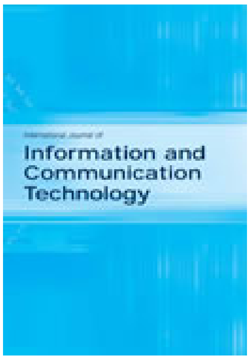
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Personalised cultural creative product design using user profile and personalised data diffusion model

Dongxu Yang, Yongxin Guo, Ronghui Liu

DOI: 10.1504/IJICT.2025.10075148

Article History:

Received:	22 September 2025
Last revised:	22 October 2025
Accepted:	25 October 2025
Published online:	12 January 2026

Personalised cultural creative product design using user profile and personalised data diffusion model

Dongxu Yang* and Yongxin Guo

School of Art and Design,
Henan University of Urban Construction,
Pingdingshan, 467000, China
Email: 30030901@huuc.edu.cn
Email: 15038818585@163.com
*Corresponding author

Ronghui Liu

School of Computer and Data Science,
Henan University of Urban Construction,
Pingdingshan, 467000, China
Email: liurh_126@126.com

Abstract: Addressing the issue that current cultural and creative product generation methods fail to account for user emotional needs, resulting in poor image generation outcomes, this paper first employs natural language processing algorithms to automatically segment user profiles and extract demand characteristics. Deep learning algorithms are introduced to analyse the sentiment behind user demands, thereby identifying emotional inclinations expressed by users. Building upon this foundation, a novel residual block architecture is designed with a diffusion model as the core network. The noise estimation network is enhanced by incorporating a convolutional block attention module. By integrating conditional control and user profiles as the control network, the approach effectively generates cultural and creative product images that align with users' emotional expectations. Experimental results demonstrate that the proposed method achieves at least an 8.63% improvement in peak signal-to-noise ratio, enabling the generation of high-quality cultural and creative product images.

Keywords: cultural and creative product generation; user profiling; conditional diffusion model; attention mechanism; natural language processing.

Reference to this paper should be made as follows: Yang, D., Guo, Y. and Liu, R. (2025) 'Personalised cultural creative product design using user profile and personalised data diffusion model', *Int. J. Information and Communication Technology*, Vol. 26, No. 49, pp.19–35.

Biographical notes: Dongxu Yang is a teacher in the School of Art and Design at Henan University of Urban Construction, China. He obtained his Bachelor's degree from the Wuhan University of Science and Technology (2005) and Master's degree from the Hubei University of Technology (2008), China. Currently, he is also studying in the Keimyung University. He has published over ten academic papers. His research interests are included product design and design marketing.

Yongxin Guo is a teacher in the School of Art and Design at Henan University of Urban Construction, China. He obtained his Bachelor's degree (2014) and Master's degree (2008) from the Xi'an University of Technology, China. He has published four papers. His research interests include innovative design of medical products and digital-intelligent cultural innovation design.

Ronghui Liu is a teacher in the School of Computer and Data Science at Henan University of Urban Construction, China. He obtained his Master's degree from the Zhongnan University of Economics and Law (2005), China, and obtained a PhD from the Donghua University (2012). He has published over ten academic papers. His research interests include machine learning and evolutionary computation.

1 Introduction

In the thriving cultural and creative industry, consumers' demand for cultural and creative products has shifted from merely functional satisfaction to personalised and emotional value pursuit. Traditional design of cultural and creative products largely relies on designers' experience, leading to issues such as serious homogenisation and low alignment with user needs, which makes it difficult to efficiently address the differentiated demands of a large number of users, thus constraining industrial innovation and development (Liu and Zhao, 2024). Diffusion models, as an emerging type of generative model, have achieved remarkable results in image generation (Chen et al., 2025). Compared with traditional generative adversarial networks (GANs), diffusion models can better meet the requirements for image quality and creativity in cultural and creative products (Dai et al., 2024). However, to achieve truly personalised generation of cultural and creative products, it is also necessary to fully consider users' individual needs and preferences and incorporate user information into the design process (Shi and Yang, 2025). User portraits, as a method that comprehensively and accurately describes and models user characteristics, can collect and analyse multidimensional data such as users' basic information, behavioural data, interests, and preferences to construct a virtual representation of the user (Li et al., 2021a). Through user portraits, we can deeply understand users' needs, preferences, and consumption habits and provide solid basis for designing personalised cultural and creative products.

Current research on methods for generating cultural and creative products is mainly focused on the image design of these products. Early image design methods for cultural and creative products primarily relied on manual rules or content-based generation (Xu and Zheng, 2022). However, their applicability was relatively narrow, and they had limited processing capabilities in complex scenarios and when dealing with poor image quality. In recent years, many bottleneck issues faced by traditional image processing techniques have been systematically resolved. Convolutional neural networks (CNNs), with their powerful feature expression capabilities, have gradually become the core solution for image design of cultural and creative products. CNN's convolutional layers automatically extract low-level features and high-level semantic features from images or text through local receptive fields and weight sharing mechanisms. Furthermore, by leveraging data-driven approaches, feature decoupling, and interactive optimisation, CNNs not only enhance the generation efficiency and diversity of cultural and creative products but also drive the deep integration of traditional culture with modern technology. Han et al. (2022) adopted a strategy of multi-scale feature fusion, using an encoder to extract global and mid-level features to jointly optimise pixel-level loss and image-level classification loss, generating semantically relevant product images. Zhang and Gao (2025) introduced channel attention modules to explicitly model dependencies between feature channels, achieving dynamic enhancement of key features and suppression of redundant information. Compared with traditional CNN methods, image generation techniques based on GANs can produce higher quality product images more accurately, especially in scenarios involving complex semantic structures; the generated results show a significant improvement in visual detail richness. GANs can generate cultural and creative products by integrating multimodal data such as text, images, and speech. In contrast, CNNs typically process each modality separately and achieve similar results through post-processing fusion. By learning from vast amounts of game assets, GANs automatically generate photorealistic 3D characters, saving developers significant design time. Conversely, CNNs rely on pre-trained models to extract fixed features, resulting in lower generation efficiency. Li et al. (2021b) further proposed a progressive GAN based on self-attention mechanisms by dynamically focusing on key regions of the product image during generation, while balancing generation quality and computational efficiency through multi-stage training strategies. Compared with the advantages of GANs in local detail generation, Transformer architecture provides new solutions for product image generation through its global context modelling capability. Liang et al. (2024) extract features from product images using CNN and generate images meeting user requirements via transformers, but the generated quality is not high. Liang (2024) designed a multimodal product image design algorithm combining Transformers and GANs, capable of generating high-quality and diverse images.

Compared with deep neural network models, diffusion models gradually denoise based on image noise distribution; this method can still maintain strong detail restoration and artefact suppression capabilities even in cases of high or complex noise. Chen and Carrillo (2011) proposed applying the denoising diffusion probabilistic model to product image generation tasks, generating high-quality product images through a gradual noise removal process. Yang et al. (2023) conducted the diffusion process in latent space, reducing computational costs, especially suitable for generating high-resolution image tasks. Li et al. (2024) further introduced conditional guided diffusion models (guided diffusion models), effectively controlling the generation process through conditional constraints, achieving high-quality image generation under specific conditions. Sun et al. (2020) incorporate CNN into conditional diffusion models to reverse-infer and reconstruct the original distribution of images, further enhancing noise modelling capabilities. Traditional cultural and creative product image design methods are often designed for vague user groups and struggle to meet personalised user needs. User portraits indirectly and accurately depict these implicit demands through behavioural data (such as browsing and collecting), enabling the model to understand and generate aesthetics that satisfy user semantics. Cao et al. (2023) analyse user portraits and emotional needs using a BERT model, integrating the results of emotional analysis into conditional generation modules to guide models in generating cultural and creative product images that better align with user emotions. Building on this, Wang et al. (2025) propose a unified framework for image-to-image translation based on conditional diffusion models by incorporating self-attention mechanisms, achieving high-fidelity product image generation.

In summary, existing methods for generating cultural and creative products have not sufficiently considered users' own emotional needs, leading to low-quality generated images. To address the aforementioned issues, this paper proposes a personalised cultural and creative product generation method that integrates user portraits with conditional diffusion models. First, this study uses the TextRank algorithm to extract keywords from user data and applies deep learning algorithms to extract demand features and perform sentiment analysis on relevant content identified through these keywords, thereby obtaining users' demands and emotional tendencies toward products. The extracted demand-and emotion-related positive and negative text contents are respectively clustered to summarise users' main expressed opinions, thus constructing a user portrait. Based on this, a cultural and creative product generation method based on an improved conditional diffusion model is designed. Achieve finer-grained spatial control by providing additional images composed of product images that meet the demands. To address issues with the backbone network, redesign the residual block structure to resolve the problem of model gradient disappearance and deepen the model's expressive power thus improving its performance. Add a convolutional block attention module (CBAM) attention module in the noise estimation network; through enhanced feature attention and spatial attention, enable the model to focus better on important image regions and capture relationships between different channels. At the same time, combine conditional control with user portraits as the control network by effectively fine-tuning to add locally spatial input conditions to the pre-trained diffusion model, thereby effectively generating product images that meet user needs. Experimental results show that the peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) of the proposed method are 28.39 and 0.942 respectively, which outperform comparison methods and can generate high-quality cultural and creative product images meeting user demands.

2 Relevant theory

2.1 User profile

The user role is the core starting point of cultural and creative product design. By building typical user profiles, designers can accurately identify the cultural needs, aesthetic preferences and functional demands of the target group. User roles serve as the compass for cultural and creative product design. Through demand insight, cultural translation, scene adaptation, and emotional resonance, they drive cultural creativity to transform from abstract concepts into concrete products. This user-centred design thinking can not only enhance the market competitiveness of products, but also promote the dynamic inheritance and innovative development of culture, achieving a dual increase in cultural and commercial value. There is a close relationship between user portraits and the design of cultural and creative products. As precise tools for extracting user needs, user portraits can provide systematic support to the cultural expression, functional design, market positioning, and communication strategies of cultural and creative products. Scholars define user portraits as fictional yet specific descriptions about target users; one user portrait can represent a group of individuals sharing similar behavioural characteristics (Miaskiewicz and Kozar, 2011). User portraits have the following advantages in the generation and design of cultural and creative products.

1 User portraits can make design goals more targeted. User portraits are a method that integrates large amounts of data and interview results from multiple users into a virtual representation, allowing designers to focus on one target user rather than a group of target users.

2 User portraits help prioritise design work based on the needs of target users. By defining the scope of target user needs, user portraits allow designers to consider only those user needs within this range during the design process, rather than focusing on what attributes and features the product should have (Moustafa, 2024). This allows designers to concentrate on determining functional priorities and limiting the set of functions during implementation, thus making products better meet user needs and improve user satisfaction.

3 User portraits can enhance designers' empathy toward users. Placing oneself in the position of the user helps designers better understand what users desire under specific circumstances and how they will use the product.

2.2 The basic principles of diffusion models

Diffusion models learn data distributions through a gradual denoising process, generating samples with rich details such as natural textures and colour transitions, and offering much greater diversity than GANs (Cao et al., 2024). Training GANs requires balancing the capabilities of the generator and discriminator, which often leads to issues like vanishing gradients or mode collapse, i.e., generators producing only a few types of samples. In contrast, diffusion models are trained based on a fixed noise prediction process without adversarial competition, offering higher stability and suitability for large-scale data training. Diffusion models are a class of generative models based on probability estimation, using conditional probability distributions to describe the relationship between noisy images and clean images, generating images by maximising conditional probabilities. The processing of diffusion models consists of two stages: forward diffusion and reverse denoising.

The forward diffusion process involves gradually adding Gaussian noise to an input image until it closely resembles a Gaussian noise distribution after multiple steps (Cachay et al., 2023). Given a data sample $x_0 \sim q(x)$, define a Markov process that incrementally adds noise to the target data, resulting in $(x_1, x_2, \ldots, x_T)$. After T additions, the final data distribution x_T becomes an independent Gaussian distribution.

$$q(x_t | x_{t-1}) = N\left(x_t; \sqrt{1-\beta_t} x_{t-1}, \beta_t I\right) \tag{1}$$

where $\{\beta_t \in (0,1)\}_{t=1}^{T}$ is a hyperparameter for variance, and the mean is jointly determined by β_t and x_{t-1}. Equation (1) has low iteration efficiency. To improve forward diffusion efficiency, introduce $\alpha_t = 1 - \beta_t$ and $\bar{\alpha} = \prod_{s=0}^{t} \alpha_s$, transforming equation (1) as follows.

$$q(x_t) = N\left(x_t; \sqrt{\bar{\alpha}_t}, x_0, (1-\bar{\alpha}_t) I\right) \tag{2}$$

Equation (2) can sample noisy data x_t at any time step t. Sample z from the standard distribution $N(0, I)$, and through parameter reparameterisation, obtain the noise-adding equation.

$$x_t = \sqrt{\alpha_t} x_{t-1} + \sqrt{1-\alpha_t} z_{t-1} \tag{3}$$

The result of the forward diffusion process is data that approximates a Gaussian distribution, while the reverse denoising process is the process of gradually denoising Gaussian noise to restore the original data. Given a noisy image x_t, construct a neural network to fit the reverse Markov process of $p_\theta(x_t - 1 | x_t)$, and gradually denoise until it restores into a clear image. The parameterised posterior distribution $p_\theta(x_t - 1 | x_t)$ is defined as follows.

$$p_\theta\left(x_t - 1 | x_t\right) = N\left(x_{t-1}; \mu_\theta\left(x_t, t\right), \sigma_t^2 I\right) \tag{4}$$

The mean μ_θ and variance σ_t^2 in equation (4) can be derived using the Bayesian equation, with the representations of the variance and mean as follows.

$$\begin{cases} \sigma_t^2 = \dfrac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t \\ \mu_\theta = \dfrac{1}{\sqrt{\alpha_t}}\left(x_t - \dfrac{1}{\sqrt{1-\bar{\alpha}_t}} z_\theta\left(x_t, t\right)\right) \end{cases} \tag{5}$$

where z_θ is Gaussian noise predicted by the neural network, and θ are neural network parameters. Through parameter reparameterisation, the final iterative formula for the reverse denoising process can be derived as follows.

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\alpha_t}} z_\theta\left(x_t, t\right)\right) + \sigma_t z \tag{6}$$

3 User persona development for cultural and creative products

3.1 User clustering based on the K-means algorithm

This article constructs user portraits by crawling user data from e-commerce platforms and automatically performs user portrait clustering and information extraction using natural language processing algorithms, analyses users' emotions, and can generate effective user portraits to facilitate the subsequent generation of personalised cultural and creative products that meet users' needs. It uses merchant backend data provided by e-commerce platforms to obtain users' browsing history, purchase records, reviews, etc. These data can directly reflect users' behavioural preferences and consumption habits. The collected data are cleaned to remove invalid, duplicate, and erroneous data, with sensitive information being anonymised to protect user privacy. At the same time, the data undergo standardisation processing to ensure consistency and comparability of data from different sources, after which the processed data is integrated into a unified dataset.

Each user is assigned an identification (ID) associated with all the keywords extracted from their posted content along with the corresponding word frequencies. The keywords and their frequencies represent users' behavioural preferences and intensity, serving as features for the user. In this article's keyword extraction process, it is based on the TextRank algorithm (Zhang et al., 2020) for extracting keywords. During the user clustering phase, first, pre-processing of the user data needs to be performed. A bag-of-words model is used to represent each user, transforming each user into a vector. The bag-of-words model refers to combining all non-repeating input keywords as a single vocabulary, with the total number of words in the bag representing the dimensionality of the vector. If a word appears in the user's keywords, its position becomes the count; if it does not appear, it is recorded as 0. Ultimately, a vector expressing the user's information is obtained.

After representing users as vectors, each user can be viewed as a point in the vector space with different positions. The distance between different users in the vector space becomes measurable. Two points that are closer together have higher semantic similarity in text and thus represent more similar users. User clustering is performed using the K-means algorithm (Lai et al., 2025). The steps of the algorithm are as follows.

1 Specify the number of clusters *n*, automatically generate *n* category centroids based on the specified number $k_1, k_2, \ldots, k_n$, and initialise the sample set to $C_t = \varnothing$, $t = 1, 2, \ldots, n$.

2 Sequentially calculate the distance between each point in the vector space and the centroid. The equation is as follows. Points with the smallest distances are assigned to their corresponding centroids' categories $C_{\lambda_i} = C_{\lambda_i} \cup \{x_i\}$.

$$d_{i,j} = \|x_i - k_i\|_2^2 \tag{7}$$

3 Based on the clustering results, re-generate cluster centres using the following equation. Repeat step (2) and step (3) until convergence.

$$k_j = \frac{1}{|C_j|} \sum\nolimits_{x \in C_j} x \tag{8}$$

3.2 User requirement feature extraction based on natural language processing technology

After users are clustered into groups, feature extraction must be performed for each category of user data obtained from clustering. Feature extraction is a research topic in the field of natural language processing. Sentences expressing similar content have similar features. Building models that identify sentence features allows computers to understand sentences with specific characteristics and helps process natural language more broadly. When users discuss cultural and creative products, they express their needs and expectations for such products. The article uses a support vector machine (SVM) model for classification training and constructs models suitable for demand extraction to perform data analysis. During the training phase, labelled texts of demands and non-demands are used as a training set input into the classification model for training. Once the recognition accuracy exceeds 90%, it is considered capable of identifying text features that express demands. Subsequently, other texts requiring identification can be input into the classification model to determine whether the current text belongs to demands or non-demands.

Firstly, perform word segmentation on the text and use term frequency-inverse document frequency (TF-IDF) (Wang, 2024) to convert it into a text vector as shown below. Here $tf_{i,j}$ is the frequency of this keyword across all texts and $idf_{i,j}$ represents inverse document frequency.

$$tfidf_{i,j} = tf_{i,j} \times idf_{i,j} \tag{9}$$

Subsequently, sentence vectors are input into a pre-trained SVM model. SVM can find the most suitable separation hyperplane in feature space to maximise the margin between positive and negative samples on the training set; after predicting classification with SVM, it is possible to determine whether a sentence expresses user demands, ultimately extracting sentences that express users' needs.

3.3 User sentiment analysis based on deep learning

Sentiment analysis can help understand users' emotional tendencies and identify existing problems with cultural and creative products, which can assist product providers in implementing targeted improvements to optimise service experiences. The purpose of performing sentiment analysis on the relevant content about cultural and creative products published by users in this paper is mainly to understand the attitude preferences that each category of people has towards such products.

When dealing with text sentiment analysis problems, the task is converted into a classification problem of determining whether a text expresses positive or negative sentiment, and a TextCNN classification model is used for category prediction. After performing sentiment analysis on the texts, they are classified into content expressing positive sentiments and content expressing negative sentiments, corresponding to parts where users are satisfied with cultural and creative products and parts where they are dissatisfied. Following this, main user viewpoints are summarised. Under each individual characteristic keyword of a group of users, there is a large amount of data that expresses various contents, including many texts expressing demand features and sentiment tendencies. Some of these data express similar viewpoints while others express different viewpoints. Simply listing information cannot allow readers to quickly understand the expressed information, so it is necessary to cluster large volumes of text content and summarise the main viewpoints conveyed within them.

In this paper's text clustering, user clusters are first filtered based on keywords ranked in the top 30 by frequency. Then, corresponding original social data texts from users are indexed from an existing database using these keywords. For all sentences, they are converted into vector form, and their similarity is calculated through cosine values as shown below.

$$\cos(\theta) = \frac{\sum_{i=1}^{n}(x_i \times y_i)}{\sqrt{\sum_{i=1}^{n}(x_i)^2} \times \sqrt{\sum_{i=1}^{n}(y_i)^2}} \tag{10}$$

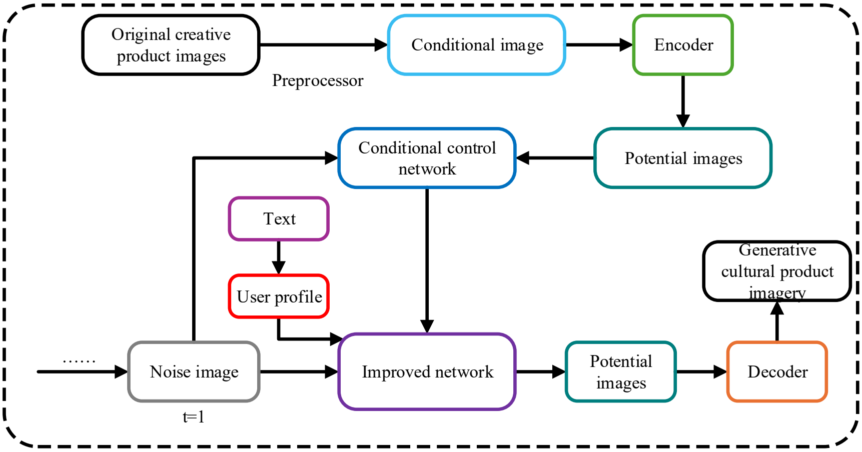
All the similarity values form a similarity matrix, and K-means algorithm is used to perform clustering analysis on it by continuously iterating calculations of matrix attraction and membership degrees to generate clustering results. Through text clustering, the overall viewpoints expressed by texts corresponding to keywords can be obtained, helping understand users' actual lifestyle patterns. For extracted demand texts and sentiment positive/negative texts, the same algorithm is applied for clustering analysis to obtain users' main product demands and positive and negative sentiment viewpoints.

4 Generation of personalised cultural and creative products based on user profiles and conditional diffusion model

4.1 Overview of personalised cultural and creative product generation models

The core competitiveness of cultural and creative products lies in their unique cultural connotations and personalised emotional expression. Traditional e-commerce platform recommendation mechanisms can only match existing products, making it difficult to meet users' deep-level, unmet individualistic creation needs. However, the current generation methods based on conditional diffusion models have shortcomings in controlling image spatial composition; they cannot precisely express complex image layouts solely through textual prompts. Generating an image that accurately matches subjective intentions usually requires numerous cycles of trial and error, significantly increasing operation time. The backbone UNet (Liu et al., 2025) model of diffusion-based models gradually reduces the size of feature maps through stacked convolutional and pooling operations, resulting in relatively low model efficiency. It cannot maintain high accuracy when training with greater depth, nor can it effectively identify meaningful channel information and positional information from feature maps.

Figure 1 The proposed personalised cultural and creative product generation method (see online version for colours)

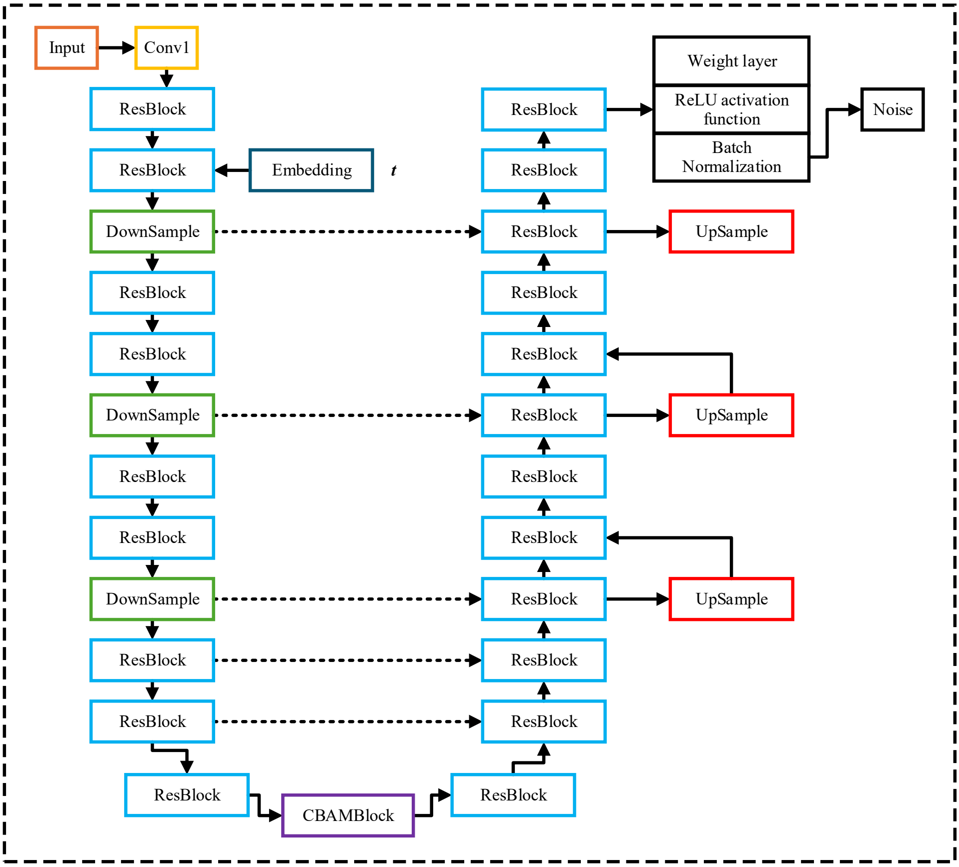


To address these issues, this paper proposes a personalised cultural and creative product generation method based on user profiling and an improved conditional diffusion model. The proposed personalised cultural and creative product generation method is shown in Figure 1. Finer-grained spatial control is achieved by providing additional images that specify the required image composition. Regarding problems in the backbone network, we redesign the residual block structure to solve the issue of vanishing gradients, deepen model expression capabilities thus enhancing performance. A CBAM attention module is added to the noise estimation network; it enhances feature and spatial attention so that the model can better focus on important regions in cultural and creative product images as well as capture relationships between different channels. At the same time, by combining with a conditional control network and effectively fine-tuning local input conditions into pre-trained diffusion models, we efficiently generate images specific to user profiles.

4.2 Improved noise estimation network

During the training process of conditional diffusion models, issues such as vanishing or exploding gradients often arise. To effectively avoid these problems, the model typically introduces ResBlock (Luo et al., 2022) as a 'bridge' for network transmission; it is precisely the core component in building ResNet. In the process of importing and processing training data, by accurately fitting the data and applying reasonable regularisation methods, the model can demonstrate strong generalisation capabilities on unseen data.

Figure 2 The improved noise estimation network structure (see online version for colours)



The skip connection in traditional ResBlock blocks uses an identity mapping approach, i.e., directly adding the input to the output. This design often leads to internal covariate shift problems within the ResBlock block, where the input distribution for each layer changes during training. To address this, this paper designs a new residual block. It incorporates batch normalisation (BN) into the ResBlock block, normalising the input data at each network layer so that the distributions have consistent mean and variance. This allows subsequent layers to no longer need to adapt to input variations in lower-level networks, achieving decoupling between layers. As a result, the distribution of input data for each layer becomes more stable, thus increasing learning speed. The model with BN added converges more easily and exhibits stronger generalisation capabilities.

The new residual block retains the original convolutional layer to extract features, maintain spatial structure, and reduce the number of model parameters. The ReLU function is chosen as the activation function for the model; it is a piecewise linear function $F(x) = \max(0, x)$, with a relatively simple structure and very efficient computation, which can accelerate training to some extent and is easy to insert into the model. In the improved residual block during backpropagation, gradients can be fully propagated backward without vanishing gradient issues.

To improve the performance of the noise estimation network and enhance the quality of generated images, a lightweight convolutional attention module CBAM is introduced into the noise estimation network. It is an attention module for CNNs that enhances the model's ability to model image features. Adding the CBAM attention module in the middle layer of the network considers both channel attention and spatial attention, selectively adjusting feature responses across different channels and spatial positions to improve the performance of the model.

The core components of CBAM include two modules: channel attention module and spatial attention module. The channel attention module adjusts the weights of channel features by learning correlations between channels, while the spatial attention module adjusts the weights of spatial features by learning correlations between different spatial positions in feature maps. The combination of these two modules enables the CBAM attention mechanism to focus more comprehensively on important information in images. Assuming $F \in \{R^{1 \times H \times W}\}$ as input and $M_C \in \{R^{c \times 1 \times 1}\}$ is one-dimensional convolution for channel attention module, the output of channel weight data is $c \times 1 \times 1$, then there are the following formulas.

$$F' = M_C(F) \otimes F \tag{11}$$

$$F'' = M_s(F') \otimes F' \tag{12}$$

where F' is the channel attention output, $F \in \{R^{1 \times H \times W}\}$ is two-dimensional convolution in spatial attention module. The intersection operation between channel attention output and spatial attention result is performed to obtain the final output of F''. As a lightweight module, CBAM can be directly embedded into network structures without additional parameters or computations, which can effectively improve model performance. The improved noise estimation network structure is shown in Figure 2.

4.3 Generation of cultural and creative products controlled by user profile conditions

In practical cultural and creative product generation, starting solely from text prompts makes it difficult to accurately express complex spatial layouts or shapes; therefore, generating an image that meets users' emotional needs requires numerous attempts. To address this issue, the ControlNet model (Zhao et al., 2023) is added on the improved network structure, locking parameters of pre-trained noise prediction networks and cloning them into a trainable copy in the control network. Introducing user profile condition information onto the already locked network achieves optimisation of the pre-trained network.

To prevent harmful noise from being added to the deep features of diffusion models and to avoid damaging the trainable copy by noise during training, the trainable copy and the locked model are connected using a zero convolution layer with weights initialised to zero that can continuously increase during training. Suppose *x* and *y* are two-dimensional feature maps, $x \in R^{h \times w \times c}$, where *h* is height, *w* is width, *c* is channel number. Adding ControlNet into pre-trained neural network blocks without changing the parameters in the original neural block *θ*, function *f* represents operations of the neural network block, for a residual block, the formula is as follows.

$$y = f(x; \theta) \tag{13}$$

Function *z* is a 1 × 1 convolution layer with weights and biases initialised to zero. Constructing a convolution layer using θ_{z_1} and θ_{z_2} as parameters, θ_c represents control parameters based on user needs, leading to equation (14), where + indicates feature addition.

$$y_c = f(x; \theta) + z\left(f\left(x + z\left(c; \theta_{z_1}\right); \theta_c\right); \theta_{z_2}\right) \tag{14}$$

In initial training, since the weights and bias parameters of the zero convolution layer are initialised to zero, therefore $y_c = y$. At the same time, due to this setting, there is no impact in forward propagation; harmful noise does not affect neural network layers. Moreover, because $z(f(x + z(c; \theta_{z_1}); \theta_c); \theta_{z_2})$ equals zero, the added neural network can accept user profiles as input conditions.

$$\frac{\partial y}{\partial \omega} = x, \frac{\partial y}{\partial x} = \omega, \frac{\partial y}{\partial b} = 1 \tag{15}$$

After backpropagation, the zero convolution layer within the model gradually changes into optimised values through learning processes, becomes non-zero and influences outputs. Suppose the zero convolution layer is $y = wx + b$, where *w* and *b* are weights and bias respectively, *x* is the input feature. The gradients for each term are as follows.

$$\frac{\partial y}{\partial \omega} = x, \frac{\partial y}{\partial x} = \omega, \frac{\partial y}{\partial b} = 1 \tag{16}$$

Since model parameters are locked, there is no need for training, which accelerates training speed without affecting the model, and each optimisation improves model performance.

5 Experimental results and performance analysis

5.1 User profile construction results

The experiment uses the CultureCreative-10K dataset collected from a large domestic cultural and creative e-commerce platform as referenced in Huang (2024). This dataset includes anonymised behavioural data of 10,000 active users, high-resolution images (512×512) for over 50,000 cultural and creative products along with their text descriptions, category tags, and 150,000 user-item interaction records. The dataset is divided into training set, test set, and validation set in a ratio of 6:3:1. The model's learning rate is set to 0.0003, Gaussian noise with standard deviation 0.35 is added, and the diffusion step length is selected as $t = 500$. To ensure optimal model performance, the batch sample size is chosen as 8. Experiments are conducted using the deep learning framework PyTorch (1.7.0), with CUDA version 11 and Python version 3.8.5.0.

By analysing the dataset from the cultural and creative e-commerce platform, user behaviour analysis results were obtained, as shown in Figure 3. It can be observed that user activity remains relatively stable during weeks 1–7, but browsing volume and add-to-cart quantity show a significant increase in weeks 8–9. This might be attributed to marketing activities by cultural and creative product shops. User purchase quantities tend to rise in subsequent weeks. We quantify such indicators based on the concept of user profile tags into user behaviour data to assist the following personalised cultural and creative product generation model in achieving better results.

In addition, this paper conducts a visual analysis of the user profile clustering results. t-distributed stochastic neighbour embedding (t-SNE) is a commonly used visualisation method that primarily maps similarity relationships in high-dimensional data to low-dimensional space through dimensionality reduction, thereby obtaining the main components of the selected feature layers and revealing key embedded user demand characteristics from the original user profiles. The visual analysis results are shown in Figure 4, where L0, L1, L2, L3, L4, and L5 represent functional demands, emotional demands, cultural demands, social demands, investment demands, and environmental protection demands respectively. Figure 4(a) shows that features of original user data are highly chaotic and difficult to distinguish effectively. However, after K-means clustering processing, the features of different types of user demand show obvious separation, with each type of user demand feature clustering well, which confirms that the proposed method has good feature learning capability and can promote better classification of user demands, thus building more accurate profiles.

Figure 3 User behaviour analysis results (see online version for colours)

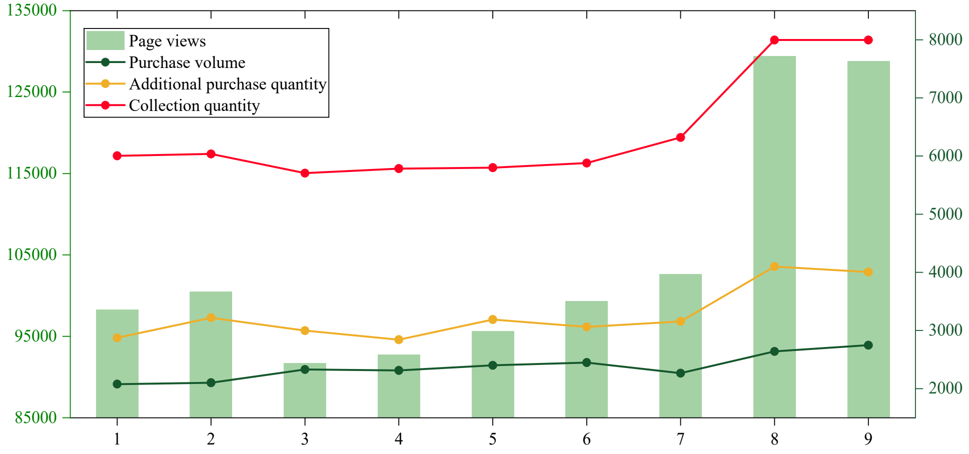
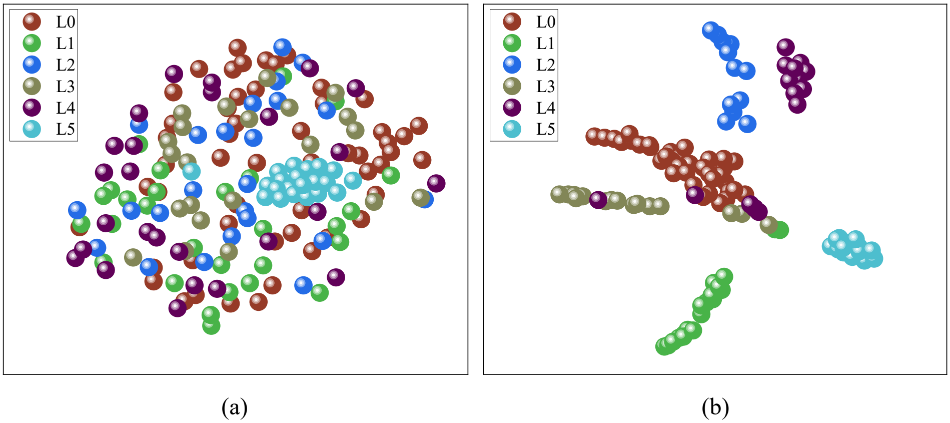


Figure 4 Visualisation of user demand characteristics, (a) distribution of original user demand data, (b) distribution of clustered user demand data (see online version for colours)



5.2 *Analysis of cultural creative product image generation results*

To verify the effectiveness of the proposed UPCDM method, five representative generative algorithms SAMGAN (Li et al., 2021b), CTRANS (Liang et al., 2024), CNNDIF (Sun et al., 2020), BCDIF (Cao et al., 2023), and SADIF (Wang et al., 2025) are selected for quantitative comparison on the experimental dataset. Quantitative metrics include PSNR, SSIM, learned perceptual image patch similarity (LPIPS), and Fréchet inception distance (FID). A higher PSNR value indicates better pixel-level matching and approaches infinity when two images are identical. Compared to PSNR, SSIM aligns more closely with subjective evaluation trends; a higher SSIM value implies better performance in generating cultural creative product images. Lower values of LPIPS and FID indicate superior perceptual quality. The image generation results comparison of different methods is shown in Table 1. UPCM achieves PSNR and SSIM values of 28.39 and 0.942, representing increases of at least 8.63% and 3.74%, respectively, compared with SAMGAN, CTRANS, CNNDIF, BCDIF, and SADI. Compared to the baseline methods, UPCM reduces LPIPS and FID by at least 15.89% and 8.44%, respectively. The method incorporates a novel residual module design and adds a lightweight CBAM attention module within the noise estimation network, which enhances image generation capability and performance. Moreover, combining the ControlNet architecture with user profiling requirements effectively accomplishes the task of generating images for cultural creative products based on specific needs.

This paper also conducts ablation experiments on each component of UPCDM, replacing the improved ResBlock block with the original ResBlock by removing it and denoted as UPCDM/M1. Removing the CBAM module is denoted as UPCDM/M2. Removing the user portrait conditional control module is denoted as UPCDM/M3. The ablation experiment results of different components are shown in Table 2. The generation method without the user portrait conditional control module performs worst on all indicators, indicating that this module has a key impact on the final generation effect. Although

the performance of removing the CBAM module and removing the improved ResBlock block is higher than UPCDM/M3 on various indicators, it still significantly underperforms compared to the complete method UPCDM. Based on the above analysis, each component in UPCDM plays a decisive role in generating cultural product images. UPCDM with all integrated components can generate high-quality cultural product images.

Table 1 The image generation results comparison

Method	*PSNR*↑	*SSIM*↑	*LPIPS*↓	*FID*↓
SAMGAN	16.19	0.814	0.262	5.09
CTRANS	19.02	0.842	0.239	4.85
CNNDIF	20.97	0.851	0.218	4.43
BCDIF	22.05	0.885	0.174	4.25
SADIF	26.41	0.908	0.151	3.91
UPCDM	28.39	0.942	0.127	3.58

Table 2 Experimental results of the dissolution of each component in the proposed method

Method	*PSNR*↑	*SSIM*↑	*LPIPS*↓	*FID*↓
UPCDM/M1	26.59	0.920	0.151	3.92
UPCDM/M2	24.31	0.893	0.176	4.16
UPCDM/M3	21.07	0.845	0.236	4.41
UPCDM	28.39	0.942	0.127	3.58

6 Conclusions

In the thriving cultural and creative industry, meeting users' personalised needs has become key to enhancing product competitiveness. To address the current challenges of cultural and creative products generation methods failing to meet user demands and generating poor image quality, this paper first designs a data-driven user profile construction method. This approach can target a large number of potential users for cultural and creative products on e-commerce platforms and automatically cluster user profiles and extract user demand characteristics through natural language processing algorithms. It also uses a TextCNN classification model to analyse the sentiment in user demands, obtaining the main product needs expressed by users as well as their positive and negative opinions. Based on this, an innovative method for personalised cultural and creative products generation is proposed that integrates user profiles with an improved conditional diffusion model. This method takes a diffusion model as its primary network and designs a new structure of residual block to effectively enhance the performance of image generation. The noise estimation network is improved by adding a CBAM attention module, which enhances the model's ability to extract key information in images and further improves image quality. By integrating conditional control with user profiles as the control network, it efficiently generates cultural and creative products that meet users' emotional needs. Experimental results show that the proposed method achieves improvements of at least 8.63% and 3.74% in PSNR and SSIM, respectively, compared to baseline methods, effectively enhancing the quality of generated images for cultural and creative products.

Acknowledgements

This work is supported by the 2025 Henan Provincial Department of Science and Technology Soft Science Project (No. 252400410618).

Declarations

All authors declare that they have no conflicts of interest.

References

Cachay, S., Zhao, B., Joren, H. and Yu, R. (2023) 'Dyffusion: a dynamics-informed diffusion model for spatiotemporal forecasting', *Advances in Neural Information Processing Systems*, Vol. 36, pp.45259–45287.

Cao, H., Tan, C., Gao, Z., Xu, Y., Chen, G., Heng, P-A. and Li, S.Z. (2024) 'A survey on generative diffusion models', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 36, No. 7, pp.2814–2830.

Cao, S., Chai, W., Hao, S., Zhang, Y., Chen, H. and Wang, G. (2023) 'Difffashion: reference-based fashion design with structure-aware transfer by diffusion models', *IEEE Transactions on Multimedia*, Vol. 26, pp.3962–3975.

Chen, H., Xiang, Q., Hu, J., Ye, M., Yu, C., Cheng, H. and Zhang, L. (2025) 'Comprehensive exploration of diffusion models in image generation: a survey', *Artificial Intelligence Review*, Vol. 58, No. 4, pp.85–99.

Chen, Y. and Carrillo, J.E. (2011) 'Single firm product diffusion model for single-function and fusion products', *European Journal of Operational Research*, Vol. 214, No. 2, pp.232–245.

Dai, M., Feng, Y., Wang, R. and Jung, J. (2024) 'Enhancing the digital inheritance and development of Chinese intangible cultural heritage paper-cutting through stable diffusion LoRA models', *Applied Sciences*, Vol. 14, No. 23, pp.32–51.

Han, S., Shi, Z. and Shi, Y. (2022) 'Cultural and creative product design and image recognition based on the convolutional neural network model', *Computational Intelligence and Neuroscience*, Vol. 20, No. 1, pp.25–42.

Huang, J. (2024) 'Personalized recommendation method for cultural creative products in tourism cities based on user profiles', *Procedia Computer Science*, Vol. 243, pp.1133–1142.

Lai, H., Huang, T., Lu, B., Zhang, S. and Xiaog, R. (2025) 'Silhouette coefficient-based weighting k-means algorithm', *Neural Computing and Applications*, Vol. 37, No. 5, pp.3061–3075.

Li, M., Zhang, J., Hong, Y., Xie, X., Meng, Z. and Gu, S. (2024) 'Advanced product personalization in the blockchain-enabled Metaverse: a diffusion model for automatic style generation', *IEEE Internet of Things Journal*, Vol. 12, No. 8, pp.10304–10315.

Li, X., Su, J., Zhang, Z. and Bai, R. (2021a) 'Product innovation concept generation based on deep learning and Kansei engineering', *Journal of Engineering Design*, Vol. 32, No. 10, pp.559–589.

Li, Z., Shu, S., Shao, J., Booth, E. and Morrison, A.M. (2021b) 'Innovative or not? The effects of consumer perceived value on purchase intentions for the palace museum's cultural and creative products', *Sustainability*, Vol. 13, No. 4, pp.84–92.

Liang, J. (2024) 'The application of artificial intelligence-assisted technology in cultural and creative product design', *Scientific Reports*, Vol. 14, No. 1, pp.69–80.

Liang, Y-S., Chen, C-Y., Li, C-T. and Chang, S-M. (2024) 'Personalized product description generation with gated pointer-generator transformer', *IEEE Transactions on Computational Social Systems*, Vol. 12, No. 1, pp.52–63.

Liu, L. and Zhao, H. (2024) 'Research on consumers' purchase intention of cultural and creative products – metaphor design based on traditional cultural symbols', *PloS One*, Vol. 19, No. 5, pp.78–94.

Liu, P., Long, Q., Luo, Y., Qu, H., Zhai, S. and Guo, F. (2025) 'Improved U-net network-based segmentation method for fast diffusion measurement', *Optics and Lasers in Engineering*, Vol. 186, pp.92–105.

Luo, Z., Sun, Z., Zhou, W., Wu, Z. and Kamata, S-I. (2022) 'Rethinking ResNets: improved stacking strategies with high-order schemes for image classification', *Complex & Intelligent Systems*, Vol. 8, No. 4, pp.3395–3407.

Miaskiewicz, T. and Kozar, K.A. (2011) 'Personas and user-centered design: how can personas benefit product design processes?', *Design Studies*, Vol. 32, No. 5, pp.417–430.

Moustafa, A.W. (2024) 'Cultural design in relation to consumer product design', *International Design Journal*, Vol. 14, No. 1, pp.419–430.

Shi, L. and Yang, X. (2025) 'Personalized recommendation algorithm for cultural and creative products based on fuzzy decision support system', *International Journal of Computational Intelligence Systems*, Vol. 18, No. 1, pp.102–116.

Sun, X., Hou, S., Cai, N. and Ma, W. (2020) 'Product information diffusion model and reasoning process in consumer behavior', *Heliyon*, Vol. 6, No. 12, pp.18–31.

Wang, Y. (2024) 'Research on the TF-IDF algorithm combined with semantics for automatic extraction of keywords from network news texts', *Journal of Intelligent Systems*, Vol. 33, No. 1, pp.20–33.

Wang, Z., Ping, Y., Lu, Z., Wang, J. and You, F. (2025) 'DM-CAM: a novel product design method combining diffusion models and class activation mapping for enhanced user-driven design', *Journal of Engineering Design*, Vol. 8, No. 4, pp.1–36.

Xu, X. and Zheng, J. (2022) 'Evaluation of cultural creative product design based on computer-aided perceptual imagery system', *Computer-Aided Design & Applications*, Vol. 19, No. S3, pp.142–152.

Yang, C., Liu, F. and Ye, J. (2023) 'A product form design method integrating Kansei engineering and diffusion model', *Advanced Engineering Informatics*, Vol. 57, pp.58–69.

Zhang, M. and Gao, H. (2025) 'Exploration of feature detection and fusion for cultural and creative product based on deep learning', *Computer-Aided Design and Applications*, Vol. 25, No. 1, pp.118–132.

Zhang, M., Li, X., Yue, S. and Yang, L. (2020) 'An empirical study of TextRank for keyword extraction', *IEEE Access*, Vol. 8, pp.178849–178858.

Zhao, S., Chen, D., Chen, Y-C., Bao, J., Hao, S., Yuan, L. and Wong, K-Y.K. (2023) 'Uni-controlnet: all-in-one control to text-to-image diffusion models', *Advances in Neural Information Processing Systems*, Vol. 36, pp.11127–11150.