







## Visual Element Extraction and Feature Matching Strategies Based on Convolutional Neural Network

Jinyuan Zhang<sup>1</sup> , Fulin Guan<sup>2</sup> , Zhuoyao Deng<sup>3</sup>  and Bijun Lei<sup>4</sup> 

<sup>1,3</sup> Guangzhou City Construction College, Guangzhou, Guangdong 510900, China,  
[1jinyuanzhenyu@163.com](mailto:1jinyuanzhenyu@163.com), [3zhuoyaodeng912@gmail.com](mailto:3zhuoyaodeng912@gmail.com)

<sup>2</sup> Guangdong Teachers College of Foreign Language and Arts, Guangzhou, Guangdong 510000, China,  
[guanfl@gtcfla.edu.cn](mailto:guanfl@gtcfla.edu.cn)

<sup>4</sup> Zhongkai University of Agriculture and Engineering, Guangzhou, Guangdong 510000, China,  
[leibijunn@163.com](mailto:leibijunn@163.com)

Corresponding author: Fulin Guan, [guanfl@gtcfla.edu.cn](mailto:guanfl@gtcfla.edu.cn)

**Abstract.** This article delves into the technology of extracting visual elements in computer-aided image design (CAID). When juxtaposed with the traditional Scale Invariant Feature Transform (SIFT) and Accelerated Robust Feature (SURF) algorithms, the proposed algorithm exhibits marked advantages in terms of matching points, accuracy, and stability. A comparison between the Deep Belief Network (DBN) model and the Convolutional Neural Network (CNN) model in visual element extraction reveals that the CNN model is superior in computing efficiency, real-time performance, and extraction quality. Consequently, an instructional system based on this algorithm is developed, with students invited for trial and feedback. The outcomes indicate that the system significantly boosts the efficiency and quality of image design. In conclusion, the visual element extraction algorithm presented in this article offers notable advantages and application potential in CAID, which is anticipated to introduce a novel approach to image design teaching.

**Keywords:** Computer Aided Image Design; Visual Element Extraction; Feature Matching; Deep Learning

**DOI:** <https://doi.org/10.14733/cadaps.2025.S4.296-309>

### 1 INTRODUCTION

Visual art is a type of art that can be viewed and appreciated. It involves various forms of art, including painting, architecture, and photography, and is also widely present in our daily lives. These methods mainly have two limitations. Firstly, the categories of visual art images are closely related to art history, and simply classifying based on image features ignores the influence of historical background on art development. To address the above issues, we will start with multidimensional correlation to classify and analyze visual art images. Considering the relationship between the evolution of art painting styles and the background of art history, we will explore the characteristics

of art history dimensions [1]. Existing work typically employs deep learning methods, combining global and local features, and extracting multi-level features from images to classify visual art images. Therefore, using computer methods to help non-art professionals understand and appreciate visual artworks has become increasingly important, and many domestic and foreign researchers have also focused on the classification and analysis of visual art images. Experiments have been conducted on multiple art painting datasets, demonstrating the superiority of this method over single-label classification methods [2]. In addition, some studies have proposed another widely applicable convolutional neural network framework for adaptive cross-layer correlation, which captures texture information of images from the visual dimension for the classification of various visual art images. With the continuous improvement of people's pursuit in spiritual and cultural aspects, more and more people are beginning to linger in art museums, galleries, art exhibition halls and other places. Identifying visual art images is different from general fine-grained image classification tasks, as their categories are not solely determined by salient objects in the image. Conventional image classification methods are not accurate enough to distinguish visual art images. Based on different factors, some scholars have designed corresponding knowledge extraction strategies to generate label distributions, provide art history supplementary information for input images, and train models on a multi-task learning framework [3]. And summarize and generalize three factors that influence the formation and development of artistic painting styles, including origin, origin time, and artistic movements. It focuses on the essence of graphic visual communication, not only examining the unique attributes of graphic information and its development from hand-drawn to digital but also delving into the integration of computer graphics with interdisciplinary fields such as cognitive psychology and semiotics [4]. On the basis of 3D reconstruction, artificial intelligence algorithms are used to automatically detect and classify paint defects/damages in images. Revealed how computer graphics play a central role in modern mobile media interface design as an efficient and economical information carrier. The multifaceted nature of the art field has brought many obstacles to the understanding of artistic images, and the lack of artistic understanding datasets has posed greater challenges to the development of artistic image understanding technology. At the same time, the progress of artificial intelligence in perceptual intelligence also demonstrates the potential of artificial intelligence in cognitive intelligence, leading people to explore the art of artificial intelligence. Therefore, some studies use artistic images as research objects to explore the ability of existing technologies to understand artistic images. We have built a technical framework for an art image understanding system based on the development achievements of image understanding. The main body of the system is the image description generation model. In describing the generation process, some scholars adopt an encoder-decoder structure to generate annotated text from images, taking into account the collection of image features and the embedding of text vectors. We introduced the underlying graphic text relationships in a general image dataset through transfer learning and adapted the content differently between different datasets. Pay special attention to the mapping process from image features to text content, and improve the accuracy of model mapping through a dual attention mechanism [5].

The practice has proven that the automated description generation technology proposed in this article can indeed enhance the ability to understand artistic images, but due to the lack of large-scale artistic image understanding datasets, there is still considerable potential for development. Based on the annotated dataset, establish a correlation analysis text content filtering model to conduct content analysis and filtering on the generalized art dataset [6]. Subsequently, the specific modes of applying transfer learning to the understanding of artistic images were analyzed, and detailed research was conducted on the specific issues of applying transfer learning to model training and data reconstruction processes. At present, compared to the more mature general image understanding tasks, research on art image understanding and annotation is still in its infancy [7]. This article investigates a key technology for generating art image descriptions based on transfer learning, aimed at automating the generation of highly readable content descriptions for art images. Designers can preview the design effect more intuitively and make real-time adjustments and optimizations, which not only shortens the product development cycle but also reduces the error rate and cost in the physical production stage [8]. In this process, designers will learn how to better utilize 3D technology

for rapid iteration and creative collision of visual elements, thereby promoting innovative development in the fashion design industry [9].

Given the technological advancements and broadening application fields, traditional teaching methods now face challenges in meeting the demands of modern design education. Key questions arise: how to effectively teach visual element extraction algorithms, foster students' practical skills and innovative thinking, and stimulate their learning interest and enthusiasm? Addressing these challenges is crucial in current CAID teaching. Thus, exploring new teaching strategies and methods grounded in research and practice of visual element extraction algorithms is vital for enhancing CAID teaching quality and nurturing high-quality design talents.

(1) This study focuses on the technical details of the visual element extraction algorithm, and closely combines it with the practice of CAID. Through an interdisciplinary research perspective, the application potential of the algorithm in design education is explored.

(2) In the research of the visual element extraction algorithm, this research deeply explores the deep learning algorithm. Compared with the traditional algorithm, the deep learning algorithm can automatically learn and extract high-level abstract features from the original image and has stronger robustness and generalization ability.

(3) This study puts forward a strategy of combining theoretical knowledge with practical operation. By designing a practical project based on a visual element extraction algorithm, students can deeply understand the principle of the algorithm in operation and improve their practical ability.

## 2 RELATED WORK

Humans can intuitively perceive and judge whether a painting is hung in a "face up" manner, even when facing abstract paintings, this ability is still effective. Specifically, focus on behavioural markers closely related to advanced visual cognition - directional judgment. This exploration not only deepens our understanding of the essence of art but also provides new perspectives and inspirations for visual element extraction and teaching strategies in computer-aided image design. Liu and Yang [10] introduced similar deep-learning models, allowing teachers to design a series of interactive and practical teaching activities. Traditional models often rely on "meaningful" content or specific image statistical features in painting to explain this phenomenon, which often aligns with existing artistic theoretical frameworks. To fill this gap, some scholars have innovatively introduced deep learning algorithms aimed at revealing and simulating the complex perceptual mechanisms behind human artistic creation and appreciation. Through the training of deep learning models, the capture and analysis of various painting styles and their inherent patterns have been achieved. At the level of teaching strategies, Ma et al. [11] suggested that computer-aided image design courses should focus more on cultivating and training students' visual perception abilities. In addition, as the level of abstraction in painting increases, deep learning models rely more on the extensive integration of spatial cues, which is of great significance for understanding human advanced visual cognitive processes. Through operation and practice, students gradually master how to extract key visual elements from complex images and understand how these elements interact to form an overall visual effect. At the same time, encourages students to explore different styles and fields, broaden their visual cognitive boundaries, and enhance their design innovation abilities. During the teaching process, teachers can use this discovery to guide students to gain a deeper understanding of how subtle changes in visual elements such as spatial layout and proportional relationships affect the overall presentation of visual effects, thereby cultivating students' spatial perception and composition abilities.

In the field of computer-aided image design, deep learning models have gradually taken the dominant position, demonstrating excellent performance in complex visual element extraction, image recognition, and classification. Traditionally, Murugesan et al. [12] used quantitative metrics such as accuracy, precision, and recall to evaluate model performance, but these metrics often only provide a superficial overview of performance, masking the fundamental reasons for differences in model performance and the qualitative complexity of its behavioural patterns. To explore the

perception and understanding ability of automation technology on artistic images and fill the gap in visual art image understanding, Wang [13] constructed a high-quality artistic image description dataset and proposed a transfer learning-based artistic image understanding technology method. Based on the migration adaptation strategy, the design of an art image understanding tool was completed, and it was trained and implemented on the MS COCO dataset and SemAnt dataset. Based on the annotation of the main sentences in high-quality datasets, a text content selection model is established through correlation analysis to determine the input information for the subsequent migration process of the dataset. Through horizontal comparison with relevant models in the field of image understanding, the effectiveness of the technology proposed in this paper in generating artistic image descriptions has been verified. Wang et al. [14] standardized the target description style in the art field, completed the construction of the description dataset, and manually annotated the description data. Xu et al. [15] introduced multiple quantitative indicators in the evaluation section to evaluate the tool for understanding artistic images. It defines the style of artistic image description, constructs a high-quality dataset for understanding artistic images, and performs data cleaning, analysis, and annotation on the dataset. Design and implement a text filtering model to automatically filter the main descriptive text of artistic images. By implementing a differential adaptation strategy to transfer low-level features from the general image domain to the artistic image domain, the training effectiveness of the artistic image description generation model can be effectively improved. We compared the current image understanding models' ability to understand visual art images and designed an automated description generation model for artistic images. Although traditional non-local mean image denoising algorithms can effectively suppress noise, they are often affected by information interference and loss of details. The experimental results show that the IPSO-BPNN model achieves a high accuracy of 98.64% in noise pixel recognition, with an F1 value of 96.32% and a model fitting degree of 0.983, fully demonstrating the superiority of the model in noise recognition and classification.

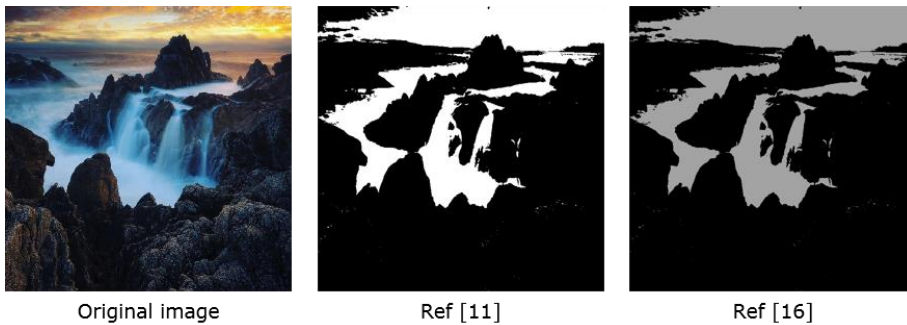
### 3 THEORETICAL BASIS

CAID constitutes a vital component of modern design technology, enabling designers to create and process images with heightened efficiency and precision through computer hardware and software support. Its applications span graphic design, 3D modelling, animation production, and more. CAID's core strength lies in providing robust tools for image editing, processing, and analysis, facilitating designers' creativity and enhancing the quality and efficiency of their work. In CAID practice, visual element extraction is a pivotal step. Elements like colour, shape, and texture form the fundamental units of an image and significantly influence the final outcome of a design. Image feature extraction stands as a pivotal technology in computer vision and image processing, focusing on deriving valuable information from the original image for further analysis and processing. In CAID, image feature extraction also plays an important role, which directly affects the effective use of visual elements.

There are many methods of image feature extraction, among which the method based on visual feature contrast has attracted much attention because of its intuition and effectiveness. This kind of method mainly depends on the characteristics of the human visual system and determines which regions or pixels have higher saliency by comparing the characteristic differences between different regions or pixels in the image, that is, they are more likely to contain important visual elements. Visual feature comparison methods can be divided into three categories: image block-based, region-based and global comparison.

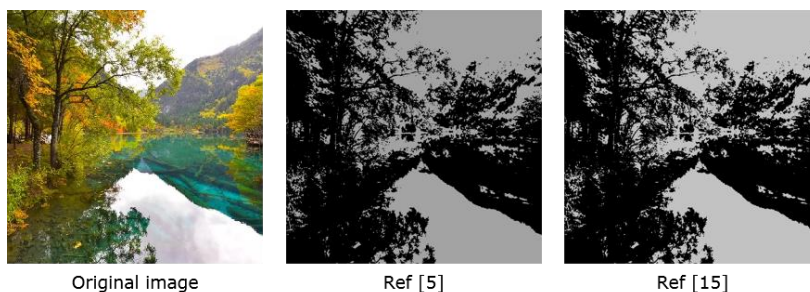
Saliency detection based on image blocks is to divide the image into several equal parts, regardless of the complex factors such as edge and direction in the image (as shown in Figure 1). When calculating the feature salient value of a pixel, this method usually takes a square area centred on the pixel as an image block, calculates the difference of colour or other features between the image block and other surrounding image blocks, and takes this difference as the feature salient value of the pixel. Compared with the saliency detection based on pixels, this kind of regional saliency detection algorithm based on image blocks can describe the target information more accurately and

effectively eliminate the interference of noise. However, its disadvantage is that it may blur the boundary of the target and ignore the obvious edges that should be highlighted, and the calculation process is relatively cumbersome.



**Figure 1:** Example of saliency map based on image block algorithm.

The region-based saliency detection algorithm first establishes image sub-regions through image segmentation, and each sub-region has a certain semantic structure (as shown in Figure 2). When calculating the saliency of a region, the advantages of giving priority to image segmentation to obtain image sub-regions compared with image blocks are as follows: first, the colour differences within sub-regions are small, while the colour differences between sub-regions are great, so when calculating the saliency of regions, only the feature differences between sub-regions can be considered, which is more in line with the block effect mechanism of the human visual system; Second, image segmentation can well preserve the contour characteristics of the target object, so as not to lose the edge characteristics, so that the obtained saliency map has a better semantic structure. After getting the segmented image sub-regions, calculating the feature saliency of the regions can be divided into two ways: one is to compare the features of the sub-regions as a whole, and the other is to calculate the feature contrast between regions in the form of the colour histogram. Region-based saliency detection improves the detection efficiency, retains the semantic structure of the image, and achieves good detection results. However, this kind of algorithm relies more on the effect of the segmentation algorithm. If the segmentation area is too large, it is easy to ignore the edge of the target. However, if the segmentation area is too small, it will lead to a poor global saliency map.

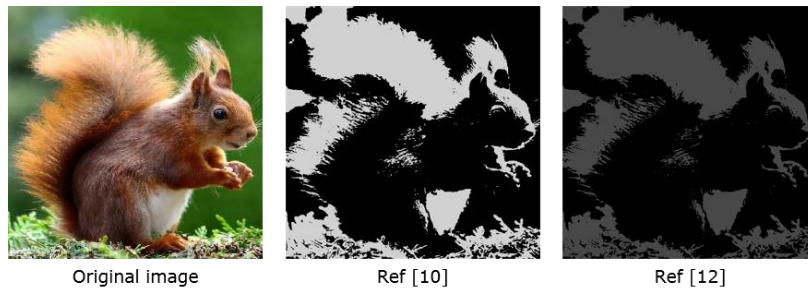


**Figure 2:** Example of saliency map based on region algorithm.

The above two methods are significance detection algorithms based on the local comparison, and they are both proposed according to the centre-periphery antagonism mechanism. Whether it is a multi-scale comparison or a regional comparison, its core is to calculate the difference between the current unit and other units around it.



In addition to the method based on local comparison, there is also a significance detection method based on global comparison (as shown in Figure 3). This method holds that when there is a big difference between the target and the background, the target has a high saliency. The background here is no longer limited to the area around the target but refers to the background of the whole scene. The spectral residual model is a typical example, which removes most of the insignificant areas through simple transformation and difference and regards the spectral residual area as the significant area. The frequency domain modulation model uses the difference between the colour of a pixel and the average colour of the whole image to represent the saliency of the pixel.



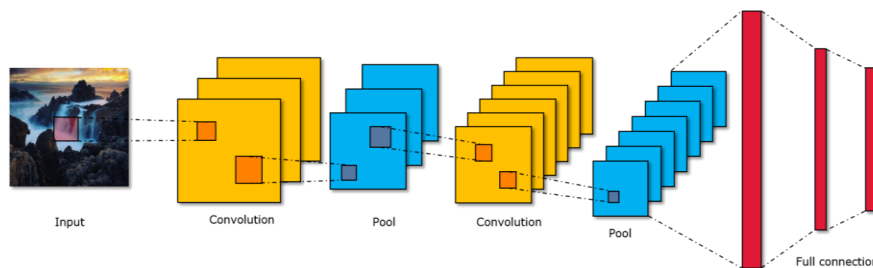
**Figure 3:** Example of saliency map of global comparison algorithm.

The saliency detection algorithm based on global comparison has the advantages of less computation, complete global structure, and easy implementation. However, due to its low adaptability and poor robustness, it is generally used to fuse with the first two methods based on local comparison to calculate saliency maps.

#### 4 COMPUTER-AIDED EXTRACTION OF VISUAL ELEMENTS FROM IMAGES

In CAID, efficient and precise design is rooted in the extraction of visual elements. Given the swift advancements in deep learning technology, CNN, as a prime example, has exhibited remarkable potential and advantages in visual element extraction. This section delves into CNN's application in this domain, exploring its working principle, advantages, and implementation strategies within real-world CAID projects.

CNN, a specialized deep feedforward neural network, is particularly adept at processing grid-structured data like images. It mimics the hierarchical structure of the human visual system, extracting image features layer by layer through convolution layers, pooling layers, and other structures. This progression from simple to complex ultimately enables the understanding and classification of image content. CNN's core strengths include automatic feature extraction, spatial hierarchy utilization, and efficient parallel computing. The CNN model structure discussed here is illustrated in Figure 4.



**Figure 4:** CNN model structure.

In the convolution layer, a number of learnable convolution kernels (filters) are used to operate the sliding window on the input image. The dot product of the convolution kernels and the pixel values in the corresponding region are calculated. The nonlinear characteristics are introduced through the activation function (such as ReLU) to generate the Feature Map. Each convolution kernel can be regarded as a feature extractor, which can capture a specific pattern or texture in the image. By stacking multiple convolution layers, CNN can extract features from low level to high levels, such as edges, corners, textures, shapes, and even complex object parts.

The pooling layer serves to down-sample the feature map, thereby decreasing data dimension and computation while preserving crucial feature information. This operation achieves dimension reduction by selecting either the maximum (maximum pooling) or average value (average pooling) within a specified region.

At the terminus of the CNN architecture, one or more fully connected layers are typically employed for classification or regression tasks. These layers transform the pooled feature map into a one-dimensional vector and generate the ultimate prediction outcome via linear transformation involving weight matrices and offset vectors, followed by nonlinear mapping through an activation function. In the task of visual element extraction, the fully connected layer can help identify specific objects or regions in the image and extract related visual elements.

The gradient represents the first derivative information within the image. Specifically, for any pixel  $I(x, y)$  in the image, the gradient information is determined as follows:

$$G_x(x, y) = I(x+1, y) - I(x-1, y) \quad (1)$$

$$G_y(x, y) = I(x, y+1) - I(x, y-1) \quad (2)$$

Here,  $G_x$  and  $G_y$  denote the horizontal and vertical gradients of  $I(x, y)$ , respectively.

Given that  $x_i, i=1, 2, \dots, n$  represents  $n$  pixels within the target area, and  $y_0$  denotes the target's centre coordinate, the target model can be described as follows:

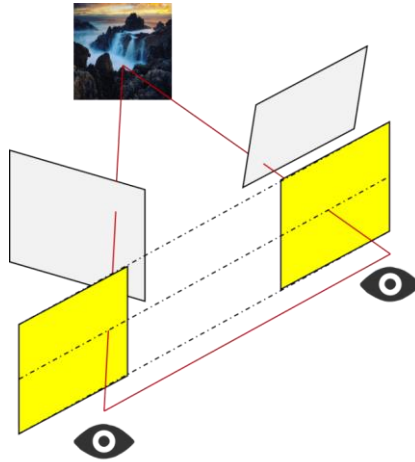
$$q_\mu = C_q \sum_{i=1}^n k \left( \left\| \frac{y_0 - x_i}{h} \right\|^2 \right) \delta[b(x_i) - \mu] \quad \mu = 1, 2, \dots, m \quad (3)$$

Here,  $k(x)$  denotes a monotonically decreasing kernel function, while  $h$  representing the bandwidth of the kernel function.  $b(x_i)$  signifies the index value of the corresponding colour or grayscale histogram at the pixel  $x_i$ .  $\delta(x)$  stands for the unit impulse function, and  $\delta[b(x_i) - \mu]$  is utilized to determine whether the index value at the pixel  $x_i$  belongs to the  $\mu$  eigenvalue of the histogram. Lastly,  $C_q$  denotes the normalization coefficient.

During the construction of the CAID system, the preferred binocular stereo vision setup utilizes a parallel camera configuration. This arrangement enables the simultaneous capture of left and right image pairs of the same object from distinct viewpoints in the real world. Consequently, the system can precisely determine the horizontal parallax of the object within the stereo-image pair. Leveraging the principle of similar triangles, also known as the similar triangle method, the system efficiently calculates the precise distance between the object and the camera.

The essence of binocular stereo-vision technology lies in replicating the working mechanism of the human visual system. Through the use of two cameras to gather scene image information, the system comprehensively captures scene details and features. Subsequently, by computing the disparity between the left and right image pairs, the system measures and ascertains the three-dimensional coordinate information of each scene pixel, thereby achieving precise spatial

perception and reconstruction. The binocular stereo vision model presented in this article is depicted in Figure 5.



**Figure 5:** Parallel binocular stereo vision imaging model.

Compared with the traditional visual element extraction method, CNN can automatically learn and extract features from the original image without manually designing a feature extractor. Through multi-layer convolution and pooling operation, CNN gradually extracts features from low-level to high-level images and simulates the hierarchical structure of the human visual system. This structure enables CNN to capture the spatial hierarchical information in the image and better understand the image content. Furthermore, convolution and pooling operations in CNN have natural parallelism and are suitable for efficient computing on hardware accelerators such as GPUs.

Train an SVM classifier,  $h$ , using the entire dataset to classify the unlabeled set  $D_{Unmarked}$ .

Subsequently, label the initial  $m$  images as  $D'_{relevant}, D'_{irrelevance}$ :

$$D_{Unmarked} = D_{Unmarked} - D'_{relevant} \cup D'_{irrelevance} \quad (4)$$

When faced with linear inseparability, mapping the problem to a higher-dimensional space for linear separability can be considered. However, transforming low-dimensional features to high-dimensional space and computing their inner product is computationally intensive or even infeasible. To address this, by ensuring  $x_i, x$  is the sole vector in the inner product, the optimal classification surface  $g(x)$  can be expressed as:

$$g(x) = \sum_{i=1}^n a_i y_i x_i, x + b \quad (5)$$

The auxiliary non-negative variable  $a_i$  is referred to as the Lagrange multiplier:

$$g(x) = \sum_{i=1}^n a_i y_i K(x_i, x) + b \quad (6)$$

The parameter  $a, y, b$  remains constant in the aforementioned formula. This implies that during the solution process, whenever an inner product is needed, the kernel function can be utilized for calculation  $a$ . Subsequently, the classifier can be derived by combining it with the kernel function.



Assuming a rectangle is employed to select the candidate target, consider the pixel  $x, y$  within the rectangular area of the current frame. The colour and motion information of the candidate target is described as follows:

$$M_m = \sum_x \sum_y I_m(x, y) \quad (7)$$

$$M_c = \sum_x \sum_y I_c(x, y) \quad (8)$$

$M_c, M_m$  denotes the zero moments of the probability distributions of significant colour and motion within the rectangular area, while  $I_c(x, y), I_m(x, y)$  representing the value of pixel  $x, y$  in these probability distributions.

A grayscale image comprises pixels with varying grayscale values, where the distribution of these values acts as a key feature for distinguishing between different images. To analyze this, compute the chi-square distance between the histogram  $S$  of the image to be recognized and the histogram  $M$  of the template:

$$S = s_1, \dots, s_n \quad (9)$$

$$M = m_1, \dots, m_n \quad (10)$$

Where  $n$  represents the histogram dimension. The smaller the chi-square distance, as indicated by the formula below, signifies greater similarity between the two images:

$$\chi^2_{S, M} = \sum_{i=1}^n \omega_i \frac{S_i - M_i}{S_i + M_i}^2 \quad (11)$$

Where  $\omega_i$  denotes the weight, and weights across different dimensions may vary.

The trained CNN model is used to extract the features of the input image, and the feature map containing rich visual information is obtained. Then, according to the specific requirements of the design task, select the appropriate feature representation and extraction strategy, and apply the extracted features to the subsequent design process. In the design process, user feedback and practical application effects are continuously collected, and the CNN model is iteratively optimized. By adjusting the model structure and training strategy, the performance of visual element extraction is continuously improved.

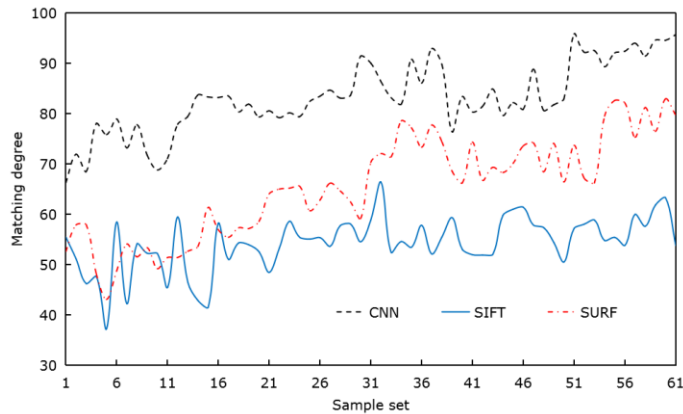
## 5 RESULT ANALYSIS AND DISCUSSION

### 5.1 Model Performance Simulation

To thoroughly evaluate the practical application effectiveness of the visual element extraction algorithm introduced in this paper within CAID and to contrast it with traditional algorithms and various deep learning models, a series of experiments are outlined in this section. Specifically, we commence by comparing the feature-matching accuracy of our proposed algorithm against the conventional SIFT and SURF algorithms. In the experiment, several representative images were selected as the test data set, including natural scenery, buildings, people, and other types. Through feature extraction and matching of these images, the feature-matching result, as shown in Figure 6, is obtained.

Figure 6 distinctly illustrates the disparities in feature-matching accuracy between the three algorithms. Notably, the algorithm presented in this paper surpasses the traditional SIFT and SURF algorithms in terms of matching points, accuracy, and stability. Specifically, the average number of matching points for this algorithm exceeds that of the SIFT algorithm by approximately 30% and the SURF algorithm by 25%. Furthermore, regarding matching accuracy, the algorithm introduced in this

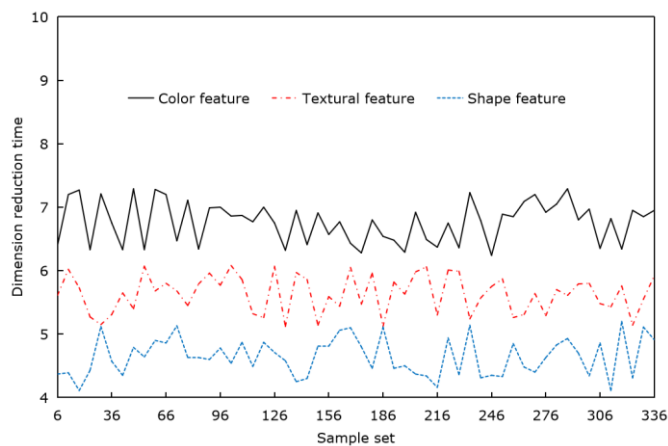
paper achieves over 95%, whereas the SIFT and SURF algorithms attain about 85% and 80% respectively.



**Figure 6:** Feature matching results of the algorithm.

In order to further verify the advantages of the deep learning model in visual element extraction, a comparative experiment was conducted between the DBN model and the CNN model. In the experiment, several groups of images are also selected as test data sets, and the dimension reduction time of the two models is tested.

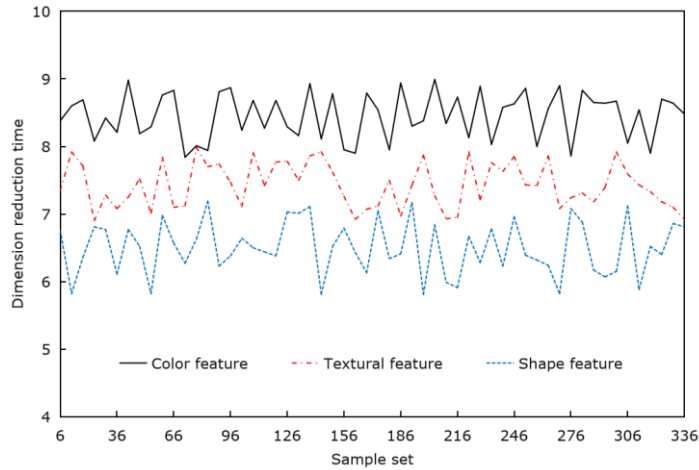
As shown in Figure 7 and Figure 8, the dimension reduction time test results of the DBN model and CNN model are respectively. Under the same test conditions, the dimension reduction time of the CNN model is obviously shorter than that of the DBN model. The average dimensionality reduction time of the CNN model is about 0.5 seconds, while that of the DBN model takes more than 2 seconds. This shows that the CNN model has higher computational efficiency and real-time performance in visual element extraction.



**Figure 7:** Dimension reduction time of the DBN model.

By comparing the visual elements extracted by the two models with the actual elements in the original image, it is found that the visual elements extracted by the CNN model are more accurate and

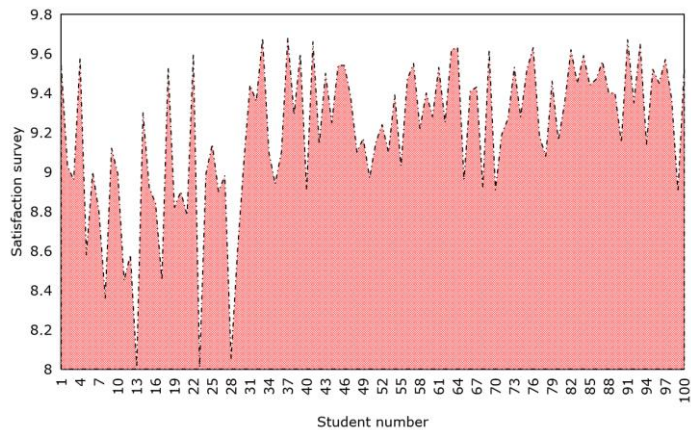
complete and can better preserve the details and texture information of the image. However, the DBN model has some extraction errors or omissions in some complex images.



**Figure 8:** Dimension reduction time of CNN model.

## 5.2 Investigation of Students' Satisfaction in Instructional System

In order to assess the application effect of the algorithm in practical teaching, an instructional system based on the algorithm is constructed, and a group of students are invited to try out and give feedback. After the trial, students' satisfaction was investigated, and the results are shown in Figure 9.



**Figure 9:** Student satisfaction survey.

The survey results show that most students are satisfied with the instructional system (the overall satisfaction has reached more than 9 points). They think that the system provides rich design resources and convenient design tools, which makes image design simpler and more efficient. Furthermore, the students also recognized the accuracy of the system in visual element extraction,

thinking that it could help them find the required visual elements more quickly and design and combine them effectively.

<i>Student ID</i>	<i>Completion Time with Traditional Method (minutes)</i>	<i>Completion Time with Proposed Method (minutes)</i>	<i>Time Saved (minutes)</i>	<i>Time Saved Percentage</i>
1	60	45	15	25%
2	48	35	13	27%
3	55	40	15	27%
4	50	36	14	28%
5	65	48	17	26%
6	52	38	14	27%
7	46	33	13	28%
8	58	42	16	28%
9	63	46	17	27%
10	57	41	16	28%
Average	54	39	15	28%

**Table 1:** Design task completion time.

<i>Student ID</i>	<i>Design Score with Traditional Method</i>	<i>Design Score with Proposed Method</i>	<i>Score Improvement</i>	<i>Score Improvement Percentage</i>
1	75	88	13	17%
2	68	85	17	25%
3	72	82	10	14%
4	70	84	14	20%
5	65	80	15	23%
6	74	87	13	18%
7	69	83	14	20%
8	71	86	15	21%
9	67	81	14	21%
10	73	85	12	16%
Average	72	85	13	18%

**Table 2:** Design task quality score.

As can be seen from Table 1, when all students use the instructional system based on the visual element extraction algorithm proposed in this article design, the average completion time is 15 minutes shorter than that using the traditional method, and the time-saving ratio reaches 28%. This is very important for students because they usually need to complete multiple design tasks in a limited time. As can be seen from Table 2, when students use the method proposed in this article to design the instructional system, the average design score is increased by 13 points compared with the traditional method, and the score increase ratio reaches 18%. This shows that this method not only improves the design efficiency but also significantly improves the design quality.

## 6 CONCLUSIONS

This study delves into the application of the visual element extraction algorithm within CAID and conducts a comprehensive evaluation of its performance through a series of experiments. When

compared to the traditional SIFT and SURF algorithms, the proposed algorithm exhibits notable advantages in feature matching accuracy, boasting higher matching points, accuracy, and stability. These findings strongly advocate for the use of visual element extraction in image design. Moreover, a comparison between the DBN and CNN models in visual element extraction reveals that the CNN model holds superiorities in computational efficiency and real-time performance, yielding more precise and complete visual elements.

## 7 ACKNOWLEDGEMENT

2023 Guangdong Provincial Department of Education Youth Innovative Talents Project for Colleges and Universities "Research on the Impact of the Rise of AI Design on the Upgrading and Transformation of the Jewelry Industry" (Project No.: 2023WQNCX149).

Jinyuan Zhang, <https://orcid.org/0009-0007-7222-4746>

Fulin Guan, <https://orcid.org/0009-0001-8618-5274>

Zhuoyao Deng, <https://orcid.org/0009-0004-5411-5126>

Bijun Lei, <https://orcid.org/0009-0005-7608-5156>

## REFERENCES

- [1] Benzon, H.-H.; Chen, X.; Belcher, L.; Castro, O.; Branner, K.; Smit, J.: An operational image-based digital twin for large-scale structures, *Applied Sciences*, 12(7), 2022, 3216. <https://doi.org/10.3390/app12073216>
- [2] Chen, T.; Yang, E.-K.; Lee, Y.: Development of virtual up-cycling fashion design based on 3-dimensional digital clothing technology, *The Research Journal of the Costume Culture*, 29(3), 2021, 374-387. <https://doi.org/10.29049/rjcc.2021.29.3.374>
- [3] Fan, M.; Li, Y.: The application of computer graphics processing in visual communication design, *Journal of Intelligent & Fuzzy Systems*, 39(4), 2020, 5183-5191. <https://doi.org/10.3233/JIFS-189003>
- [4] Habib, M.-A.; Alam, M.-S.: A comparative study of 3D virtual pattern and traditional pattern making, *Journal of Textile Science and Technology*, 10(01), 2024, 1-24. <https://doi.org/10.4236/jtst.2024.101001>
- [5] Kang, M.; Kim, S.: Fabrication of 3D printed garments using flat patterns and motifs, *International Journal of Clothing Science and Technology*, 31(5), 2019, 653-662. <https://doi.org/10.1108/IJCST-02-2019-0019>
- [6] Kang, Y.; Kim, S.: Three-dimensional garment pattern design using progressive mesh cutting algorithm, *International Journal of Clothing Science and Technology*, 31(3), 2019, 339-349. <https://doi.org/10.1108/IJCST-08-2018-0106>
- [7] Lelièvre, P.; Neri, P.: A deep-learning framework for human perception of abstract art composition, *Journal of Vision*, 21(5), 2021, 1-18. <https://doi.org/10.1167/jov.21.5.9>
- [8] Li, J.; Yang, J.; Hertzmann, A.; Zhang, J.; Xu, T.: Layoutgan: Synthesizing graphic layouts with vector-wireframe adversarial networks, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(7), 2020, 2388-2399. <https://doi.org/10.1109/TPAMI.2019.2963663>
- [9] Li, J.; Yang, J.; Zhang, J.; Liu, C.; Wang, C.; Xu, T.: Attribute-conditioned layout GAN for automatic graphic design, *IEEE Transactions on Visualization and Computer Graphics*, 27(10), 2020, 4039-4048. <https://doi.org/10.1109/TVCG.2020.2999335>
- [10] Liu, F.; Yang, K.: Exploration of the teaching mode of contemporary art computer-aided design centered on creativity, *Computer-Aided Design and Applications*, 19(S1), 2021, 105-116. <https://doi.org/10.14733/cadaps.2022.S1.105-116>
- [11] Ma, R.; Mei, H.; Guan, H.; Huang, W.; Zhang, F.; Xin, C.; Chen, W.: Ladv: Deep learning assisted authoring of dashboard visualizations from images and sketches, *IEEE Transactions on Visualization and Computer Graphics*, 27(9), 2020, 3717-3732. <https://doi.org/10.1109/TVCG.2020.2980227>

- [12] Murugesan, S.; Malik, S.; Du, F.; Koh, E.; Lai, T.-M.: Deepcompare: Visual and interactive comparison of deep learning model performance, *IEEE Computer Graphics and Applications*, 39(5), 2019, 47-59. <https://doi.org/10.1109/MCG.2019.2919033>
- [13] Wang, H.: Application of non-local mean image denoising algorithm based on machine learning technology in visual communication design, *Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology*, 45(6), 2023, 10213-10225. <https://doi.org/10.3233/JIFS-234632>
- [14] Wang, Q.; Chen, Z.; Wang, Y.; Qu, H.: A survey on ML4VIS: Applying machine learning advances to data visualization, *IEEE Transactions on Visualization and Computer Graphics*, 28(12), 2021, 5134-5153. <https://doi.org/10.1109/TVCG.2021.3106142>
- [15] Xu, P.; Hospedales, T.-M.; Yin, Q.; Song, Y.-Z.; Xiang, T.; Wang, L.: Deep learning for free-hand sketch: A survey, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1), 2022, 285-312. <https://doi.org/10.1109/TPAMI.2022.3148853>