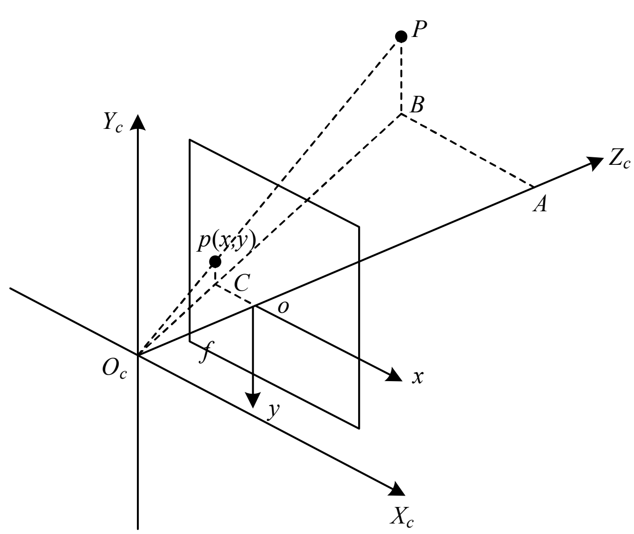

Fig. 5. The image physical coordinate system and the camera coordinate system

A point P in space is projected to the point $p(x, y)$ in the image coordinate system $o-xy$. Since the image physical coordinate system is parallel to the camera coordinate system in the ideal imaging model, triangle ABO_c is similar to triangle OCOc, and triangle PBO_c is similar to triangle pCO_c. The conversion process of the two coordinate systems can be expressed by the similarity relation as:

$$\frac{AB}{oC}=\frac{AO_c}{oO_c}=\frac{PB}{pC}=\frac{X_c}{x}=\frac{Z_c}{f}=\frac{Y_c}{y} \tag{8}$$

$$x=f\frac{X_w}{Z_w}y=f\frac{Y_w}{Z_w} \tag{9}$$

$$Z_c\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}=\begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{10}$$

3) Camera Coordinate System to World Coordinate System. The world coordinate system is used to describe the actual position of the target object relative to the camera, and the conversion relationship between the two is shown in Figure 6. The point $\left(X_c, Y_c, Z_c\right)$ under the camera coordinate system can be converted to $\left(X_w, Y_w, Z_w\right)$ under the world coordinate system by this conversion relationship.

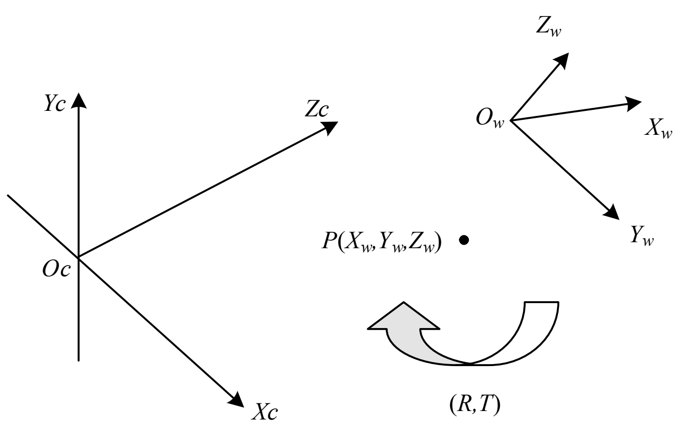

Fig. 6. Camera coordinate system and world coordinate system

The transformation from the camera coordinate system to the world coordinate system can be regarded as a rigid body transformation, and the connection between the two coordinate systems can be established by translation and rotation operations. The mathematical relationship of the transformation process can be expressed as:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}=\begin{bmatrix} R & T \\ Q^T & 1 \end{bmatrix}\begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \tag{11}$$

where R is the rotation matrix of 3×3, T is the translation vector of 3×1, and $Q^T=\left(0,0,0\right)^T$.

Combining the transformation process from (1) to (3), the schematic diagram of the transformation process of the four coordinate systems can be simplified as shown in Fig. 7.

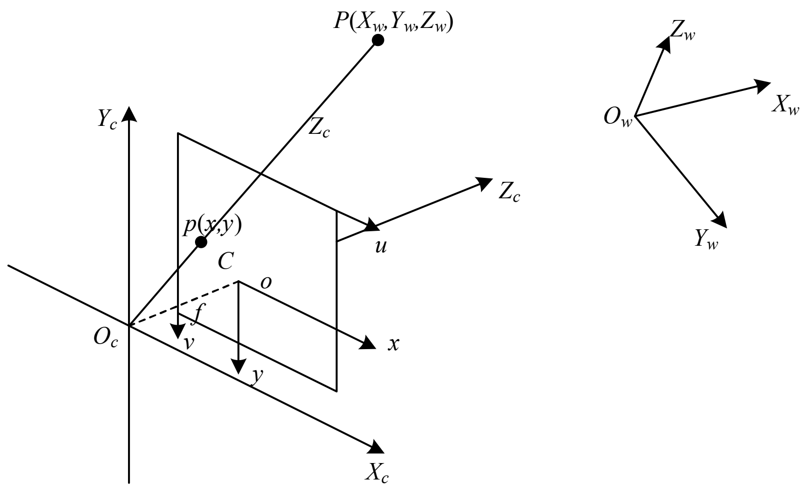

Fig. 7. The transformation of four coordinate systems

Combining Eqs. (7), (10), and (11) one can introduce the mathematical model corresponding to the world coordinate system on the pixel coordinate system as:

$$Z_c\begin{bmatrix}u\\v\\1\end{bmatrix}=\begin{bmatrix}\frac{1}{d_x}&0&u_o\\0&\frac{1}{d_y}&v_O\\0&0&1\end{bmatrix}\begin{bmatrix}f&0&0&0\\0&f&0&0\\0&0&1&0\end{bmatrix}\begin{bmatrix}R&T\\O^T&1\end{bmatrix}\begin{bmatrix}Y_w\\Y_w\\Z_w\\1\end{bmatrix}$$
$$=\begin{bmatrix}\alpha_x&0&u_0&0\\0&\alpha_y&v_0&0\\0&0&1&0\end{bmatrix}\begin{bmatrix}R&T\\O^T&1\end{bmatrix}\begin{bmatrix}Y_w\\Y_w\\Z_w\\1\end{bmatrix}=M_1M_2\begin{bmatrix}Y_w\\Y_w\\Z_w\\1\end{bmatrix} \tag{12}$$

In Equation (12), $M_1=\begin{bmatrix}\alpha_x&0&u_0&0\\0&\alpha_y&v_0&0\\0&0&1&0\end{bmatrix}$ is the internal parameter of the camera, $M_2=\begin{bmatrix}R&T\\Q^T&1\end{bmatrix}$ is the external parameter of the camera, and $\alpha_x=\frac{f}{d_x}$ and $\alpha_y=\frac{f}{d_y}$ represent the magnification factor when the pixel coordinates of the imaging point are converted to physical coordinates after normalizing the focal length of the camera, respectively.

2.3. Scene-driven parallel multiplexed feature fusion pyramid

In target recognition, in order to obtain rich semantic information with a larger sensory field, many classical models generally downsample with a large multiplier, but the use of a larger downsampling multiplier often results in the loss of small target feature information. In the pictures collected by UAV, the size of the target is not fixed, when the larger size target and the smaller size target appear at the same time or the larger size target and the smaller size target appear to be sticky, often due to the small target feature information has disappeared in the deep feature map dispersion, and finally the model focuses on the recognition of the large target and ignores the small target, resulting in inaccurate recognition. Therefore, this section proposes a scene-driven parallel multiplexed feature fusion pyramid structure.

2.3.1. Definition of scene-driven coefficients. When recognizing the information collected by the UAV, it is not possible to determine which frame has only a large target, which frame has a small target, and which frame has both a large target and a small target, so in this section, we make a judgment by defining the scene-driven coefficients, and perform the feature fusion based on the scene-driven coefficients.

The background color of the Jiangnan garden scene explored in this paper is relatively single, so the background interference can be removed by the color space, and then the size gap of the foreground target in the current recognition picture can be determined according to the area size of the foreground target. The specific calculation method of scene driving factor g is as follows:

1) Convert the image from RGB color space to HSV color space, calculate the average value of all pixel points in the image with H components as $\bar{H}$, and filter out the pixel points with H components in the range of $\left(\bar{H}-5,\bar{H}+5\right)$ through the mask.

2) Morphological operations (binarization, open operation, closed operation, etc.) are performed first, and then the target contour is obtained after edge extraction.

3) Calculate the number of pixel points within all contours in the image, and let the contour with the highest number of pixel points contain $P_{\max}$ pixel points, and the contour with the lowest number of pixel points contain $P_{\min}$ pixel points, and the scene driving coefficients are as shown in Equation (13).

$$g = sig\,\mathrm{mod}\left(\frac{P_{\max}}{P_{\min}}\right) \tag{13}$$

The larger $g \in (0,1)$ and g indicate that the larger the difference between $P_{\max}$ and $P_{\min}$, the larger the difference in foreground target size, the larger the probability of losing small target features in the deep feature map; the smaller g indicates that the smaller the difference between $P_{\max}$ and $P_{\min}$, the smaller the difference in foreground target size, the smaller the probability of losing small target features in the deep feature map.

2.3.2. Multiplexed feature fusion methods. For most models, the larger the downsampling multiple, the richer the semantic information and the larger the receptive field, and in image classification, the downsampling multiple is too large and the classification effect is often very good. For example, in the ResNets network, the downsampling multiple is 32 times, and the network structure is also very deep, so the network can obtain richer semantic information and a larger sensory field when detecting large targets, which makes the detection accuracy higher. However, this network structure and downsampling times are very unfriendly to the recognition of small targets, and the features of small targets are easily disappeared after large times of downsampling. In recent years, many solutions have been proposed for the problem of low accuracy of small target recognition, such as the feature pyramid (FPN), which fuses shallow and deep features by constructing a pyramid structure, thus solving the problem of small target features disappearing easily after large times of downsampling. However, for small targets, their feature information has already disappeared in the deep feature map, and it is difficult to recover the feature information after the up-sampling operation, so FPN still has the problem of small target feature loss. The targets in this scenario are all small targets, so to address the above problem, this section proposes the scene-driven parallel multiplexed feature fusion pyramid (SDPMFPN).

Since the intermediate feature maps are rich in semantic and detail information, and the shallow feature maps correspond to a low downsampling multiplier, which makes the small target feature information well preserved. Therefore, the SDPMFPN structure proposed in this section recovers the small target feature information by (1) up-sampling and down-sampling the intermediate feature graph. (2) Adding parallel branches to fuse shallow features.

The structure of SDPMFPN in Backbone uses ResNets structure, and the main path follows the original FPN feature extraction structure, and the bottom-up input is the output part of the backbone network. The selected feature map is defined as $\{F_2, F_3, F_4\}$ corresponding to downsampling multiplicity of $\{8,16,32\}$, respectively, the top-down output feature map is $\{P_2, P_3, P_4\}$, respectively, and the parallel branch corresponds to feature map $\{R_2, R_3\}$. The specific fusion process is shown in Eqs. (14), (15), and (16):

$$P_4^* = Conv_{3\times3}\left\{P_4 \oplus g \cdot Downsample\left(P_3\right)\right\} \tag{14}$$

$$P_3^* = \mathrm{Conv}_{3\times3}\left\{P_3 \oplus R_3\right\} \tag{15}$$

$$P_2^* = Conv_{3\times3}\left\{P_2 \oplus g \cdot Upsample\left(P_3\right) \oplus R_2\right\} \tag{16}$$

where, $Conv_{k\times k}$ represents the convolution operation, k represents the convolution kernel size, g represents the scene driving coefficients, $\oplus$ represents the feature fusion method, P_i^* represents the feature map after the fused features, R_i represents the feature map with parallel branches, $Upsample$ represents the upsampling operation, and $Downsample$ represents the downsampling operation. The corresponding feature map of the pyramid is defined as shown in Eqs. (17), (18):

$$P_4 = Conv_{1\times1}\left\{F_4\right\} \tag{17}$$

$$P_i = Conv_{3\times3}\left\{Conv_{1\times1}\left\{F_i\right\} + Upsample\left(P_{i+1}\right)\right\} \tag{18}$$

where $Conv_{1\times1}$ denotes the reduction of the number of feature map channels to 256 using 1×1-convolution. The feature maps of the pyramid layer of parallel branches are defined as shown in Eqs. (19), (20):

$$R_3 = Conv_{1\times1}\left\{F_3\right\} \tag{19}$$

$$R_i = Conv_{3\times3}\left\{Conv_{1\times1}\left\{F_i\right\} + Upsample\left(R_{i+1}\right)\right\} \tag{20}$$

This structure solves the feature conflict problem of the original feature pyramid when facing a large difference in target size through the effective fusion between different layers, realizes a better feature extraction effect, and strengthens the extraction ability for small targets.

3. Research on the design dataset based on visual recognition for plan generation of Jiangnan gardens

3.1. Rationality validation of the data enhancement framework for the generation of Jiangnan garden floor plans

In order to prove the importance of each part in the Jiangnan garden plan data enhancement framework based on binocular vision system constructed by the research, ablation experiments are conducted for the 2 key steps of outer environment cutting and element hierarchical recognition, and each part is proved to be indispensable by semantic segmentation evaluation metrics.

3.1.1. Data preparation for ablation experiments. For the ablation experiments, 2 sets of new training data sets with 2 sets of standard test sets were constructed on the basis of the data enhancement framework. Two sets of 200 Jiangnan garden plans were selected as the training data, respectively Train_Non.Mask.SEG_200 and Train_Mask.SEG_200. Meanwhile, 50 Jiangnan garden plans were selected as the test set, and the test sets were untrained. 2 sets of standard test sets were Test_Non.Mask.SEG_50 and Test_Mask.SEG_50, respectively. SEG_50 with Test_Mask.

3.1.2. Semantic segmentation effect evaluation metrics. Traditional classification metrics such as accuracy, precision, recall, F-socre, etc. are only concerned with the accuracy of pixel classification. In the field of image semantic segmentation, spatial matching is a more important criterion. Therefore, the study introduces intersection and merger ratio (IOU) and mean intersection and merger ratio (MIOU) as the evaluation metrics of semantic segmentation accuracy. Compared to other metrics, MIOU not only evaluates both individual sample and overall performance in multi-sample tasks, but also applies to the case of sample imbalance. For example, in this study, structures and roads have a smaller sample size than other elements, but MIOU can also be accurately evaluated.IOU refers to the ratio of the intersection and concatenation of the true and predicted values of two categories, and MIOU refers to the average IOU of multiple categories.The formula is as follows:

$$IOU = \frac{TP}{FN + TP + FP} \tag{21}$$

$$MIOU = \frac{1}{k+1} \sum_{i=1}^{k} \frac{TP}{FN + TP + FP} \tag{22}$$

where: k is the number of categories; TP is true positive; FP is false positive; FN is false negative; TN is true negative.

3.1.3. Ablation experiments to reduce the effect of uncorrelated features. To address the problem that the data enhancement process is highly susceptible to the influence of irrelevant features, the study considers the external environment of the design site in the plan view of the Jiangnan Garden as a complex interference factor, so the external environment is used as a variable in the ablation experiment. The same Pix2Pix training strategy is adopted, firstly, 2 datasets of Train_Non.Mask.SEG_200 and Train_Mask.SEG_200 are trained respectively to obtain the unmasked planar map segmentation model and the masked planar map segmentation model. Secondly, the corresponding segmentation models are tested by inputting the Test_Non.Mask.SEG_50 and Test_Mask.SEG_50 2 standard test sets. Finally, the test results are evaluated for each element IOU and overall MIOU metrics.

The metrics evaluation results are shown in Figure 8. The MIOU of the unmasked planar map segmentation model on the training set is 0.293, and the MIOU of the masked planar map segmentation model is 0.385. After masking, the IOU of all the other 5 types of elements is improved except for the road, which still has an IOU of 0.231. The recognition accuracy of the road element is not significantly improved due to the masking problem of the floor plan itself and the limitation of the segmentation model. The structure element has various forms in the plan view, with few common features, and the area of the structure in the plan view accounts for a very small proportion, which makes it difficult to recognize. After adding the outer environment mask to reduce irrelevant features, the recognition accuracy of structures is significantly improved.

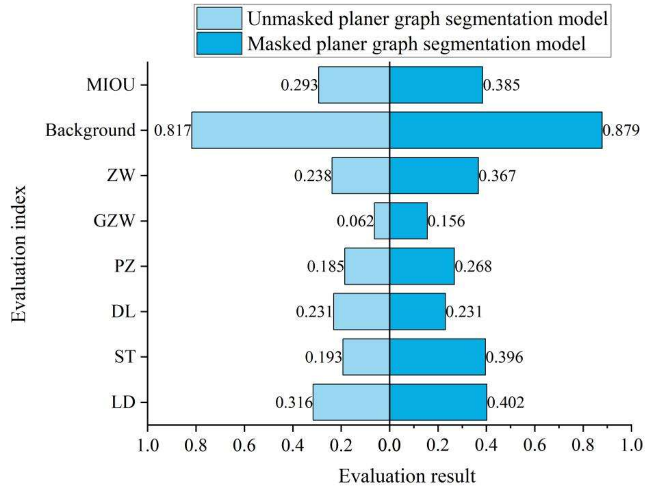

Fig. 8. Results of ablation experiments to reduce the influence of irrelevant features

3.1.4. Ablation experiments for semantic segmentation of floor plans. The algorithms trained by the GAN framework generate synthetic data that are usually difficult to achieve the reasonableness and accuracy needed by the planar generation algorithms. Therefore, instead of adopting the strategy of direct planar map segmentation, the study constructed the optimization method described above. In order to prove that the constructed optimization method is more suitable for the generation study than the direct segmentation, ablation experiments are conducted for the 2 types of methods.

For the training and testing results of the direct segmentation model, the study has been done in the ablation experiment of reducing the influence of irrelevant features, i.e., masking the planar graph segmentation model. Therefore, Test_Mask.SEG_50 is input into the hierarchical recognition model in the data enhancement framework based on binocular vision system analysis and SDPMFPN constructed in this paper, and the metrics evaluation is compared with the test results of the masked plan view segmentation model.

The results of the metrics evaluation are shown in Fig. 9. The results show that the MIOU of the data enhancement framework constructed in this paper is not substantially improved over the masked planar graph segmentation model, with an average intersection and merger ratio of 0.376 for the data enhancement framework without Mask, and an average intersection and merger ratio of 0.389 for the data enhancement framework with Mask, which is slightly higher than the MIOU of 0.386 for the masked planar graph segmentation model. For the water elements, which are more difficult to recognize, the IOU improves to 0.537. While the IOU for certain elements such as structures, paving, and green spaces decreases. There are three reasons for this:

(1) The mask planar segmentation model is not a classical semantic segmentation model, but a generative model. Due to the classical strategy of GAN, the generator will generate many plausible results, so the color of the segmented planar map is not in strict accordance with the standard color card, but there will be a lot of “critical colors”, which leads to the need to calculate the IOU by defining the color range to identify the color. When calculating the IOU, it is necessary to define the color range to identify, so more pixels are identified, resulting in an inflated IOU value;

(2) The color of the segmentation map obtained by the data enhancement framework is standard, but the edge optimization experiment will expand the pixels recognized by each element, and the FP will increase, resulting in a decrease in the IOU;

(3) Compared to extracting the elements separately, merged extraction (masked planar map segmentation model) enables the neural network to learn the underlying logic and connections between the elements, which helps to improve the accuracy of segmentation. Since the green space is at the bottom layer and is heavily occluded, the separately extracted green space image has more vacancies inside due to occlusion. Therefore, the study chose to add the site edge Mask at the bottom

layer to get the complete semantic segmentation plan view data. Instead of reducing the accuracy, this step makes the image more complete and is suitable for planar generation studies.

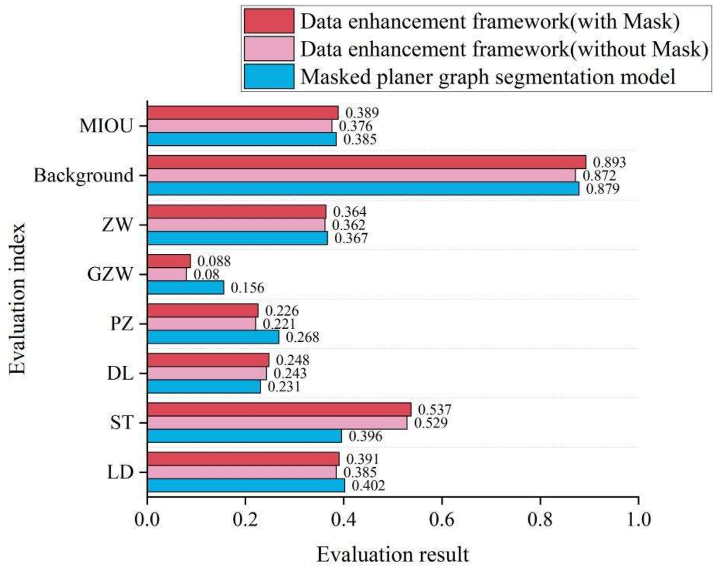

Fig. 9. Results of ablation experiments on semantic segmentation

Compared with the merged segmentation strategy of the masked planar graph segmentation model, the dataset obtained from the optimization framework of this study is more suitable for the training of generation design related algorithms, such as initially solving the problems of intermittent segmentation, "critical color" and noise caused by the non-semantic segmentation model, and in particular, the image after adding Mask is more suitable to be used as a dataset for the planar generation research.

3.2. The application of data enhancement framework for the generation of Jiangnan garden floor plan

Tens of thousands of floor plans of Jiangnan landscape gardens were collected by crawling on famous landscape garden design websites, such as Pinterest, gooood, mooool, and Arch Daily. Combined with multiple rounds of K-means clustering and manual categorization, a total of 8973 floor plans meeting the research criteria were screened out. These floor plans are inputted into the data enhancement framework constructed in this study, and the accurate semantic segmentation images of 8973 floor plans are finally obtained.

The acquired 8973 complete sets of segmented floor plans are recombined with the original floor plans as Train_SEG_8973, and trained to obtain the planar fast segmentation model. The training effect is tested using Test_Mask.SEG_50. The experimental evaluation results of the planar recognition application are shown in Fig. 10.

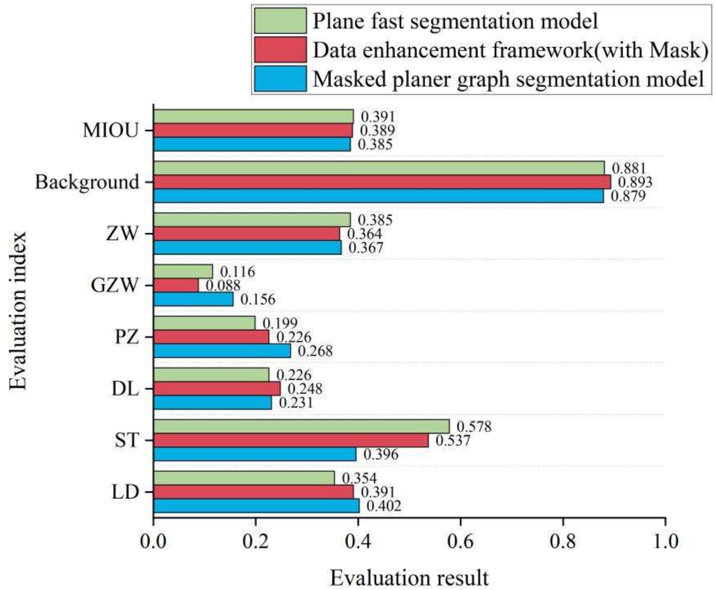

Fig. 10. Experimental evaluation results of planar recognition application

The planar fast segmentation model did not improve the MIOU significantly as did the data augmentation framework optimized in this study.The planar fast segmentation model with an MIOU of 0.391 only improved the MIOU by 0.005 over the masked planar map segmentation model on the initial training set after increasing the dataset by tens of times.The results of the study show that for semantic segmentation tasks, the training dataset needs to have a high degree of accuracy in order to get the desired generated results, therefore, even though the study expanded the amount of data, it was difficult to improve the model performance because the accuracy was not up to the level of manual Photoshop drawing. Although the segmented images obtained by the model after expanding the samples are more complete and standardized, which is more suitable for planar generation research, more accurate segmentation methods need to be further explored.

4. Analysis of the association between southern garden design elements and aesthetic meaning

4.1. The Application of Visual Recognition in the Design of Jiangnan Gardens

In this study, we first use the design proposal dataset to train the algorithm based on binocular vision recognition and SDPMFPN, so that the algorithm generates diversified design proposals; and then we visualize the internal hidden space of the algorithm using Principal Component Analysis (PCA) dimensionality reduction method, and analyze the features by which the algorithm guides the generation of diversified design proposals, in order to explore how the neural network understands the design features.

The algorithm is trained with a large amount of data to acquire different design features and generate different solutions. Among them, the w vector contains the features of each generated scheme. Therefore, the key part of this study is to explore w-vectors, including “w-vector data characterization” and “w-vector semantic information analysis”.

4.1.1. Algorithm training. Jiangnan garden design scheme is a static plan image data, this study adopts the visual recognition technology proposed in the article, and the training adopts 512×512 resolution, running on dual-card NVIDIA 3090 GPUs with 64 GB of video memory.Among them, in order to better acquire the design law, the data mirror enhancement function is turned on for the training of the generalized dataset; and for the training of the directional dataset, the data mirror enhancement function is not turned on for keeping the The red line of the field remains unchanged, the data mirroring enhancement function is not enabled.

4.1.2. Results of w-vector semantic information analysis. In this section, w-vectors are truncated using the truncation technique: 10 consecutive floating point numbers with truncation factors set from 0 to 0.9 (in steps of 0.1) are used to generate 1,000 w-vectors based on the 100 random z-vectors from the previous section, which are co-constrained to the algorithm. Further embedded image analysis allows visualization of the average scheme generated by the generator. The oriented generator can summarize the common design framework from multiple scenarios, which can represent the most suitable design solution for the site, and other diversity scenarios are based on this framework for fine-tuning the design content. Figure 11 shows the average w-vector location.

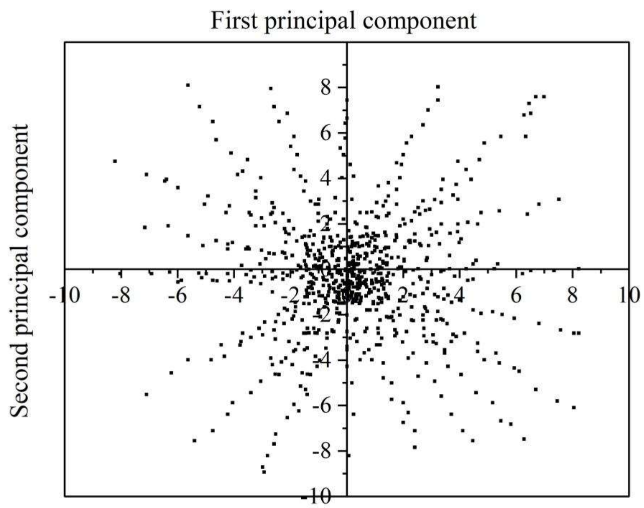

Fig. 11. Average w vector position

As can be seen from the distribution of the mean vector w positions in Fig. 11, the mean vectors are mainly concentrated between the principal components of -2 and 2. The larger the component, the more sparsely the vectors are distributed.

The complete w vectors (for 1) of the directed generators are red-labeled and clustered to obtain 100 diversified design solutions with complete design features. The distribution of diversified solutions is shown in Fig. 12, and the clustering of diversified design solutions is shown in Fig. 13. Regulated from 0 to 0.9, the changes of w vector under the influence of different truncation coefficients are analyzed. Affected by the truncation coefficients, the scheme features change in a specific direction, and the semantic features contained in this w vector can be analyzed based on the direction of the change of the scheme features.

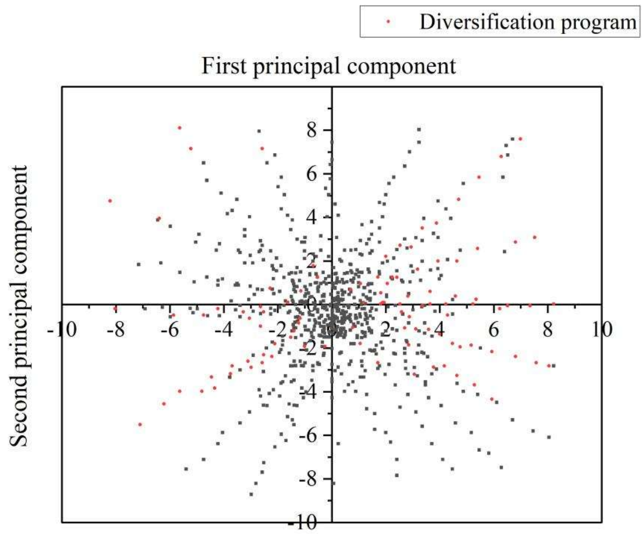

Fig. 12. Diversification scheme distribution

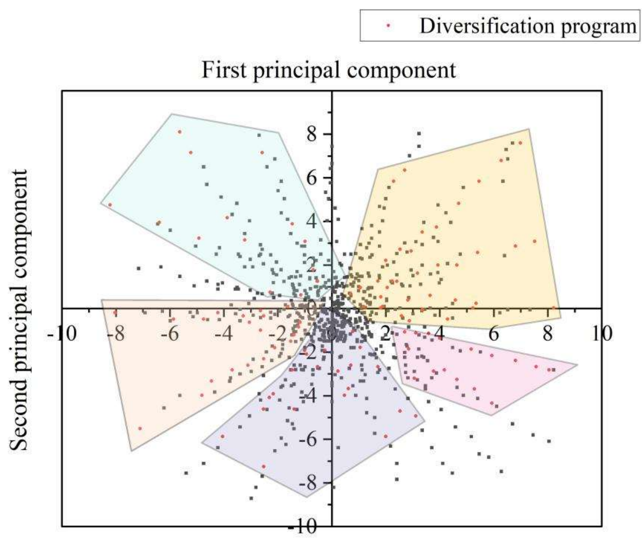

Fig. 13. Diverse design scheme clustering

The generated diversified design solutions with complete design features are mainly distributed in the range of larger values of the principal components, and the diversified design solutions generated between [-2,2], where the mean vector w is denser, are inversely not many, and the clustering divides the diversified design solutions generated by the directional generator into five classes.

4.2. Aesthetic Contextual Correlation Analysis

4.2.1. Data analysis methods. Applying statistical analysis to analyze data helps to better understand the state of objective phenomena, reveal the nature and regularity of objective phenomena, and predict the activities of objective phenomena in the future. The application of SPSS 26.0 analysis software in the way of one-way ANOVA to analyze the statistical data, can be in the difference test to determine whether the control variables between more than two groups of data have a significant impact on the observed variables, analysis of the research will be the aesthetic mood of the spatial connectivity value, the control value, the average depth of the value of the degree of integration, respectively, are set as a control variable, the observation variable is the design of the Jiangnan gardens elements, and the control variables were analyzed by one-way ANOVA with the observation variables two by two, respectively. The evaluation index for testing the significance of difference is r. When the significance value r < 0.05 indicates that the difference is significant, and the significance r < 00.1 indicates that the significant characteristics of the difference are more obvious. The following analysis will first test the normal distribution of the data, according to the data normal distribution or not, and then further categorize the data for correlation test.

4.2.2. Test for normal distribution of data. Doing field research and analysis on a landscape garden in Jiangnan area, SPSS26.0 was applied to analyze the data of the actual research and the feature data quantified in the space above, to study the correlation as well as the correlation coefficient. First of all, the data were tested for normality, and Figures 14-Figure 17 show the normality detection of connection value, control value, average depth value and integration in the test variables of the samples, respectively.

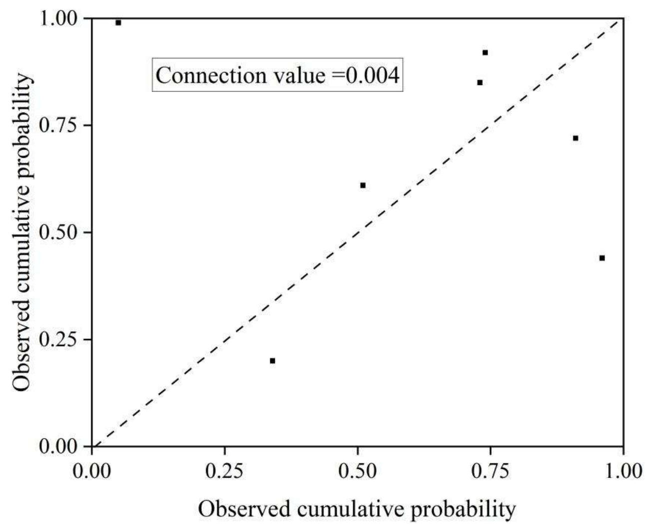

Fig. 14. Connection value positive detection P-P plot

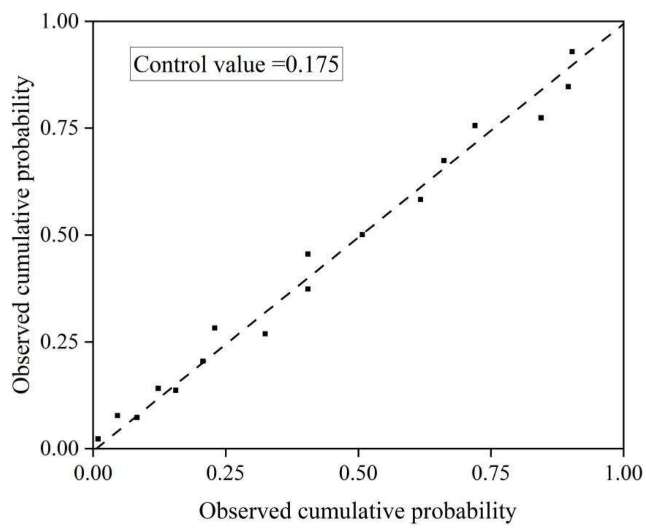

Fig. 15. Control value positive detection P-P plot

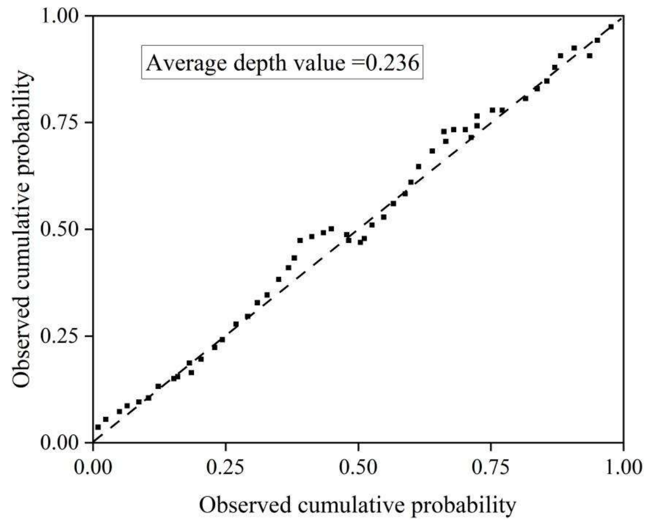

Fig. 16. Average depth value positive detection P-P plot

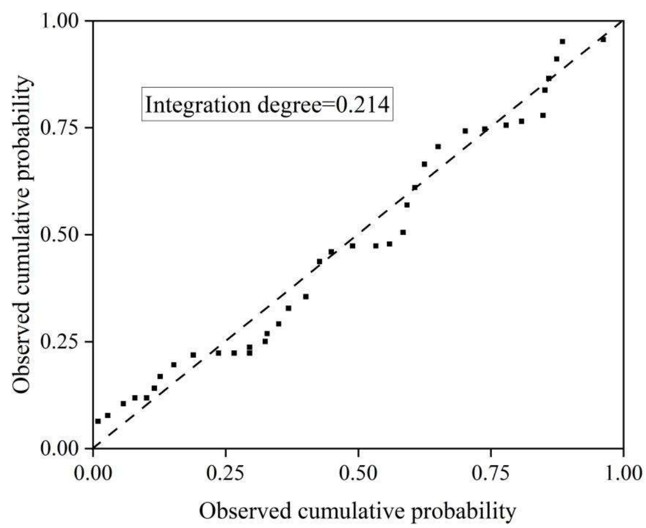

Fig. 17. Integration degree positive detection P-P plot

The quantification of spatial values based on visual recognition will be tested for normal distribution of the data, for the normal distribution test of the connection values, the sample data is less than 0.05 for non-normal distribution, while its control value, average depth value and integration are 0.175, 0.236 and 0.214 respectively, which are all greater than 0.05 and conform to normal distribution.

4.2.3. Correlation coefficient analysis. From the perspective of aesthetic mood to explore the spatial organization characteristics and its relationship, through the organization and summary of the design elements of the Jiangnan garden, the spatial distribution of the design elements of the numerical value and the above quantitative data on the space of the Jiangnan garden to organize the formation of a statistical table, the summary of the data to be analyzed as an analytical reference. Using asymptotic significance - two-tailed test, the detection of P-value conforms to 0.05>P>0.01, the detection of variables at the 0.05 level of correlation is significant; when P<0.01, the detection of variables at the 0.01 level of correlation is significant, which indicates that the correlation is higher, and r is the correlation coefficient.

There are 873 design element data about Jiangnan gardens collected by the research experiment, and the results of correlation analysis and calculation of spatial eigenvalue correlation between design element data and aesthetic mood are as follow Table 1.

Table 1. Inspection data of Jiangnan garden

	Connection value	Control value	Average depth value	Integration degree
r	0.173*	0.301**	-0.278*	0.325**
p	0.042	0.009	0.018	0.004

**At the 0.05 level (two-tailed), the correlation is significant.

**At the 0.01 level (two-tailed), the correlation is significant.

The correlation coefficient of connection value and Jiangnan garden design elements is 0.173, p=0.042, significant correlation at 0.05 level; the correlation coefficient of spatial control value and Jiangnan garden design elements is 0.301, p=0.009, significant correlation at 0.01 level; the correlation coefficient of average depth value is -0.278, p=0.018, significant correlation at 0.05 level , negatively correlated, and the correlation coefficient for integration was -0.325, p=0.004, significant at the 0.01 level of correlation.

5. Conclusion

The article proposes a visual recognition system based on binocular vision and scene-driven parallel feature fusion pyramid structure. On this basis, a data enhancement framework for Jiangnan garden plan is constructed, and the rationality of the data enhancement framework for Jiangnan garden plan generation is verified and applied. Then the algorithm is trained by using the design scheme dataset, so that the algorithm generates diversified design schemes and analyzes the semantic information of w-vectors. Finally, the spatially quantized feature data and the data from the field research are used to analyze the correlation between the design elements of Jiangnan gardens and the aesthetic context.

In the ablation experiment, the MIOU of the masked planar map segmentation model on the training set is 0.385, and that of the unmasked planar map segmentation model is 0.293, and the IOUs of LD, ST, PZ, GZW, and ZW are enhanced to 0.402, 0.396, 0.268, 0.156, 0.367, and 0.879, respectively, after masking.The data enhancement framework constructed in this paper has a MIOU (0.389 after adding Mask, 0.376 without adding Mask) is not substantially improved over the masked planar map segmentation model, and the MIOU of the planar fast segmentation model after expanding nearly 20 times of the samples is only 0.391, which indicates that the data enhancement framework constructed in this paper is more suitable for generating the training of design-related algorithms.

For the control value, average depth value and integration degree of Jiangnan garden space sample data, the positive test is 0.175, 0.236 and 0.214, respectively, are greater than 0.05, in line with the normal distribution. There is a significant correlation between the connection value, spatial control value, average depth value, and integration degree in the aesthetic context of Jiangnan gardens and the data on the design elements of Jiangnan gardens. The correlation coefficients of connection value, spatial control value and integration degree of the research samples are 0.173, 0.301 and 0.325, respectively, which are positively correlated with the garden design elements; the correlation coefficient of average depth value is -0.278, which is negatively correlated with the design elements of Jiangnan gardens.

This paper focuses on the design elements of Jiangnan gardens and combines them with the aesthetic mood of the garden space, based on exploring the matching of the garden design elements with the aesthetic characteristics of the space, to provide a basis for the optimization of the design of the garden space.

References

[1] Hongjun, Z. H. O. U., Zhenzhen, G. U. O., Yunyi, W. U., & Huaiyun, K. O. U. (2024). Reality and Virtuality: Exploration of Contemporary Reconstruction of Jiangnan Historic Gardens. Landscape Architecture, 31(8), 55-64.

[2] Sun, C., Jiang, Z., & Yu, B. (2024). How to interpret Jiangnan gardens: a study of the spatial layout of Jiangnan gardens from the perspective of fractal geometry. Heritage Science, 12(1), 353.

[3] Chen, H., Tan, Z., & Sun, P. (2024). Research on Wind Environment Simulation in Five Types of "Gray Spaces" in Traditional Jiangnan Gardens, China. Sustainability, 16(17), 7765.

[4] Wang, C. (2023). Research on Gardening Art from the Perspective of Different Aesthetic Forms:--Taking the Example of Private Gardens in Jiangnan of the Ming Dynasty. Highlights in Art and Design, 3(2), 104-109.

[5] Liu, Y. (2023). The historical research of Chinese architecture and gardens under the transformations of historiography in the twentieth century. Journal of Asian Architecture and Building Engineering, 22(5), 2724-2732.

[6] GU, L. (2018). TRENDS IN CHINESE GARDEN-MAKING: THE QIANLONG EMPEROR (1711-99) AND GARDENS OF JIANGNAN. Garden History, 46(2), 184-195.

[7] Yiman, L. I., & Xiaomei, Y. U. A. N. (2023). "Apricot Grove Garden in Lingnan": The Formation of Jiangnan Scenery Imagery by Literati Gatherings for Apricot Blossom. Landscape Architecture, 30(10), 128-134

[8] Yanlin, Z. H. A. N. G. (2020). Study on the Art of Stone Piling in Jiangnan Gardens under the Influence of Taoist Aesthetic Thought in Ming and Qing Dynasties. Journal of Landscape Research, 12(5), 83-86.

[9] Yang, L. (2016, November). To Study the Evolution of Private Gardens of the Bridges. In 4th International Conference on Management Science, Education Technology, Arts, Social Science and Economics 2016 (pp. 476-479). Atlantis Press.

[10] Xiao, F. (2024). Analyzing the Gardening Art of Yanshan Garden with Yuanye (The Craft of Gardens). Scientific and Social Research, 6(3), 155-160.

[11] Xu, Y., & Suwanthada, P. (2024). The Suzhou Classical Gardens: An Analysis of the GardeningIdeas of Cultural Ecology and Humanity Restoration. Journal of Roi Kaensarn Academi, 9(12), 704-720.

[12] Liang, W., & Nasir, N. B. M. (2024). From Chinese Painting to Jiangnan Gardens: Inheritance and Innovation of Artistic Style. Journal of Education and Educational Research, 8(1), 27-30.

[13] Kai, G. (2022). Literati Gardens of the Jiangnan Region: Characters and Mutations. In Routledge Handbook of Chinese Architecture (pp. 102-120). Routledge.

[14] Balık, G., & Lökçe, D. E. N. İ. Z. (2019). On the Relationship of Landscape and Painting. AM Journal of Art and Media Studies, 19.

[15] Noh, L. M. B. M., Razak, H. A., & Xing, L. (2023). The Aesthetic Development of Traditional Chinese Landscape Painting in Contemporary Landscape Painting. Art and Performance Letters, 4(7), 53-63.

[16] Baomar, Z. S. A. A., & Sanad, R. (2021). A study of landscape painting development–Past, present and future perspectives. New Trends and Issue Proceedings on Humanities and Social Sciences, 8(1), 01-11.

[17] Yang, H. (2022). Critique and Discussion on Some Typical Gardens. In A Treatise on the Garden of Jiangnan: A study on the Art of Chinese Classical Garden (pp. 551-673). Singapore: Springer Nature Singapore.

[18] Tao, J., & Rodloytuk, P. (2023). Landscape Narratives of the Garden in Central Guangdong-The Case Study of the Keyuan Garden in the Qing Dynasty. Journal of Namibian Studies: History Politics Culture, 33, 3106-3118.