Applied to Art and Design Scene Visual Comprehension and Recognition Algorithm Research

Yuxin Shi
College of Arts and Media Science and Technology, College of Hubei University of Arts and Science,
Xiangyang 441025, Hubei, China

***Abstract*—Combining advanced intelligent algorithms to improve the scene visual understanding and recognition method for art design can not only provide more inspirations and creative materials for artists, but also improve the efficiency and quality of art creation, and provide scientific and accurate references of artworks. Focusing on the art design scene visual understanding and recognition problem, a scene visual understanding and recognition method based on the intelligent optimisation algorithm to optimise the structural parameters of the multilayer perception machine is proposed. Firstly, the scene visual recognition method is outlined and analyzed, and the application scheme of multilayer perceptron in the understanding and recognition problem is designed; then, for the problems of the multilayer perceptron model, such as the training does not generalize, combined with the Pond's optimization algorithm, the training parameters of the multilayer perceptron model are optimized, and the visual understanding and recognition scheme of the art design scene is designed; finally, the proposed model is verified with the image dataset, and the scene visual understanding and recognition accuracy reaches 0.98, compared with other models, the proposed method has higher recognition accuracy. This research solves the problem of scene visual understanding and recognition, and applies it to the field of art design to improve the efficiency of art design assistance.**

Keywords—Art design; scene visual understanding and recognition; multilayer perceptron; pond goose algorithm; image dataset

I. Introduction

Traditional art design methods are carried out using software or hand-drawing and so on [1]. With the rapid development of multimedia technology, the art design method based on computer technology has gradually been highly valued by experts and scholars in the field of art design [2]. At present, art design software presents diversified types and functions, which enriches the creativity of visual effects, improves the working method of art design, and becomes one of the future trends of art design research [3]. Scene visual understanding and recognition, as one of the key technologies most widely used in the field of art, uses computer vision and image processing technology to deeply understand and analyse the scenes in images and videos [4]. Art design that combines scene visual understanding and recognition algorithms can provide artists with more inspiration and creative materials, and can also improve the efficiency and quality of art creation and provide scientific and accurate references for artwork [5]. Therefore, the study of art design methods combining scene visual understanding and recognition algorithms is an important theoretical research significance for art design to assist decision-making and creation. According to the principle of design process, the research on scene visual understanding and recognition for art design generally includes the research contents of scene image segmentation, extraction of target features, and understanding and recognition model construction [6]. Scene image segmentation processes the image from colour segmentation, morphological processing, etc.; extracting target features captures the key information of the image, i.e., shape, colour, texture, edges, etc.; as a key part of the scene visual understanding and recognition problem [7], deep learning or machine learning algorithms are trained and constructed based on the annotated feature sample set. According to the principle of the core approach, scene visual understanding and recognition methods can be classified into search-based scene visual understanding and recognition methods [8], template matching-based scene visual understanding and recognition methods [9], and language model-based scene visual understanding and recognition methods [10].Sandrine et al [11] obtained image descriptions by constructing the mapping relationship between images and texts using similarity to obtain the compliant utterances; Mustafa et al [12] proposed a visual scene description scene based on coordinate position by analysing a large number of images and annotation information; Daniel et al [13] improved convolutional neural network using encoding-decoding network to further improve the accuracy of the scene visual understanding recognition model; Chen et al [14] combined the current state of the art of domestic and international research and used encoding-decoding network and attention mechanism model to construct and analyse the visual scene understanding model. Although the current scene visual understanding and recognition algorithms for art design have achieved a lot of results and application progress, there are still some challenges, such as the generalisation ability needs to be improved, more sensitive to the smile change of the data, the real-time needs to be further improved, and the consumption of computational resources is more [15].

This paper focuses on the scene visual understanding and recognition problem for art design, combines the intelligent optimization algorithm to optimize the network structure parameter framework paradigm [16], and puts forward a scene visual understanding and recognition method based on the optimization of the Pond's Goose algorithm to improve the multilayer perceptual machine model. Aiming at the characteristics of the scene visual understanding and recognition problem for art design, the scene visual understanding and recognition model algorithm and its optimisation strategy are analysed and introduced, and at the same time, the method of Pond's Goose algorithm optimisation recognition is applied to the scene visual understanding and recognition problem for art design, and the art-related data is used to compare the other five recognition models, which verifies the high efficiency of the proposed method and the high accuracy and real-time performance.

II. Overview of Identification Methods

A. Analysis of the Problem

1) Introduction to the issue: In the field of art design, scene visual understanding recognition algorithms are mainly used to automatically identify and classify images in art works, or automate the processing of art works to improve the efficiency and quality of art creation [17]. In the art market, scene visual understanding algorithms can be used for art appraisal and valuation to provide a more scientific and accurate reference for art transactions. The main applications of scene visual understanding recognition algorithms in the field of art design are shown in Fig. 1.

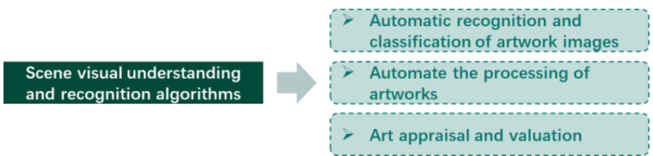

Fig. 1. Scene visual understanding recognition algorithm application.

In this paper, the scene visual understanding recognition algorithm is used to extract the target features in the complex scene image, complete the combination design of the extracted target, and realise the art assistance, so the research in this paper is mainly used to automatically identify and classify the images in the art work. For the problem of target recognition and classification in complex images, this paper adopts multi-layer perceptron to construct the mapping relationship between target features and target information, as shown in Fig. 2.

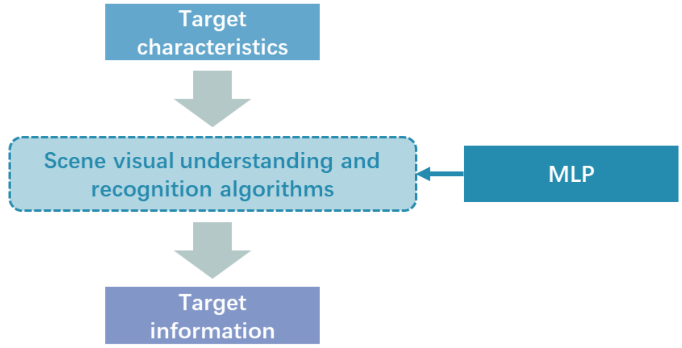

Fig. 2. Scene visual understanding and recognition solution architecture.

2) Multimodal feature extraction: In order to obtain as much information about the image target features and improve the accuracy of the understanding recognition algorithm, this paper firstly rubs the coarse segmentation technique [18] for colour segmentation of the scene image; then, the effective description region is processed morphologically; finally, the scale invariant feature transform technique is used to extract the target features in the candidate region. The multimodal feature extraction is shown in Fig. 3, and the principle of Scale Invariant Feature Transform (SIFT) [19] technique is shown in Fig. 4.

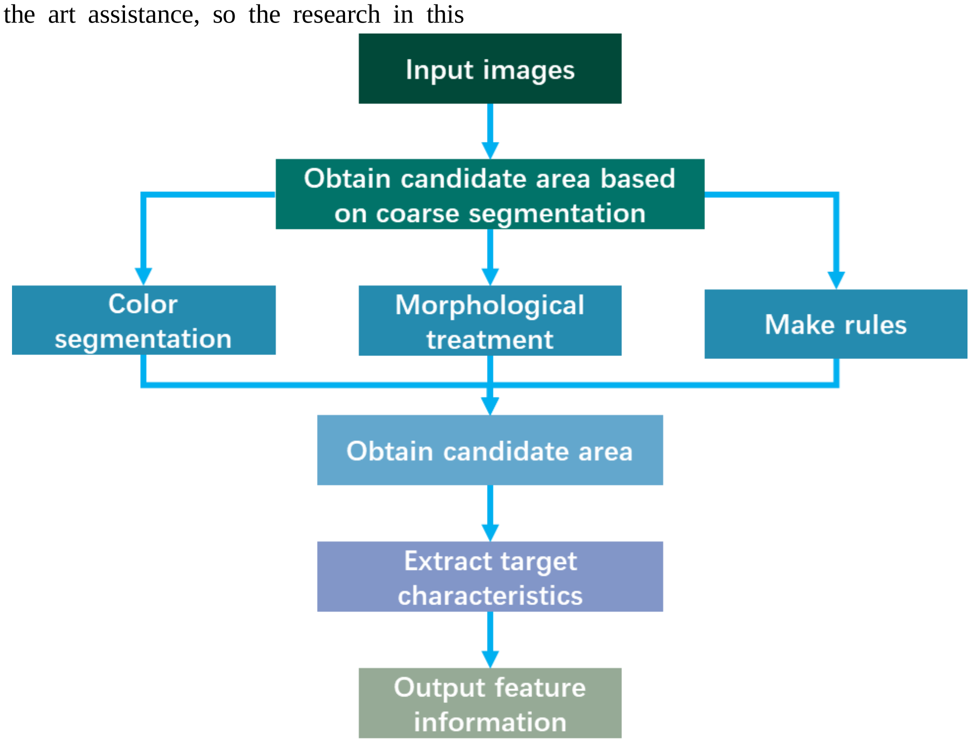

Fig. 3. Multimodal feature extraction.

Fig. 4. Principle of SIFT technology.

B. Overview of Multilayer Perceptron

Multilayer perceptrons (MLP) [20] is the introduction of multiple hidden layers on top of a single-layer neural network. MLP networks can acquire complex relationships between inputs through hidden layer nodes, the inputs undergo an activation function, which produces the final prediction through the output layer. MLP networks require at least three layers of artificial neurons. The MLP input layer will be make each node connected to the hidden layer according to the input sample set vector dimensions. The hidden layer generally uses a sigmoid function to make the feature data from the hidden layer to the output layer very smooth.

$$f(x) = 1 \Big/ 1 + \exp\left(-\sum_j w_j x_j - b\right) \tag{1}$$

Where, $f(x)$ denotes the hidden layer output, w_j denotes the node weights and b denotes the node bias.

The parameters of MLP network include the number of hidden layer nodes, hidden layer node excitation function, and connection weight. For the selection of the number of hidden layer nodes, this paper adopts an experimental approach, taking different numbers of nodes respectively, observing the recognition accuracy, and taking the number of nodes with the largest accuracy. For the hidden layer node excitation function, this paper adopts the sigmoid function as the activation function [21]; for the connection weight, MLP generally adopts the back propagation algorithm [22]. Backpropagation can be used to train feed-forward artificial neural networks with any layers and any number of hidden units, but the practical limitations of computational power will constrain the ability of backpropagation, therefore, this paper adopts the Pond's algorithm [23] as the optimisation algorithm for the selection of MLP network structure parameters.

$$C(w,b) = \frac{1}{2n}\sum_x \left\| y(x) - a \right\|^2 \tag{2}$$

$$w_k = w_k - \frac{\eta}{m}\sum_j \frac{\partial C_{X_j}}{\partial w_k} \tag{3}$$

$$b_k = b_k - \frac{\eta}{m}\sum_j \frac{\partial C_{X_j}}{\partial b_k} \tag{4}$$

where w and b denote the structural weights and biases of the MLP network, respectively.

MLP is applied to many fields, including image recognition, speech processing, language analysis, etc., and is mainly applied to complex nonlinear mapping relationship processing.

C. Application of Multilayer Perceptron Machine in Comprehension Recognition

The application of MLP to the problem of English recognition for visual understanding of scenes oriented to the field of art and design is mainly shown in Fig. 5. The input vector of MLP is the target features extracted from the image, and the output vector is the category of the target.

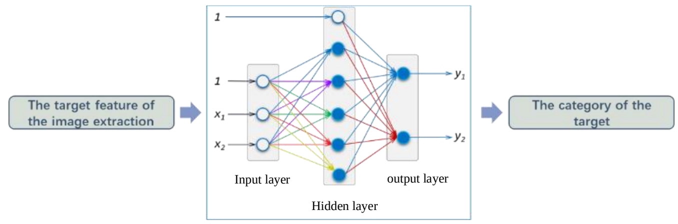

Fig. 5. MLP applications

III. Recognition Method Optimisation Algorithm

In order to improve the recognition accuracy of the multilayer perceptron network, this paper chooses the pond goose algorithm as the recognition method optimisation algorithm, the specific optimisation process is as follows:

A. Pond's Goose Algorithm

The Gannet Optimization Algorithm (GOA) [24] is a natural heuristic optimization algorithm that mimics the behaviour of the pond goose. The algorithm is used to explore optimal solutions in the search space by mathematically modelling the U- and V-diving behaviours of the Pond Goose during foraging as well as sudden rotations and random wandering. The GOA is designed to balance the capabilities of global exploration and local exploitation in order to improve the efficiency of solving engineering optimisation problems.

1) Initialisation phase: The GOA algorithm uses a random initialisation strategy for population generation, which generates uniformly distributed GOA population locations from each dimension using upper and lower bounds of the search space.

2) Exploration phase: The exploration phase of the GOA algorithm simulates the U- and V-diving behaviour of the pond goose after finding prey in the air (shown in Fig. 6) for global search.

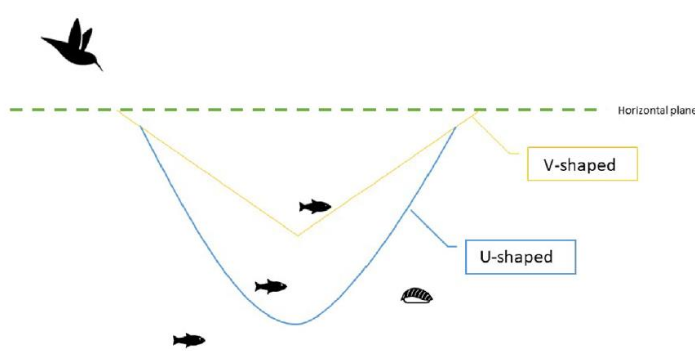

Fig. 6. The GOA algorithm U- and V-dive behaviour.

Pond geese dive in a U-shaped dive when prey is at greater depths:

$$a = 2\times\cos\left(2\times\pi\times r_2\right)\times t_1 \tag{5}$$

Pond geese dive in a V-dive when prey is in a relatively shallow position:

$$a = 2\times V\left(2\times\pi\times r_3\right)\times t_1 \tag{6}$$

$$V\left(x\right)=\begin{cases}-\dfrac{1}{\pi}\times x+1 & x\in\left(0,\pi\right)\\ \dfrac{1}{\pi}\times x11 & x\in\left(\pi,2\pi\right)\end{cases} \tag{7}$$

$$t_1 = 1-\frac{t}{T_{\max}} \tag{8}$$

Where t denotes the number of contemporary iterations, $T_{\max}$ denotes the maximum number of iterations, and r_2 and r_3 denote random numbers, respectively.

The position of the pond goose is updated by introducing a random variable q and randomly selecting either a U-dive approach or a V-dive approach:

$$MX_i\left(t+1\right)=\begin{cases}X_i\left(t\right)+u_1+u_2 & q\geq 0.5\\ X_i\left(t\right)+v_1+v_2 & q<0.5\end{cases} \tag{9}$$

$$u_2 = A\times\left(X_i\left(t\right)-X_r\left(t\right)\right) \tag{10}$$

$$v_2 = B\times\left(X_i\left(t\right)-X_m\left(t\right)\right) \tag{11}$$

$$A=\left(2\times r_4-1\right)\times a \tag{12}$$

$$B=\left(2\times r_5-1\right)\times b \tag{13}$$

Where r_4 and r_5 denote random numbers, u_1 is a random number between -a and a, v_1 is a random number between -b and b, $X_i\left(t\right)$ denotes the position information of the i-th song pond goose individual in the t-th iteration, $X_r\left(t\right)$ denotes the position of the randomly selected pond goose individual, $X_m\left(t\right)$ denotes the average of the position of all the pond geese individuals, and the calculation is as follows:

$$X_m\left(t\right)=\frac{1}{N}\sum_{i=1}^{N}X_i\left(t\right) \tag{14}$$

3) Development phase: The GOA algorithm development phase simulates the capture behaviour of the pond goose in the water, and based on the capture ability of the pond goose decides whether to perform a random swim or a sudden rotation to capture the prey (as shown in Fig. 7).

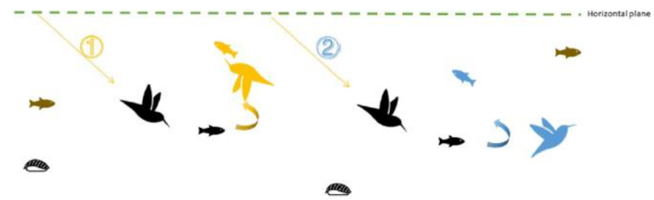

Fig. 7. GOA algorithm development phase.

When the pond goose enters the water, its catching ability C is greater than or equal to c, it will suddenly rotate to catch fish; its catching ability C is less than c, it will give up catching fish and march randomly, the Levy flight model is used to simulate the marching of the pond goose, and the specific model of the position update is as follows:

$$MX_i\left(t+1\right)=\begin{cases}t_1\times\delta\times\left(X_i\left(t\right)-X_{best}\left(t\right)\right)+X_i\left(t\right) & C\geq c\\ X_{best}\left(t\right)-\left(X_i\left(t\right)-X_{best}\left(t\right)\right)\times P\times t_2 & C<c\end{cases} \tag{15}$$

$$C=\frac{1}{R\times t_2} \tag{16}$$

$$t_2 = 1+\frac{t}{T_{\max}} \tag{17}$$

$$R=\frac{M\times vel^2}{L} \tag{18}$$

$$L=0.2+\left(2-0.2\right)\times r_6 \tag{19}$$

$$\delta = C\times\left|X_i\left(t\right)-X_{best}\left(t\right)\right| \tag{20}$$

$$P=Levy\left(D\right) \tag{21}$$

$$Levy\left(D\right)=0.01\times\frac{\mu\times\sigma}{\left|v\right|^{\frac{1}{\beta}}} \tag{22}$$

$$\sigma=\left(\frac{\Gamma\left(1+\beta\right)\times\sin\left(\dfrac{\pi\beta}{2}\right)}{\Gamma\left(\dfrac{1+\beta}{2}\right)\times\beta\times 2^{\frac{\beta-1}{2}}}\right)^{\frac{1}{\beta}} \tag{23}$$

Where, r_6 is a random number, M denotes the mass of the pond goose, vel denotes the velocity of the pond goose with the value of 1.5 m/s, c is a constant, which generally takes the values of c=0.2 and β=1.5, $X_{best}\left(t\right)$ denotes a constant, and μ and σ are random numbers, respectively.

The pseudo-code and flowchart of the GOA algorithm are shown in Fig. 8 and Fig. 9, respectively.

```
Algorithm1: Gannet Optimization Algorithm (GOA)
Set GOA parameters;
Initialize GOA population based on random distribution;
Obtain memory matrix MX;
Calculate X fitness;
While t<=Max_t
    if rand>0.5
        for MXi do
            if q>=0.5, update position using U-shaped;
            else, update position using V-shaped;
            end if
        end for
    else
        for Mxi do
            if c>=0.2, update position with a sudden tuning;
            else, update position with Levy movement;
        end
    end
   Calculate MXi fitness, and update better position;
End while
```

Fig. 8. Pseudo-code of the GOA algorithm.

B. Optimisation of Recognition Methods by the Pond's Goose Algorithm

In order to increase the accuracy of the multilayer perceptron recognition method, this paper uses the Pond's Goose algorithm to optimise the multilayer perceptron network. The paradigm of the Pond's Goose algorithm to optimise the multilayer perceptual machine network is shown in Fig. 10, and the GOA algorithm takes w and b as the optimisation variables, and $C(w,b)$ as the fitness function. The flow of the MLP scene visual understanding recognition method based on the GOA algorithm is shown in Fig. 11, and the specific steps are as follows:

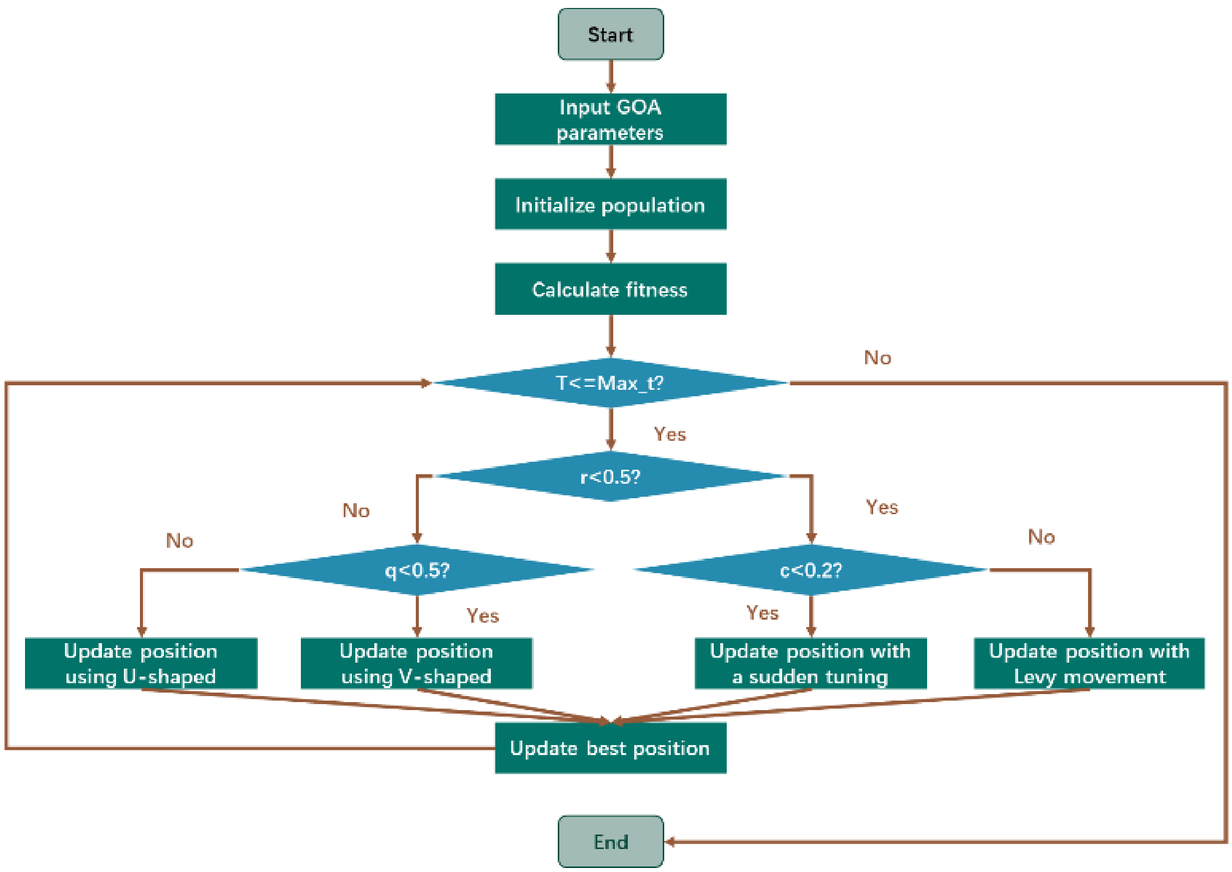

Fig. 9. Flowchart of GOA algorithm.

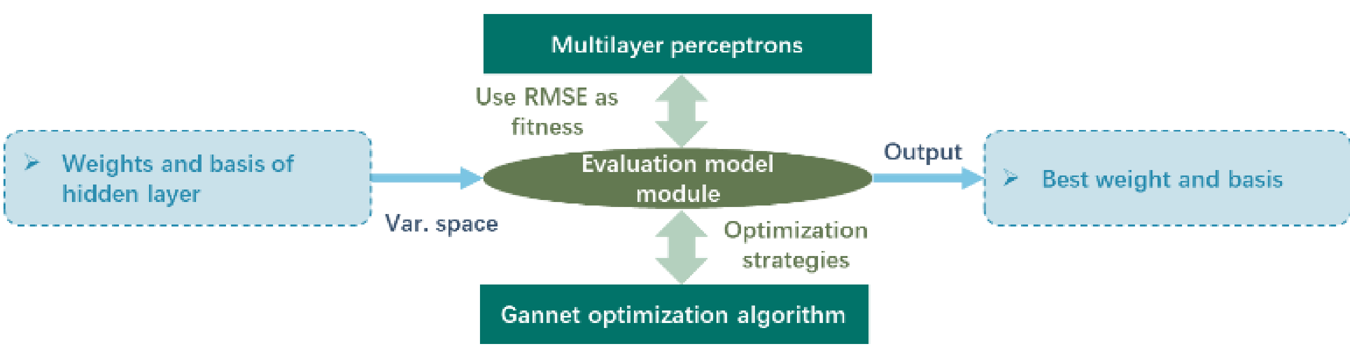

Fig. 10. GOA algorithm to optimise the MLP network paradigm.

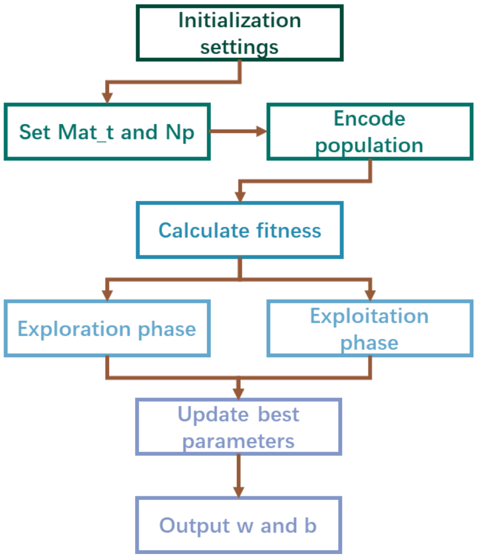

Fig. 11. GOA-MLP step-by-step diagram.

- Step 1: Initialisation setup. 1) Set the maximum number of iterations as well as the number of populations and algorithmic control parameters for the GOA algorithm optimised MLP network; 2) Initialise the GOA algorithm populations using real number coding;
- Step 2: Calculate the $C(w,b)$ value and determine the optimal structural parameters for the current number of iterations based on the error value;
- Step 3: Behavioural simulation models such as U- and V-diving behaviours during foraging, as well as sudden rotations and random wandering, were used to update information on the location of individuals in the population;
- Step 4: Calculate the fitness value, update the structural parameters, and determine whether the maximum number of iterations or the optimal solution of the GOA-MLP algorithm is no longer changing;
- Step 5: Output the structural parameters of the optimal MLP model.

IV. Understanding the Application of Recognition Methods in Art and Design

A. Application Programmes

The visual understanding and recognition method of art design scene based on GOA-MLP model performs colour segmentation and morphological processing of the scene through coarse segmentation technology and formulates rules to extract candidate scene regions, uses SIFT technology to extract internal features from the feature regions, annotates the sample data, and trains the scene visual understanding and recognition algorithm based on the GOA-MLP model, to complete the visual understanding and recognition of the scene based on the art design [25]. Focusing on the principle of visual understanding and recognition method of art design scene based on GOA-MLP model, this section designs the visual understanding and recognition scheme of art design scene based on GOA-MLP model (shown in Fig. 12), and gives the application analysis.

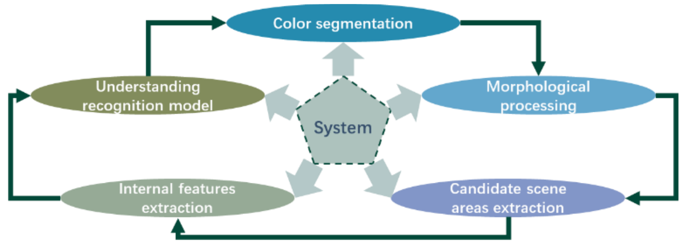

Fig. 12. Artistic design scene visual understanding and recognition programme.

In Fig. 12, the visual understanding and recognition system for art and design scenes based on GOA-MLP model consists of modules such as colour segmentation, morphological processing, extraction of candidate scene regions, extraction of internal features, and construction of scene visual understanding and recognition model.

B. Application Steps

Based on the analysis of the Art and Design Scene Visual Understanding and Recognition Programme, the modules are described below:

1) Colour segmentation module: For the colour segmentation problem, HSV colour space is used, where H is converted by the following formula:

$$H=\begin{cases} 0 & \max=\min \\ 60\times\dfrac{G-B}{\max-\min} & \max=R\,\&\,G\geq B \\ 60\times\dfrac{G-B}{\max-\min}+360 & \max=R\,\&\,G<B \\ 60\times\dfrac{G-B}{\max-\min}+120 & \max=G \\ 60\times\dfrac{G-B}{\max-\min}+240 & \max=B \end{cases} \tag{24}$$

Where $\max$ and $\min$ denote the maximum and minimum values of the pixel for each channel in the RGB colour space, respectively.

2) Morphological processing module: To address the problem of noise and breakage in colour segmented images, this subsection uses morphological processing methods to reduce the effect of noise and to obtain connected regions. In morphological processing technique, the broken portion of the candidate region is re-cracked using an expansion operation to fill the broken contour lines [26].

3) Candidate scene area analysis module: Aiming at the problem that there are still uneliminated interference regions in the image after morphological processing, this paper uses constraints to extract candidate scene regions. The specific satisfaction conditions are as follows:

$$\begin{cases} S_i \geq S_{\min} \cap S_i \leq S_{\max} \\ \dfrac{L_i}{W_i} \geq \left(\dfrac{L}{W}\right)_{\min} \cap \dfrac{L_i}{W_i} \leq \left(\dfrac{L}{W}\right)_{\max} \\ \dfrac{S_i}{L_i \times W_i} \geq \left(\dfrac{S}{L \times W}\right)_{\min} \end{cases} \tag{25}$$

Among them, L , W and S denote the width, height and area of the connected area, respectively. $S_{\min}$, $S_{\max}$ and $\left(S/\left(L\times W\right)\right)_{\min}$ denote the minimum and maximum values of the area of the connected area, respectively; $\left(L/W\right)_{\min}$ and $\left(L/W\right)_{\max}$ denote the maximum value of the aspect ratio of the connected area, respectively; and denotes the minimum value of the duty cycle.

4) Internal feature extraction module: In order to reduce the computational complexity and retain the key information, this subsection adopts the SIFT technique to extract the target features inside the candidate region. The idea of SIFT-based target feature extraction method is to take the SIFT feature point as the centre point, calculate the gradient and direction of the pixels within the range of 16×16 , and at the same time use dense sampling to obtain more information.

5) Scene visual understanding recognition model building module: In order to complete the art design method based on the scene visual understanding recognition algorithm, this paper uses GOA-MLP algorithm training to construct the scene visual understanding recognition model, and the specific training process is shown in Fig. 13. Firstly, the data is divided, then the MLP model is improved by using GOA optimisation, secondly, the MLP structure parameters are reconstructed by using the training set, and finally the understanding recognition model is used to complete the art design.

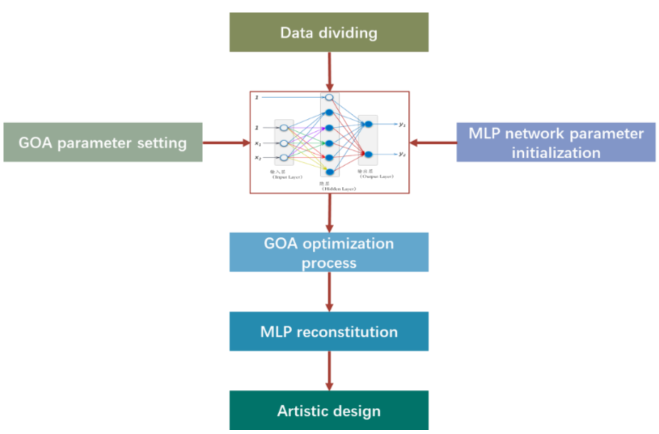

Fig. 13. Scene visual understanding recognition model building module.

V. Simulation Experiment

A. Experimental Set-up

In order to verify that the GAO algorithm improves the accuracy of the MLP model, this paper uses the sine-cosine optimisation algorithm (SCA) [27], the Harris hawk optimisation algorithm (HHO) [28], the butterfly optimisation algorithm (BOA) [29], the crow's tern optimisation algorithm (STOA) [29], and the gannet optimisation algorithm (GOA) to make comparisons with the specific parameter settings as shown in Table I. The structure of the MLP model to be optimised includes one input layer, three hidden layers and one output layer, the activation function is a Sigmoid function, the number of nodes in the hidden layer is divided into, the number of populations of SCA, HHO, BOA, STOA and GOA algorithms is 100, the maximum number of iterations is 500, and the condition for satisfying the optimal result output is to reach the maximum number of iterations.

TABLE I. Experimental Parameter Settings

Arithmetic	Parameterisation
MLP	The activation function is Sigmoid function and the classifier is Softmax, including input layer, hidden layer (3 layers), output layer, hidden layer nodes are [30]
SCA-MLP [27]	The MLP structure parameters are set as above, with a set to 2
HHO-MLP [28]	The MLP structure parameters are set as above, E0 is a random number and E1 decreases linearly from 2 to 0
BOA-MLP [29]	The MLP structural parameters are set as above, with a transition probability of 0.6, a force index of 0.1 and a perceptual mode value of 0.01
STOA-MLP [30]	Sa decreases linearly from 2 to 0.
GOA-MLP	MOP_max is 1, MOP_min is 0.2, Alpha is 5, Mu is 0.499

This paper uses the scene visual understanding feature extraction method parameter settings are shown in Table II.

TABLE II. Parameter Settings for Visual Feature Extraction Methods

Project	Parameter values	Note
Optimal thresholds	90	RGB colour space
Optimal lower and upper thresholds	200, 280	HSV colour space
S_{max}	400	Maximum area of connected areas
S_{min}	(L x W)/2	Minimum area of connected areas
$(L/W)_{max}$	0.2	Maximum Aspect Ratio for Connected Areas
$(L/W)_{min}$	4.6	Minimum aspect ratio of connected areas
$(S/(L\times W))_{min}$	0.7	Duty Cycle Min.

The hardware platform for the experimental development environment is Pentium IV 3.7 GHz CPU and 4 RAM with a memory size of 3 GB. The software environment includes VisualStudio, Win10 operating system, Multigen Creator, Visual C++ 4.0 programming language, Matlab2021a programming language.

The dataset was adopted from the image dataset of the Massachusetts Institute of Technology (MIT), including 2688 scene images, including scenes of mountains, cities, coasts, flowers, etc. as shown in Fig. 14.

Fig. 14. Sources of data sets.

B. Analysis of Performance Test Results

In order to verify the effectiveness as well as the efficiency of the method proposed in this paper, this paper firstly analyses the visual processing methods of scenes oriented to artistic design, and secondly, tests and analyses the recognition methods for comparison.

1) Scene visual processing analysis: According to the above simulation environment, parameter settings and data set for target information extraction, the extraction results are obtained as shown in Fig. 15.

Obviously, as obtained from Fig. 15, the scene visual understanding algorithm can effectively extract the floral information of the known image, solve the problem of poor clarity of the output image, reduce the influence of noise, and improve the overall design of the image.

2) *Analysis of identification test results:* In order to enhance the effect of validation experiments, this paper randomly extracts five sets of data sets from the database to carry out experimental analysis, each time the extracted data sets include 2000 images, of which the training set reaches 1500, the test set is 350, and the rest is the validation set, and the specific test results are shown in Table III, Fig. 16, and Fig. 17.

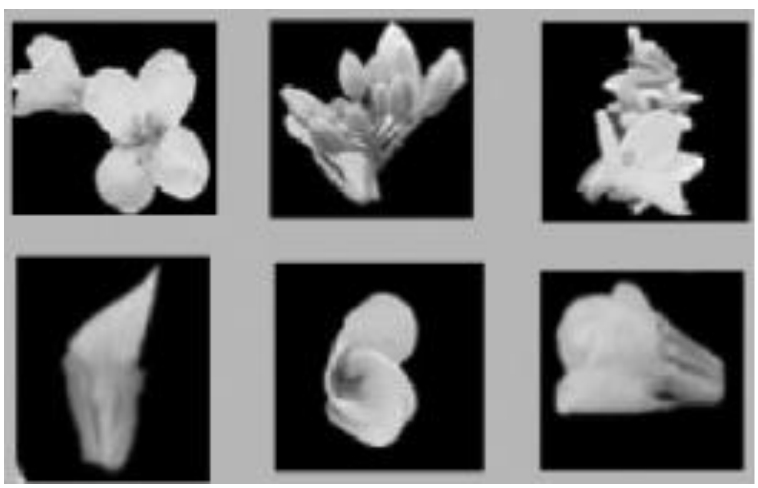

Fig. 15. Target information extraction results.

From Table III and Fig. 16, it is easy to find that from the dataset point of view, the accuracy of GOA-MLP comprehension and recognition method is better than that of MLP, SCA-MLP, HHO-MLP, BOA-MLP, and STOA-MLP.It can be seen that the accuracy of MLP comprehension and recognition method based on optimisation algorithms is better than that of MLP method, and the accuracy of GOA optimised MLP model is better than other optimisation algorithms.

Fig. 17 demonstrates the optimisation convergence curves for each algorithm of SCA-MLP, HHO-MLP, BOA-MLP, STOA-MLP, and GOA-MLP. From Fig. 17, it can be obtained that: with the increase in the number of iterations, the value of the adaptation function of the comprehension recognition accuracy of the GOA optimised MLP model increases, and the accuracy of the convergence is higher than that of the algorithms such as SCA, HHO, BOA, and STOA; and the optimised convergence of the GOA-MLP model is stable up to the vicinity of 0.98.

TABLE III. DATA SET TEST IDENTIFICATION ACCURACY RESULTS (%)

Data set	MLP	SCA-MLP	HHO-MLP	BOA-MLP	STOA-MLP	GOA-MLP
1	67.34	86.98	87.32	84.22	93.90	97.98
2	66.21	86.10	87.35	83.78	92.53	97.46
3	69.88	88.38	89.25	87.90	95.16	98.85
4	65.20	86.12	86.95	83.15	90.74	97.00
5	67.58	87.24	88.65	85.37	92.88	98.05

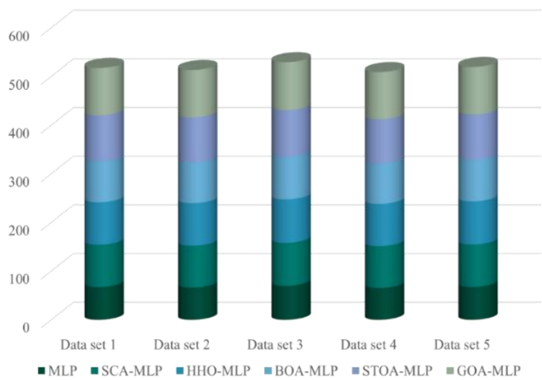

Fig. 16. Recognition results of each algorithm.

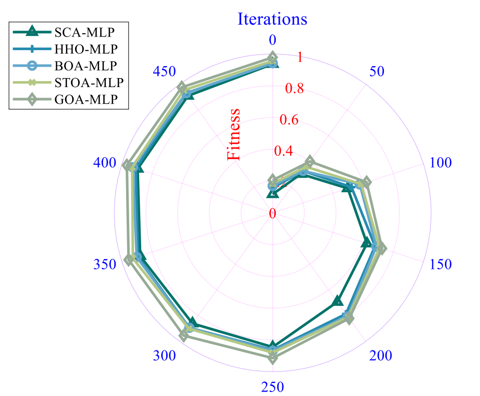

Fig. 17. Convergence process of each optimisation algorithm.

VI. CONCLUSION

Aiming at the problem of visual understanding and recognition effect of diverse scenes of art design, this paper proposes a visual understanding and recognition method of scenes oriented to art design based on GOA-MLP, which makes the GOA algorithm optimise the MLP structural parameters and improve the understanding and recognition accuracy. Experiments with MIT image datasets lead to the following conclusions: using the GOA algorithm to optimise the MLP structural parameters and using it in the art design scene visual understanding and recognition problem improves the art scene understanding and recognition accuracy and increases the art design effect. By analysing five types of image datasets, the GOA-MLP based scene visual understanding and recognition method has higher recognition accuracy. This study only focuses on the accuracy of the understanding and recognition method, and the subsequent re-recognition time method still requires in-depth research.

REFERENCES

[1] Zhang F S , Ge D Y , Song J , Xiang W J. Outdoor scene understanding of mobile robot via multi-sensor information fusion[J]. Information Integration, 2022.

[2] Glavan A ,Talavera, Estefanía.InstaIndoor and multi-modal deep learning for indoor scene recognition[J].Neural Computing and Applications,. 2022, 34(9):6861-6877.

[3] Zhang W , Wang Y , Ni B , Yang X. Fully context-aware image inpainting with a learned semantic pyramid[J]. Pattern Recognition Society, 2023:143.

[4] Rafique A A , Ghadi Y Y , Alsuhibany S A , Chelloug S A, Jalal A, Park J. CNN Based Multi-Object Segmentation and Feature Fusion for Scene Recognition[J]. Computers, Materials and Continuum (English), 2022.

[5] Nourali K , Dolkhani E .Scene text visual question answering by using YOLO and STN[J].International Journal of Speech Technology, 2024, 27(1):69-76 .

[6] Su T , Shi Y , Xie C , Luo W, Ye H, Xu L. A hybrid loss balancing algorithm based on gradient equilibrium and sample loss for understanding of road scenes at basic-level[J].Pattern Analysis and Applications, 2022, 25(4):1041-1053.

[7] Narazaki Y , Pang W , Wang G Chai W. Unsupervised Domain Adaptation Approach for Vision-Based Semantic Understanding of Bridge Inspection Scenes without Manual Annotations[J].Journal of Bridge Engineering, 2024, 29(2):4023118.1-4023118.16.

[8] Bendall R C A , Eachus P , Thompson C .The influence of stimuli valence, extraversion, and emotion regulation on visual search within real-world scenes [J].Scientific reports, 2022, 12(1):948.

[9] Arif A , Ghadi Y Y , Alarfaj M , Jalal A, Kamal S, Kim D S. Human Pose Estimation and Object Interaction for Sports Behaviour[J]. Computers, Materials and Continuum (English), 2022(7):18.

[10] Pan Y .On visual understanding[J].Frontiers of Information Technology & Electronic Engineering, 2022, 23(9):1287-1289.

[11] Sandrine B , Cohen A A , Paul F .How iconic news images travel: republishing and reframing historic photographs in Israeli newspapers[J]. Communication, 2022(1):1.

[12] Mustafa A , Russell C , Hilton A .4D Temporally Coherent Multi-Person Semantic Reconstruction and Segmentation[J]. Computer Vision, 2022, 130(6):1583-1606.

[13] Daniel S P .Recognition of Family Life by Children Living in Kinship Care Arrangements in England[J].

[14] Chen S , Demachi K , Dong F .Graph-based linguistic and visual information integration for on-site occupational hazards identification[J]. Automation in construction, 2022(5):137.

[15] Pan Y .On visual understanding[J]. Frontiers in Information and Electronic Engineering: English Edition, 2022, 23(9):1287-1289.

[16] Ogiela U , Snasel V .Predictive intelligence in evaluation of visual perception thresholds for visual pattern recognition and understanding[J]. Information Processing & Management: Libraries and Information Retrieval Systems and Communication Networks: an International Journal, 2022(2):59.

[17] Tu J , Wu G , Wang L .Dual Graph Networks for Pose Estimation in Crowded Scenes[J].International Journal of Computer Vision, 2024(3):132.

[18] Fang Y .Instantaneous visual imaging of latent fingerprints in water[J]. Science in China:Chemistry, 2022.

[19] Kishinami H , Itoyama K , Nishida K , Nakadai K. Visual Scene Reconstruction based on Echolocation with a Generative Adversarial Network[J]. of the Robotics Society of Japan, 2022, 40(4):351-354.

[20] ZHANG Xiaoyan, XIANG Mian, ZHU Li, ZHOU Bintao, LIU Hongxiao, DUAN Yaqin. Ultra-short-term wind power prediction based on MLP-BiLSTM-TCN combination[J]. Journal of Hubei University for Nationalities (Natural Science Edition),2023,41(04):513-519+529.

[21] Yanni Lu. Atmospheric temperature prediction based on MLP and Transformer model[J]. Journal of Yuncheng College,2024,42(03):43-47.

[22] HOU Kepeng,BAO Guangtuo,SUN Huafen.Application of SSA-MLP model in rocky slope stability prediction[J]. Journal of Safety and Environment,2024,24(05):1795-1803.

[23] QI Huiling, HU Hongping, BAI Yanping, HOU Qiang. Prediction of novel coronavirus outbreak based on optimised BP neural network with improved pond goose algorithm[J]. Journal of Shanxi University (Natural Science Edition),2023,46(06):1283-1292.

[24] Shieh C S .A Parallel Compact Gannet Optimization Algorithm for Solving Engineering Optimization Problems[J].Mathematics, 2023, 11.

[25] Lu K, Miao Tenghui. Research on scene visual understanding algorithm applied to art assisted design[J]. Modern Electronic Technology,2020,43(09):37-40.

[26] Z. M. Wang, X. Wang, G. Li, F. T. Zhang. A review of visual scene understanding[J]. Journal of Xi'an University of Posts and Telecommunications,2019,24(01):1-15.

[27] Liu LQ,Chen F. PCNN parameter-optimised image fusion with sine-cosine dynamic interference Harris Hawk algorithm[J]. Software Guide,2024,23(03):62-70.

[28] Murugesan S , Suganyadevi M V .Performance Analysis of Simplified Seven-Level Inverter using Hybrid HHO-PSO Algorithm for Renewable Energy Applications[J].Iranian Journal of Science and Technology, Transactions of Electrical Engineering, 2024, 48(2):781-801.

[29] Badi M , Mahapatra S .Optimal reactive power management through a hybrid BOA-GWO-PSO algorithm for alleviating congestion[J].International Journal of System Assurance Engineering and Management, 2023, 14:1437 - 1456.

[30] He Huan. Fault diagnosis of rolling bearing of wind turbine based on STOA-XGBoost[J]. Heilongjiang Science,2024,15(12):30-33.