Layout Optimization of Visual Communication Design Based on Particle Swarm Optimization Algorithm

Cheng Zhou

Hanshan Normal University, Chaozhou, Guangdong 521041, China,
zc@hstc.edu.cn

Corresponding author: Cheng Zhou, zc@hstc.edu.cn

Abstract. This article aims to explore a new method of computer-aided visual communication design (VCD) layout optimization and improve the automation and intelligence level of design layout by integrating particle swarm optimization (PSO) and deep learning technology. An innovative layout optimization framework is proposed, which combines the advantages of convolutional neural network (CNN) in image feature extraction and the ability of Rotation-Shift Transformer Networks (RSTN) in layout fine-tuning, and at the same time uses PSO algorithm to search globally to find the optimal or nearly optimal design layout scheme. The experimental results show that this method has obvious advantages in extracting the precision of image features of design sample layout, and the average precision is improved by about 15% compared with other methods. In addition, algorithms have significant advantages in verifying the retention of computational information. Especially in terms of dimensionality reduction time and feature extraction dimensions, the average speed of our algorithm has been improved by over 30%. This provides a clearer idea and verification effect for the practical design of PSO algorithm layout optimization.

Keywords: Computer-Aided Design; Visual Communication Design; Layout Optimization; Particle Swarm Optimization Algorithm
DOI: https://doi.org/10.14733/cadaps.2025.S4.211-223

1 INTRODUCTION

In recent years, the gradual decrease in computing costs and the rapid growth of market demand have led to the emergence of various complex applications based on visual communication. As one of the core issues in the field of computer vision, visual communication has broad application prospects in both military and civilian applications [1]. In civilian use, visual communication systems can replace humans to complete tedious and repetitive tasks related to visual communication, effectively reducing personnel and improving work efficiency. Therefore, it has attracted many commercial companies to invest funds and technology personnel for research. In military terms, compared to active systems such as radar, visual communication systems generally use optical devices such as cameras and operate in passive mode. It does not radiate radio waves outward, making it less likely to be detected by enemy electronic surveillance equipment, and has higher concealment and anti-interference capabilities [2]. In the field of computer vision, the main task of tracking and communication algorithms is to find continuous correspondences of image structures representing target regions or target features in a continuous video sequence. Use the external and internal edges of the target to represent it. Mainly used for situations where there are independent parts and each part is connected [3]. The target is the region of interest in the image sequence that can be analyzed. Ships in the sea, birds in the sky, cars on the road, red blood cells in the blood, etc. The algorithm first needs to solve three problems: how to represent the target in the image. Secondly, how to use image information; The third is how to model the target's actions, appearance, shape, etc. Humans can recognize targets because these regions of interest have certain features in the image sequence, which enable us to identify targets from the background and other interfering information. These complex regions can generally be represented using their abstract spatial structure, with the target represented by a central point or set of points [4]. The target is represented by simple geometric shapes such as rectangles and ellipses. Each independent part is expressed in simple geometric shapes, and there are connections between each part. Similar to the relational model, the independent parts are represented by the central axis to represent the entity. Therefore, describing these features mathematically and maximizing the use of the image information they contain is a key aspect of visual communication. The ideal image features should enable the target information to have distinguishability, reliability, independence, and low dimensionality.

The target feature information is modelled using probability density models, which can be divided into parametric and non-parametric methods. Common parameterized probability density models include Gaussian and mixture Gaussian models, while non-parametric target apparent probability density can often be estimated using histograms, kernel density estimation (KDE), etc. Templates are often formed by the simple geometric shape or contour information of the target, which has the advantage of encompassing both spatial and visual information. However, templates often only contain information in a single state, which is not suitable for situations where the target undergoes significant changes [5]. Dynamic Appearance Model: The dynamic appearance model can simultaneously model the shape and image features of the target. Similar to contour-based methods, dynamic epigenetic models model contours or internal regions of contours, which can be colour, texture, or other gradient features. Dynamic epigenetic models often require a learning process from samples, which can simultaneously establish target shape and image feature information. This model can be modelled by principal component analysis (PCA) [6]. The likelihood probability density model represents the degree of similarity between observation and propagation, while in image-based particle filter tracking problems, it represents the degree of similarity between each image subregion and the target template determined by particles. Common colour feature extraction methods include colour histogram, colour set, colour moment, colour aggregation, vector, colour correlation graph, etc. One is to select features and express them mathematically, and the other is to select functions that represent feature similarity [7]. From a human visual perspective, colour features are a fundamental visual characteristic for humans to perceive and distinguish different objects. The likelihood model plays a decisive role in tracking performance, and an excellent likelihood model should be able to distinguish between tracked and non-tracked targets effectively. In the field of visual tracking, colour is a robust feature that is suitable for describing deformable targets and has good stability against plane rotation, non-rigid bodies, and partial occlusion [8]. Deep learning features capture a wider range of visual features, such as shape contours, texture details, and lighting effects, making similarity measurement more in line with human visual perception habits. This similarity measurement method based on deep learning can not only improve the accuracy of retrieval but also make the retrieval results closer to users' visual expectations.

Although traditional pooling operations are effective in reducing data dimensions and extracting key features, their limitations lie in ignoring the rich details within the view and the complex spatial relationships between views [9]. The view level attention mechanism in 3D2SeqViews simulates this process by learning the degree of attention that different shape classes pay to each view, dynamically adjusting the visual focus to highlight key views and appropriately weaken secondary information

[10]. At the level of visual communication, this is equivalent to losing the delicate depiction of the story and the dynamic connection of the scene. The proposal of 3D2SeqViews is aimed at breaking this limitation [11]. At the same time, CNN, as a representative in the field of deep learning, has made remarkable achievements in image recognition, feature extraction and style transfer with its powerful image processing ability. Through the design of the convolution layer, pool layer and other structures, CNN can effectively extract the key features in the image, and classify and identify the image. RSTN can better model the spatial transformation relationship in images by introducing the parameters of rotation and translation transformation [12]. In the task of layout adjustment, RSTN can automatically adjust the position and angle of elements according to the input design elements and layout requirements, thus generating a more reasonable and beautiful layout scheme. Although this study mainly focuses on biological macromolecules, its visualization methods and ideas also have some enlightenment to VCD. Combining RSTN with PSO and CNN is expected to realize automatic optimization and fine-tuning of design layout and further enhance the diversity of design.

Based on the above analysis, this study aims to integrate the PSO algorithm, CNN and RSTN, and propose a computer-aided VCD layout optimization method. This method will make full use of PSO's global search ability, CNN's image recognition and feature extraction ability, and RSTN's layout fine-tuning ability to jointly build an efficient and intelligent design layout optimization framework. Through this framework, designers can explore different layout schemes more conveniently and find the optimal or nearly optimal layout results quickly.

The significance of this study lies in the following points:

(1) The combination of the PSO algorithm, CNN and RSTN is applied to the layout optimization of VCD, which provides a new idea for the research in this field.

(2) The proposed optimization method can significantly improve the design efficiency and quality, and reduce the workload of designers. In practical application, this method can help designers quickly generate various layout schemes and provide optimization suggestions, thus speeding up the design process.

(3) This study will deeply explore the application mechanism of different algorithms in VCD, and reveal their optimization effects and limitations, thus providing a useful supplement for theoretical research in related fields.

2 THEORETICAL AND TECHNICAL BASIS

Liu et al. [13] applied a deep learning-based object detection model to a visual SLAM system based on the ORB-SLAM2 system and conducted research on visual SLAM in dynamic environments. The gradual application of Simultaneous Localization and Mapping (SLAM) technology in the process of robot intelligence reflects the further improvement of the technological development level. The real environment is complex and dynamic, and dynamic targets in the scene can cause significant errors in the pose estimation of the system, reducing its positioning performance. This technology includes basic positioning and mapping functions, providing support for indoor navigation and path planning of robots. When using a camera as its underlying hardware, the research field formed is called visual SLAM. Currently, most visual SLAM systems can operate stably based on the assumption that the operating environment is static or rigid. When dealing with the secondary filtering of dynamic feature points, if the traditional epipolar geometric constraint algorithm selects an unreasonable distance threshold range from the matching point to the epipolar line, it can easily cause problems of excessive or insufficient filtering. Combining Ghost-YOLOv5s with an improved epipolar geometry constraint algorithm for dynamic feature point filtering, improving the positioning accuracy of the system, and constructing a three-dimensional semantic point cloud map without dynamic target information in dynamic scenes. Nie et al. [14] proposed an improved epipolar geometry constraint algorithm for filtering out quadratic dynamic points, combined with feature points obtained from static target boxes through object detection, and changed its threshold-solving method. The visual ORB-SLAM2 system in dynamic environments utilizes its original keyframes to construct maps with a large number of dynamic target ghosting and excessive redundant information, resulting in poor readability of the maps. Therefore, the generated point cloud map has poor readability and lacks necessary semantic information, which further affects the construction effect of the map. The existing deep learning-based object detection model YOLOv5s has a large number of convolutional layer parameters and high computational requirements. Combining with SLAM systems is not conducive to deployment in mobile or embedded edge environments. Therefore, it is proposed to apply the Ghost-YOLOv5s object detection model to SLAM systems for dynamic feature point filtering. Replace most of the convolutions in YOLOv5s with Ghost convolutions to make the model more lightweight and real-time. Use interval filtering, image similarity detection, and depth data detection to jointly screen keyframes. Accurately distinguish between dynamic and static points in the scene, filter out true dynamic points, and improve the excessive or insufficient filtering of feature points. Then, the semantic information obtained from the Ghost YOLOv5s object detection model in this article is applied to filter out dynamic information in the scene, and dense point clouds and octree maps are constructed in dynamic scenes.

The use of a fixed particle propagation radius in traditional particle filtering tracking algorithms enables real-time adjustment of particle propagation based on tracking conditions, effectively improving the sampling efficiency of particles. Sager et al. [15] proposed an adaptive particle propagation method based on tracking performance, which is a weak point. We focused on the sampling part of particle filtering and gradually optimized the particle distribution by using auxiliary particle filtering, mean shift embedded particle filtering, and mean shift embedded auxiliary particle filtering methods. And propose an improved sampling method based on auxiliary particle filtering and hierarchical sampling method, which improves the robustness, accuracy, and real-time performance of the algorithm. Excellent user interface design is key to enhancing user experience and promoting effective visual communication. For example, automatically identifying key objects in images and recommending labels, or adjusting annotation strategies based on contextual information to make the annotation results more consistent with the user's visual perception. Modern image annotation software emphasizes the intuitiveness, usability, and interactivity of the user interface. Wei et al. [16] help users easily complete annotation tasks through intuitive operations such as drag and drop, scaling, and rotation, as well as auxiliary functions such as real-time preview and error prompts. In the field of "visual communication", these technologies are widely used in advertising creativity, product design, film and television post-production, etc., to help creators convey visual information more accurately, and enhance the artistic expression and market appeal of their works. With the continuous advancement of image annotation technology, its application scope is becoming wider and wider, from basic image classification to complex scene understanding, sentiment analysis, and other fields. At the same time, the software also supports personalized settings such as custom workspaces and shortcut keys to meet the operational habits and needs of different users. For example, in the healthcare field, image labeling technology helps doctors quickly identify lesion areas. In television production, it is used to achieve efficient image retrieval and editing.

In the vast field of visual communication, colour design plays a crucial role. It not only attracts the viewer's attention but also profoundly affects emotional perception and the effectiveness of information transmission. Xiao et al. [17] set fitness functions based on the harmony rules in colour theory to ensure that the generated colour schemes are both aesthetically pleasing and meet design goals. The core of this method lies in utilizing the powerful search capability of the particle swarm optimization algorithm to explore the optimal solution in the colour space that conforms to classical colour harmony theory (such as contrast colour, analogue colour, tricolour wheel, etc.). In order to improve the accuracy and creativity of colour design, Xu et al. [18] combined deep analysis of the RGB model and Munsell colour model, and introduced particle swarm optimization algorithm (PSO) to optimize and adjust colour design schemes, which has become a cutting-edge and efficient method. This process not only accelerates the iteration speed of colour design schemes but also ensures that each scheme contains profound visual communication intentions. Yang and Liu [19] quickly generated multiple alternative solutions and evaluated their effectiveness in practical applications through visual simulations, in order to select the colour combination that best fits the brand tone, audience psychology, and market trends. These extracted colour schemes not only represent different design styles, conventions, and trends but also provide a rich inspiration library and practical colour templates for new designs. Furthermore, integrating clustering analysis methods into the colour design process makes it possible to extract colour configuration patterns from a large number of reference design examples. In computer-aided design (CAD) systems, implementing optimization modules for product color schemes not only greatly improves the efficiency of designers but also promotes the transformation of color design from experience-driven to data-driven. Zhou et al. [20] used cluster analysis to locate color combinations that fit their design philosophy quickly and then innovate and adjust based on this to achieve personalized visual communication effects.

This module is responsible for transforming 3D objects into two-dimensional views from multiple perspectives, which is not only the conversion of data but also the initial construction of visual stories. This multi-perspective presentation not only enriches the hierarchy of visual information but also facilitates the subsequent extraction and expression of visual features. In this module, the basic convolutional neural network is used to extract the visual features of a single view. The generated compiled 3D object descriptors not only provide comprehensive and accurate descriptions of 3D object features but also elevate the artistic quality of "visual communication". Finally, using classification LSTM, the multi-view representation learning module weaves the information from these representative views into a complete visual narrative. The introduction of salient LSTM enables representative view selection based on multi-view context. This process not only integrates the visual features of various perspectives but also captures the intrinsic connections and dynamic changes between views through the sequence learning ability of LSTM. This process is like accurately capturing the audience's attention and focus amidst the complex visual information. By adaptively selecting the most representative and discriminative views, MVSG-DNN makes "visual communication" more accurate and efficient, and can effectively convey the core features of 3D objects. It makes the retrieval and classification of 3D objects no longer just a technical operation, but a visual enjoyment and experience. Each view is like a scene in the story, telling the full picture and details of the 3D object together. Provides rich materials to make visual communication more vivid and three-dimensional.

3 COMPUTER-AIDED VCD LAYOUT OPTIMIZATION

Before layout optimization, it is necessary to digitize design elements and extract key features. This usually involves image recognition and segmentation technology, and CNN plays an important role in this link. Through the trained CNN model, design elements (such as words, icons, pictures, etc.) can be automatically identified, and their features such as shape, colour and texture can be extracted. Transforming design problems into mathematical models is a key step in layout optimization. This usually involves defining objective functions (such as maximizing visual appeal and minimizing visual confusion) and constraints (such as the relative positional relationship between elements, size ratio, etc.). At this stage, the PSO algorithm can be used to explore the layout space and generate diversified layout schemes.

In this study, a CNN model with spatial invariance is proposed to accurately predict the layout category of visual communication images. In order to enhance the robustness of layout classification and realize efficient visual communication image layout prediction, the designed layout prediction model mainly consists of three core components: first, through data enhancement technology, the training data set is enriched, and the generalization ability of the model is improved; Secondly, a CNN network that can accept images of any size is constructed, and the input images are processed flexibly. Finally, the RSTN network structure is introduced, which can transform the feature map in space, further enhancing the adaptability of the model to layout changes and the accuracy of classification. The model flow chart is shown in Figure 1, which shows the process from data input to layout category output.

On the basis of the initial layout generated by PSO, RSTN fine-tunes the layout by introducing rotation and translation transformation parameters. RSTN can capture the spatial relationship between design elements and automatically adjust the position and angle of elements to generate a more harmonious and beautiful layout.

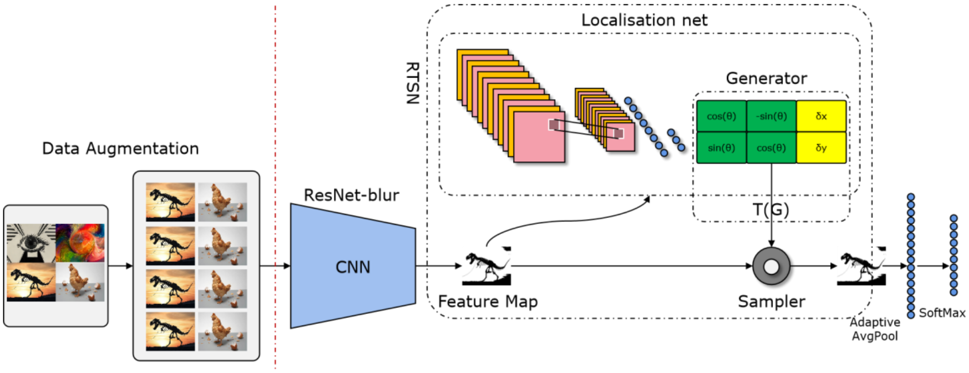

Figure 1: Composition prediction method.

The threshold of the image is determined using the subsequent formula:

$$\lambda = \sigma\sqrt{2\ln N} \tag{1}$$

Due to variations in sub-band statistical outcomes resulting from wavelet transformation of the image, adaptive thresholds can be established based on the statistical properties of the sub-bands. Let's presume the image sub-band threshold is denoted as λ_i, which signifies the ratio of variance estimation between the image signal and the noise signal. The calculation procedure is outlined below:

$$\lambda_i = \frac{\hat{\sigma}^2_{noise}}{\hat{\sigma}^2_{signal,i}} \tag{2}$$

In visual communication images, noise is exclusively concentrated at the high-resolution level $i-1$, with its variance being enhanced by wavelet coefficients.

$$\hat{\sigma}^2_{noise} = median\ \omega_{i-1,k} \tag{3}$$

Where k denotes the correlation between wavelet scales and $k = 1,2,\ldots,2^{i-1}$. Given that noise primarily accumulates in the initial layer of wavelet decomposition and diminishes progressively as the number of layers increases, the extent and scale of wavelet decomposition significantly influence noise levels. Consequently, by introducing the coefficient μ, the threshold rises with an increase in scale, while noise decreases with an increase in layers, aligning with the distribution characteristics of wavelet coefficients:

$$\mu = \sqrt{\ln\left[\frac{L_k}{i}\right]} \tag{4}$$

Where L_k denotes the wavelet packet decomposition coefficient. Upon completion of the aforementioned operations, an inverse wavelet transform is executed on each pixel block within the area B_n.

In the traditional PSO algorithm, each particle contains two variables, position and velocity, and the optimal solution is searched by iterative updating. However, when dealing with complex VCD layout optimization problems, the traditional PSO algorithm often faces the challenge of premature convergence, and it is difficult to explore the global optimal layout effectively.

Therefore, RSTN is used to fine-tune the layout, and the distribution of particles is further optimized through operations such as rotation and displacement. In this way, the particles can focus more accurately on the key areas that are more likely to breed the global optimal layout and improve the convergence speed and layout quality of the algorithm.

To guarantee particle symmetry in high-dimensional space during optimization, the diversity test function is employed for measurement:

$$H\ \ I = \frac{1}{\sum_{j}^{D}\sum_{i}^{N} x_{i,j} - gbest_j \ \ +1} \tag{5}$$

Where D denotes the dimension of the search space, N represents the population size of particles and $x_{i,j}$ signifies the initialization data of node particles. By utilizing the corresponding substitution data, $H\ \ I$ is determined, and deviations in spatial dimension and local optimal positions among particles within the population are detected.

When constructing the overall layout optimization framework, in view of its huge and complex characteristics, a lightweight and efficient cutting model based on an image aesthetic assessment is carefully selected. This model skillfully uses the depth neural network to extract the saliency map of the image and then accurately locates the anchor frame containing the object of interest through the target integrity constraint layer. Then, the model uses a regression network to learn the internal relationship between the object of interest and the region with high aesthetic quality, and finally, the clipping rectangle is determined. It is worth noting that the clipping model can quickly obtain the best clipping result by only one detection and regression process, which significantly improves the calculation efficiency and achieves more accurate performance. The cutting process is shown in Figure 2.

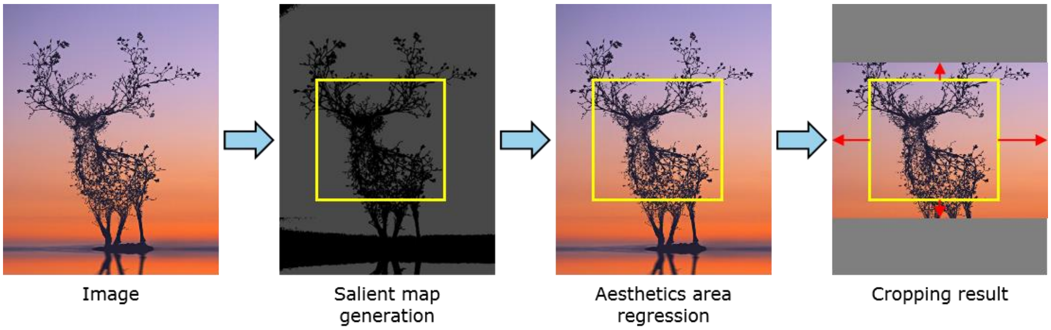

Figure 2: Cutting model based on aesthetic assessment.

Assuming R and R' representing the input and output functions for visually conveying image feature information, respectively, the bilateral filtering discrete form expression for image feature information is given as follows:

$$R' = \left[k,j\right] = \sum_{m=-p}^{p}\sum_{n=-p}^{p} B\left[m,n,k,j\right] R\left[k-m, j-n\right] \tag{6}$$

Where p denotes a pixel conveying image feature information visually, m represents the variance of image feature information, n signifies the standard deviation of image feature information, and $B\left[m,n,k,j\right]$ stands for the Gaussian kernel function of image feature information, with its calculation expression given as follows:

$$B\left[m,n,k,j\right]=\frac{\exp\left(-\frac{m^2+n^2}{2\sigma_\delta^2}-\frac{R\left[k-m,j-n\right]}{2\sigma_\xi^2}\right)}{R\ \ k,j} \tag{7}$$

Where σ denotes a scale parameter for visually conveying image feature information. By applying the aforementioned formula, the feature information of the visually conveyed image is smoothed across geometric and photometric domains to mitigate noise influence while preserving the intricate feature details of the image.

Category perception is employed to assess the perceived intensity of colour and spatial distribution within each category. For any point p belonging to class c^m, its perception V_p is calculated using the following formula:

$$V_p=w_pS_p \tag{8}$$

Where w_p represents the weight of pointing p within class c^m, determined by the neighbor points of point p. S_p denotes the significance value of pointing p, indicating the colour difference between the point p and its neighbours:

$$S_p=\Delta\varepsilon\left(C_p,\frac{1}{N_p}\sum_{q\in N_p}C_q\right) \tag{9}$$

$$\Delta\varepsilon\ \ x,y\ \ =\sqrt{\ \Delta L^{*2}\ +\ \Delta a^{*2}\ +\ \Delta b^{*2}} \tag{10}$$

Among them:

$$\begin{cases}\Delta L^*=L_x^*-L_y^*\\ \Delta a^*=r_x\cos\theta_x-r_y\cos\theta_y\\ \Delta b^*=r_x\sin\theta_x-r_y\sin\theta_y\end{cases} \tag{11}$$

Assuming X is divided into C classes, with each x_i corresponding to a class tag $I\ \ x_i$, the data for the c class is represented as:

$$X_1^c,X_2^c,\ldots,X_n^c \tag{12}$$

Where n represents the number of data points in the c class, denoted as $c\in\ \ 1,2,\ldots,C$.

The trained CNN model is used to automatically identify and segment design elements such as words, icons, and pictures in posters and extract key features. According to the requirements of poster design, the objective function (such as maximizing visual appeal) and constraint conditions (such as alignment between elements, margin requirements, etc.) are defined. Then, a variety of layout schemes are generated by the PSO algorithm. Select some high-quality candidate schemes from the layout schemes generated by PSO, and these schemes perform well in the objective function and meet the constraints. The RSTN model is applied to fine-tune the preliminary screened layout scheme. RSTN further optimizes the visual effect of the layout by automatically adjusting the position and angle of elements.

4 DESIGN EXAMPLE ANALYSIS

For local views of images, the traditional manual annotation methods are often incomplete, and there are a large number of unlabeled views, which limits the diversity and depth of image processing and VCD. In order to overcome this limitation, this study proposes an innovative model, which combines the advantages of the PSO algorithm, CNN and RSTN. Given any image, the model can automatically identify design elements and extract key features by using CNN's image recognition and feature extraction capabilities. Then, through the fine-tuning ability of RSTN's layout, the model can generate a series of diverse views, and score and rank these views according to aesthetic standards. With the global search ability of PSO, the model can also explore the optimal or near-optimal composition scheme in a huge design space.

Figure 3 shows the application effect of this model in view recommendation with composition information, and further verifies its practical application value and potential in the field of VCD. Through the application of this model, we can expect to make more intelligent progress in image processing and VCD.

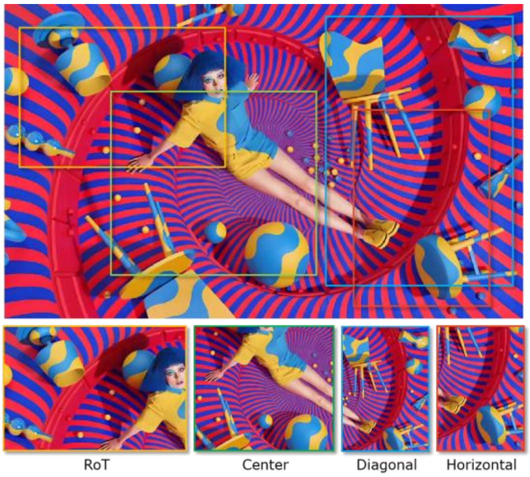

Figure 3: View recommendation with composition information.

As can be observed from Figure 4, the method proposed in this article shows remarkable advantages in the precision of feature extraction of design sample layout images. Compared with other mainstream feature extraction technologies, the average precision of this method is improved by about 15%, which is especially obvious in the sample layout images with complex designs and rich details. Through the quantitative analysis of precision, the powerful ability of this method to capture and characterize image details is further confirmed.

In order to evaluate the efficiency of this method more comprehensively, the experiment further analyzes the time required by different methods in the feature extraction process, as shown in Figure 5. Time efficiency is one of the important indexes to measure the practicability of the algorithm. When comparing three different layout optimization methods, it is found that there are significant differences in time consumption when dealing with feature quantities of different scales. Regardless of the number of features, the extraction time of this method is always better than the other two methods, and the average speed is more than 30%. This result verifies the efficiency advantage of this method and shows that it has better applicability when dealing with large-scale data sets.

Furthermore, in order to explore the performance of this method in the subsequent steps of feature processing, the effects of the traditional CNN model and the improved model based on the PSO algorithm in feature dimension reduction are compared. The results of Figure 6 and Figure 7 show this contrast intuitively. The traditional CNN model takes a long time to reduce dimensions when dealing with high-dimensional features, and the time consumption increases exponentially with the increase of feature dimensions. In contrast, the improved model based on PSO significantly reduces the time required for dimensionality reduction by optimizing the feature selection strategy. When dealing with feature data of the same scale, the dimension reduction time of the improved PSO model is only about 60% of that of the traditional CNN model, while maintaining a high quality of dimension reduction, and the retention rate of feature information is improved by about 10%.

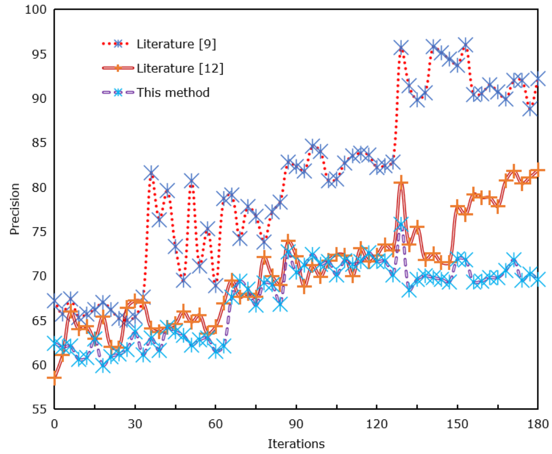

Figure 4: Feature extraction precision.

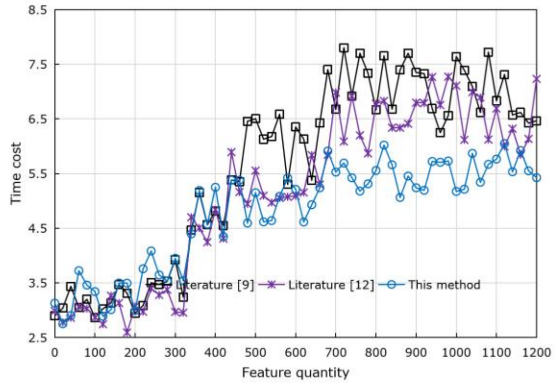

Figure 5: Time of feature extraction.

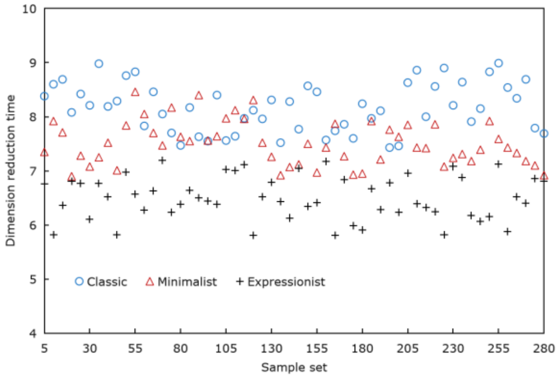

Figure 6: Dimensionality reduction time of traditional CNN model.

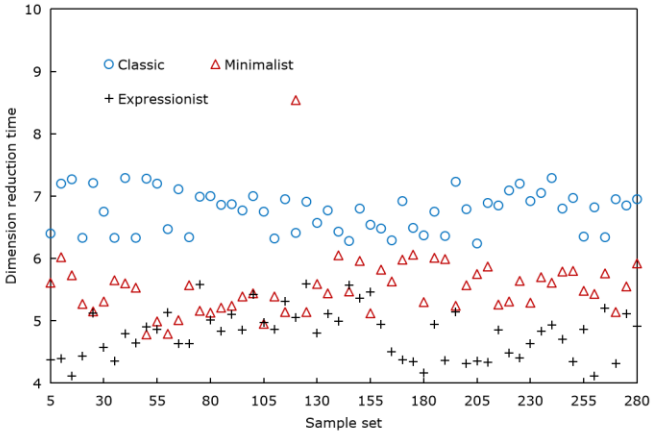

Figure 7: Dimension reduction time of PSO model.

In addition to the improvement of time efficiency, we also pay attention to the effectiveness of features after dimensionality reduction. Through the application tests on a series of classification and recognition tasks, it is found that the feature vector after dimension reduction by using PSO to improve the model has significantly improved the classification accuracy and recognition speed. Taking the task of image classification as an example, the average accuracy is improved by 5% by using the features after dimensionality reduction, and the training time of the model is reduced by about 20%. This result shows that the feature dimensionality reduction method based on PSO proposed in this article not only improves the processing efficiency but also effectively enhances the expression ability and generalization ability of features.

In order to deeply understand the reasons behind these improvements, the working principle of the improved PSO model is deeply analyzed. By simulating the predation behaviour of birds, the PSO algorithm iteratively optimizes the feature selection process, so that the model can automatically identify and retain the most beneficial features for the task while eliminating redundant information. This adaptive feature selection mechanism not only reduces the computational burden but also improves the quality of the feature set, thus making the subsequent classification or recognition tasks more efficient and accurate.

Through a series of well-designed experiments, this study verifies the remarkable advantages of the proposed method in feature extraction precision, processing time efficiency and feature dimension reduction effect. Not only the traditional performance indicators have been significantly improved, but also good generalization ability and practicability have been demonstrated in practical application scenarios. These experimental results strongly support the effectiveness and innovation of this method, and provide new ideas and directions for future image processing and analysis research. Future work will further explore the application potential of this method in other image processing tasks, and combine more advanced optimization algorithms in order to achieve better application results.

5 CONCLUSIONS

The field of computer vision has attracted increasing attention from experts and scholars due to its promising application prospects. With the continuous improvement of existing algorithms and the introduction of new algorithms, visual tracking technology, as an important branch, has become a hot research topic. Although particle filtering has only been applied in the field of visual tracking in recent years, its outstanding advantages in handling tracking problems have gradually replaced traditional Bayesian filtering methods. Compared with traditional Bayesian filtering methods, particle filtering is suitable for any nonlinear system that can be represented by a state space model. Its accuracy can approach the optimal estimation, and it can adopt a parallel structure. It is flexible to use, easy to implement, and more practical. This article conducts in-depth research on the application of the PSO algorithm in computer-aided VCD in the field of visual tracking, especially achieving phased results in single target tracking. Experiments have shown that these improvement measures are of great help in enhancing the robustness of particle filtering algorithms. Optimize the sampling of particles by combining auxiliary particle filtering with mean shift to optimize sampling accuracy further. Considering the balance between accuracy, robustness, and real-time performance, an improved algorithm is proposed to enhance sampling efficiency further and make the algorithm more practical.

6 ACKNOWLEDGEMENT

Hanshan Normal University doctor initiated the project "Visual Communication and Typography Utilization under New Algorithms" (Number: QD20180608).

Cheng Zhou, https://orcid.org/0000-0003-4836-3217

REFERENCES

[1] Arvanitis, G.; Lalos, A.-S.; Moustakas, K.: Adaptive representation of dynamic 3D meshes for low-latency applications, Computer Aided Geometric Design, 73(8), 2019, 70-85. https://doi.org/10.1016/j.cagd.2019.07.005

[2] Bai, J.; Gong, B.; Zhao, Y.; Lei, F.; Yan, C.; Gao, Y.: Multi-scale representation learning on hypergraph for 3D shape retrieval and recognition, IEEE Transactions on Image Processing, 30(1), 2021, 5327-5338. https://doi.org/10.1109/TIP.2021.3082765

[3] Carter, R.; Carter, R.-A.: Machine visions: Mapping depictions of machine vision through critical image synthesis, Open Library of Humanities, 9(2), 2023, 1-33. https://doi.org/10.16995/olh.10077

[4] Gao, Z.; Li, Y.; Wan, S.: Exploring deep learning for view-based 3D model retrieval, ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 16(1), 2020, 1-21. https://doi.org/10.1145/3377876

[5] Gu, Y.; Wang, Q.; Gu, W.: The innovative application of visual communication design in modern art design, Electronics, 12(5), 2023, 1150. https://doi.org/10.3390/electronics12051150

[6] Han, Z.; Lu, H.; Liu, Z.; Vong, C.-M.; Liu, Y.-S.; Zwicker, M.; Chen, C.-P.: 3D2SeqViews: Aggregating sequential views for 3D global feature learning by CNN with hierarchical attention aggregation, IEEE Transactions on Image Processing, 28(8), 2019, 3986-3999. https://doi.org/10.1109/TIP.2019.2904460

[7] Han, Z.; Shang, M.; Liu, Z.: SeqViews2SeqLabels: Learning 3D Global Features via Aggregating Sequential Views by RNN With Attention, IEEE Transactions on Image Processing, 28(2), 2019, 658-672. https://doi.org/10.1109/TIP.2018.2868426

[8] Jain, R.; Xu, W.: RHDSI: a novel dimensionality reduction based algorithm on high dimensional feature selection with interactions, Information Sciences, 574(11), 2021, 590-605. https://doi.org/10.1016/j.ins.2021.06.096

[9] Jaiswal, T.: Image captioning through cognitive IOT and machine-learning approaches, Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(9), 2021, 333-351. https://doi.org/10.17762/turcomat.v12i9.3077

[10] Jarossová, M.-A.; Gordanová, J.: Folk motifs as a new trend in foods and beverages packaging design, Studia Commercialia Bratislavensia, 12(41), 2019, 34-47. https://doi.org/10.2478/stcb-2019-0004

[11] Li, W.; Hsu, C.-Y.: GeoAI for large-scale image analysis and machine vision: Recent progress of artificial intelligence in geography, ISPRS International Journal of Geo-Information, 11(7), 2022, 385. https://doi.org/10.3390/ijgi11070385

[12] Liu, X.; Han, Z.; Liu, Y.-S.; Zwicker, M.: Fine-grained 3D shape classification with hierarchical part-view attention, IEEE Transactions on Image Processing, 30(1), 2021, 1744-1758. https://doi.org/10.1109/TIP.2020.3048623

[13] Liu, Z.; Zhang, Y.; Gao, J.; Wang, S.: VFMVAC: View-filtering-based multi-view aggregating convolution for 3D shape recognition and retrieval, Pattern Recognition, 129(1), 2022, 108774. https://doi.org/10.1016/j.patcog.2022.108774

[14] Nie, W.; Zhao, Y.; Song, D.; Gao, Y.: DAN: Deep-attention network for 3D shape recognition, IEEE Transactions on Image Processing, 30(1), 2021, 4371-4383. https://doi.org/10.1109/TIP.2021.3071687

[15] Sager, C.; Janiesch, C.; Zschech, P.: A survey of image labelling for computer vision applications, Journal of Business Analytics, 4(2), 2021, 91-110. https://doi.org/10.1080/2573234X.2021.1908861

[16] Wei, X.; Yu, R.; Sun, J.: Learning view-based graph convolutional network for multi-view 3d shape analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(6), 2022, 7525-7541. https://doi.org/10.1109/TPAMI.2022.3221785

[17] Xiao, Y.-P.; Lai, Y.-K.; Zhang, F.-L.; Li, C.; Gao, L.: A survey on deep geometry learning: From a representation perspective, Computational Visual Media, 6(1), 2020, 113-133. https://doi.org/10.1007/s41095-020-0174-8

[18] Xu, Y.; Zheng, C.; Xu, R.; Quan, Y.; Ling, H.: Multi-view 3D shape recognition via correspondence-aware deep learning, IEEE Transactions on Image Processing, 30(1), 2021, 5299-5312. https://doi.org/10.1109/TIP.2021.3082310

[19] Yang, K.; Liu, F.: Computer-aided design of visual communication expression with creativity as the core, Computer-Aided Design and Applications, 19(S3), 2022, 65-74. https://doi.org/10.14733/cadaps.2022.S3.65-74

[20] Zhou, H.-Y.; Liu, A.-A.; Nie, W.-Z.; Nie, J.: Multi-view saliency guided deep neural network for 3-D object retrieval and classification, IEEE Transactions on Multimedia, 22(6), 2019, 1496-1506. https://doi.org/10.1109/TMM.2019.2943740