International Journal of Information and Communication Technology

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

Deep learning-based innovative product design driven by social network data

Changhui Wang

DOI: 10.1504/IJICT.2026.10077935

Article History:

Received:	02 January 2026
Last revised:	04 February 2026
Accepted:	06 February 2026
Published online:	18 May 2026

Deep learning-based innovative product design driven by social network data

Changhui Wang

Art Design Education,
Zhengzhou Institute of Technology,
Zhengzhou 450000, China
Email: wch_0123@sina.com

Abstract: Social network data has become a vital resource driving product innovation and design. Current research struggles to fully uncover users' emotional needs toward products when dealing with unstructured, high-dimensional social data, resulting in subpar product quality. To address this, this paper first employs a multi-scale attention network to analyse product emotional needs, capturing users' emotional demands. Subsequently, a spatial cross-reconstruction module is designed within the generative adversarial network to obtain more refined features. Simultaneously, a semantic correlation attention module is designed for mapping emotional needs to product images. This extracts attribute and word encodings as semantic representations to guide image generation, enhancing semantic consistency between emotional needs and visual content. Experimental results demonstrate that the proposed method achieves 92.71% accuracy in emotional need recognition and an FID of 11.88 for product images, outperforming state-of-the-art methods and delivering outstanding performance in innovative product design tasks.

Keywords: innovative product design; social network; deep learning; emotional needs analysis; generative adversarial network; GAN.

Reference to this paper should be made as follows: Wang, C. (2026) 'Deep learning-based innovative product design driven by social network data', *Int. J. Information and Communication Technology*, Vol. 27, No. 49, pp.59–78.

Biographical notes: Changhui Wang is an Associate Professor at the School of Art and Media, Zhengzhou Institute of Technology, China. She obtained her Bachelor's degree (2007) and a Master's degree (2010) from Shaanxi Normal University. She has published 20 CN-indexed papers and one EI-indexed paper. Her research focuses on AI art design, vocational curriculum development, and teaching model innovation.

1 Introduction

The popularity of social networks has restructured the interaction logic between users and products, with users no longer being passive recipients but participants and co-creators of product iterations (Lee, 2021). Traditional product design mainly relies on designers' experience judgments and small-scale user research, which not only leads to issues such as lagging demand capture, incomplete scenario coverage, and insufficient personalisation, but also faces challenges like creative bottlenecks and long iteration cycles in the generation of product solutions, making it difficult to meet users' diversified needs for product experiences (Gai et al., 2024). Social networks can comprehensively reflect user requirements and provide full-chain decision-making support for product design. However, social network data has characteristics such as high dimensionality, fragmentation, and dense noise (Rathore et al., 2018). Traditional data mining methods struggle to deeply mine social comment text data and cannot support the automated and intelligent generation of product solutions (Bhad and Buktar, 2021). Deep learning technology provides a new solution for efficiently processing social network comment data and generating product designs (Gozuacik et al., 2021). The iterative upgrades of generative models such as generative adversarial networks (GAN), variational autoencoders (VAE), and transformers (Bordas et al., 2024) have not only achieved effective breakthroughs in key issues like multimodal data fusion and potential demand prediction, but also possess the capability to automatically generate product concepts and prototype solutions from requirement data. By deeply analysing and generating modelling of social network data through deep learning technologies, user emotional needs can be accurately analysed, injecting new vitality into innovative product design (Zhang et al., 2022).

Social network platforms are filled with a large amount of user comments on products. By mining this data, users' emotional needs can be fully understood, thus guiding merchants to design innovative products that align with their emotional preferences. The product evaluation data on social networks directly digitises users' actual needs, usage pain points, and emotional preferences, covering various dimensions of information such as explicit functional jokes, implicit experience expectations, and personalised scenario demands. By mining, analysing, and applying these data throughout the entire process of innovative product design, it can break through the limitations of designers' subjective predictions in traditional design, shifting product innovation from isolated indoor work to precise and user-centred innovation. From demand exploration, scheme design to implementation verification and iterative optimisation, it forms a complete user value support chain, while significantly increasing the success rate of innovation and reducing the cost of trial and error. Buker et al. (2023) quantified user comment data from social networks and established a mapping model between morphological feature design elements and sensory image scores using Quantitative Method I theory, assisting designers in optimising product design. Alaniz and Biazzo (2019) used semantic differential methods to obtain users' evaluations of products on social media networks and analysed the design features of the products to identify key design elements affecting sensory image evaluation. Jin et al. (2022) used crawler tools to analyse comment information from social platforms, conducted an in-depth study of user needs using a fuzzy Kano model, and thereby achieved innovative product designs. Li et al. (2018) employed latent Dirichlet allocation (LDA) topic modelling to separately analyse positive and negative reviews of new energy vehicles, extracting corresponding characteristic words. They built the mapping relationship between the feature words and product design. Comments from social networks are unstructured data, and due to characteristics such as information mixing and large volume, traditional methods have difficulty directly mining features within them. In recent years, with the widespread application of machine learning and big data technologies, it has become possible to process a large amount of online reviews and extract required information. Song (2024) used Word2Vec word embeddings as input features for a support vector machine model to determine users' sentiment tendencies in their comments. Then, he designed innovative products that matched user emotions using diffusion models. Chan et al. (2020) analysed user emotions using decision trees and introduced parametric modelling methods to achieve innovative product design. Li et al. (2021) mined user comments on electronic products from Twitter using term frequency-inverse document frequency (TF-IDF), providing references for product design. The massive user-generated social comment data on social network platforms possess unique characteristics such as unstructured nature, high noise level, dynamic nature, implicit emotions, network dependence, multi-modal integration, etc. These characteristics directly conflict with the limitations of traditional methods, making it difficult for traditional methods to effectively extract features. The data cleaning of traditional methods merely stops at simple rule filtering, lacking the ability to identify noise based on semantics, and is unable to cope with the high noise and low information density characteristics of social comment data, resulting in low product design efficiency.

Deep learning is the latest research trend in the field of machine learning, originating from artificial neural networks in the domain of machine learning. Deep learning methods have strong feature learning capabilities, allowing models with multiple layers to automatically learn high-level abstract features required by classification models, avoiding the difficulty of manually constructing features in traditional machine learning. Luo et al. (2025) utilised the convolutional neural network (CNN) and parsed user emotional hierarchies from social network comments using syntactic tree structures, introducing autoencoders to generate products that align with user emotions, significantly improving product quality. Guo and Ye (2023) proposed a parallel approach of two CNNs to capture semantic features in social text. Moreover, demand constraints are added to the Transformer to design products of better quality. Compared to generative models such as diffusion models and autoencoders, GAN generates detailed and high-fidelity design outputs through dynamic adversarial interaction between generators and discriminators. In visually dominant product design, this mechanism can produce designs closer to real-world styles, reducing common issues of blurriness or distortion in traditional generative models. Kang et al. (2023) used multi-label techniques and GAN for innovative generation design of product appearances, shortening the design cycle. Wang et al. (2025) proposed an intelligent automotive product design model based on GAN, improving design efficiency. Deng et al. (2023) analysed user emotional needs in social networks using bidirectional encoder representations from transformer (BERT) and generated innovative products using conditional generative adversarial networks (CGAN) with textual semantics as constraints. Tang et al. (2025) built a user demand recognition model based on residual network (ResNet) and established an emotion-preference-driven product design generation model by combining it with GAN. Niu and Zhou (2025), taking bamboo furniture as an example, combined deep CNN and deep convolutional GAN to construct a product emotional design generation model that can effectively generate bamboo furniture sketches meeting user emotional needs.

Through the analysis of current innovative product design research, it is clear that existing product design methods do not make sufficient use of information related to user emotional needs. Text semantics and product images show sparse mapping characteristics in space, thus affecting the correlation between images and emotional demand semantics. To address these issues, this paper proposes an innovative product design method based on deep learning driven by social network data. First, a multi-scale attention network is used for analysis of product emotional needs. By combining multi-scale feature extraction technology with hierarchical dynamic sparse attention mechanisms, rich local short-range and global long-range semantic information from user comments about products in social networks is captured. Core key information is dynamically filtered, effectively reducing the computational complexity of the model. To generate product images that meet user emotional needs, this paper adopts bilinear interpolation within GAN to link a spatial cross-reconstruction module for separating and reconstructing feature maps with different amounts of information, achieving more refined features. At the same time, a semantic association attention module is designed in the mapping between emotional demands and product images, extracting attribute encoding and word encoding as semantic representations to guide image generation, thereby improving the semantic consistency between emotional needs and visual content. Experimental results show that the proposed method improves emotional demand recognition accuracy by at least 4.48%, while reducing the Fréchet inception distance (FID) of generated product images by at least 28.82% compared to baseline methods, verifying the feasibility and effectiveness of the proposed approach.

2 Relevant theoretical analysis

2.1 Generative adversarial network

GAN is a deep learning model that generates high-quality data through adversarial training between the generator and discriminator. The model consists of two core components: the generator and the discriminator (Gui et al., 2021). The generator synthesises data samples from random noise, while the discriminator judges whether an input sample is authentic. Their adversarial optimisation process resembles the game between forgers and authentication experts. The generator continuously improves its counterfeiting techniques to deceive the discriminator, while simultaneously enhancing the discrimination capability of the discriminator until ultimately enabling the generator to output highly realistic samples. The core innovation of GAN lies in the zero-sum game training framework. The generator (G) attempts to forge realistic data to deceive the discriminator (D). Meanwhile, the discriminator strives to distinguish between real data and generated data. The two achieve dynamic optimisation through a minimax game. Deep learning algorithms such as CNN or LSTM use supervised learning or self-supervised learning, relying on fixed objective functions, which can easily fall into local optima or mode collapse. GAN continuously optimises the generator through dynamic feedback, avoiding local optima and enhancing the diversity and authenticity of the generated samples.

The generator G serves as a core component in deep neural network architecture, with an input being a random noise vector of specific dimensions and an output generating synthetic samples that simulate real data distributions. The network continuously optimises its generative capabilities through adversarial training, aiming to make the generated samples effectively deceive the discriminator. Correspondingly, the discriminator D also adopts a neural network structure, whose core function lies in accurately determining whether input samples are authentic. During model training, the discriminator receives two types of input data: real samples from datasets as positive examples and synthetic samples from the generator as negative examples. The training process for the generator and the discriminator is formalised as an optimisation problem in a two-player game (Ma et al., 2019), as shown in equation (1).

$$\min_G \max_D V(D,G) = E_{x\sim P_{\text{data}}(x)}\left[\log(D(x)\right] + E_{z\sim P_z(z)}\left[\log\left(1-D\left(G(z)\right)\right)\right] \quad (1)$$

where $E_{x\sim P_{\text{data}}(x)}$ represents the expectation of the real data distribution, while $E_{z\sim P_z(z)}$ represents the expectation of the input noise distribution. The generator G aims to minimise $\log(1 - D(G(z)))$, reducing the probability that the discriminator identifies generated samples as fake data. Conversely, the discriminator is trained to maximise $\log(D(x))$, reflecting its ability to correctly distinguish between real and generated samples.

Additionally, CGAN is an extension of the original GAN, where auxiliary conditional information y is introduced into the basic framework, allowing the generation process to be constrained by specific conditions (Baek et al., 2022). In CGAN, the training of the generator and discriminator is shown in equation (2), where y represents the conditional variable, such as a label or category, which guides the generation of data samples that are more closely related to the given condition.

$$\min_G \max_D V(D,G) = E_{x\sim P_{\text{dan}}(x)}\left[\log(D\left(x\middle|y\right)\right] + E_{z\sim P_z(z)}\left[\log\left(1-D\left(G\left(z\middle|y\right)\right)\right)\right] \quad (2)$$

2.2 Attention mechanism

The attention mechanism originates from the selective focusing characteristics of human vision. When observing complex scenes, humans naturally focus on significant targets while ignoring secondary backgrounds. This efficient information processing method inspired the modelling of attention mechanisms in machine learning. The core idea of an attention mechanism is: when processing input data, the model assigns different weights (i.e., attention scores) to different parts based on the current task's requirements; higher weights indicate that a specific part is more important for the current task (Chen et al., 2024). Integrating the attention mechanism into traditional neural networks can significantly enhance the model's performance. The attention mechanism enables the model to break free from the limitations of feature extraction in traditional networks, achieving precise focusing on key features, effective capture of long-distance dependencies, and adaptive integration of multi-source features. This fundamentally solves the core problems of traditional networks' global average attention and loss of long-distance information. The attention mechanism is not an independent deep learning model but a pluggable feature optimisation module. Its core idea stems from human visual attention, ignoring irrelevant and redundant information. By integrating this idea into traditional neural networks, the model can adaptively assign differentiated attention weights to features at different positions and dimensions, allowing the network to focus more on the core features valuable for the task and reducing the interference from irrelevant features. Ultimately, this leads to a dual improvement in model accuracy and generalisation ability.

To more specifically explain the working mechanism of attention mechanisms, we take self-attention (self-attention) as an example. First, input features are linearly transformed through trainable parameter matrices to generate query vectors Q, key vectors K, and value vectors V. Next, compute the dot product between Q and K to obtain feature similarity. Then scale and apply softmax normalisation to the similarity matrix to calculate attention weights. Finally, perform a weighted sum of V based on these weights to obtain the feature representation, d_k is the dimension of the vector. The calculation formulas are as follows.

$$output = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) \cdot V \tag{3}$$

3 Analysis of product sentiment needs in social networks based on multi-scale attention networks

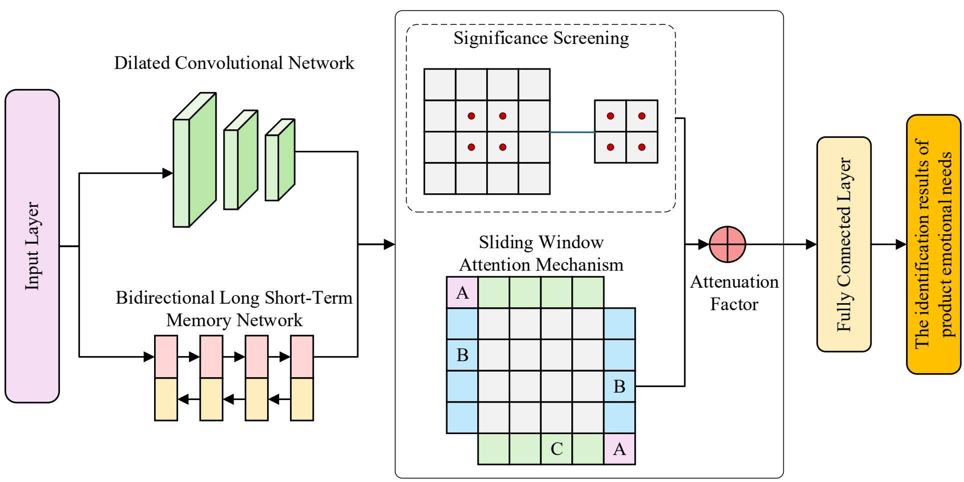
3.1 Product review text feature extraction

User-generated content in social networks contains rich emotional information, which reflects users' deep needs and experiential feelings toward products. However, for the task of text analysis on product user reviews, social network texts often involve a large amount of professional terminology, technical parameters, and detailed descriptions of product usage experiences, exhibiting obvious domain specificity and semantic complexity. This leads to traditional pre-trained language models struggling to achieve satisfactory performance when applied to such texts, as these models lack effective capabilities for capturing deep, fine-grained semantic dependencies among words in reviews. In recent years, multi-scale feature fusion strategies have attracted widespread attention in the field of semantic understanding. Features at different scales, such as lexical-level local features and sentence-level global features, often play different roles in text semantic understanding and exhibit complementary relationships. However, current multi-scale feature fusion methods mostly ignore the dynamic interactions and complementary relationships between features at different scales, lacking effective cross-scale semantic information fusion strategies, thus making it difficult to achieve true feature complementarity and collaborative optimisation.

To address the above challenges, this chapter proposes a sparse multi-scale attention network (SMSAN), as shown in Figure 1. SMSAN captures rich local short-range and global long-range semantic information in product user reviews from social networks by combining multi-scale feature extraction technology, hierarchical dynamic sparse attention mechanisms, and adaptive sparse rate control strategies. It dynamically selects core key information and effectively reduces the model's computational complexity to achieve an optimal balance between text analysis efficiency and semantic understanding accuracy.

This paper first preprocesses product review text data published by users on social platforms, removing noisy data (such as advertisements and irrelevant comments), processing missing values and outliers. The input raw product review texts are tokenised using the Chinese tokeniser provided with the a lite BERT (ALBERT) model (Areshey and Mathkour, 2024), and then mapped into Token sequences. After being processed through the word embedding layer of the ALBERT model, the Token ID sequence is transformed into a set of continuous embedding vectors that are context-dependent. ALBERT is a lightweight pre-trained language model optimised based on BERT, addressing issues such as large model parameter size, low training efficiency, and insufficient long-distance dependency capture. It retains the core advantage of BERT's deep embedding based on context semantics. Compared with shallow Word2Vec and the basic version of BERT, it has significant inherent advantages in terms of parameter size and training efficiency, long-distance semantic modelling, etc. It is particularly suitable for processing scenarios involving short, oralised text such as Chinese product review texts. From the perspective of model positioning, Word2Vec is a static word embedding model that can only generate a single word vector without context. BERT is a basic deep context embedding model, but it has problems of redundant parameters and high training costs. ALBERT is a lightweight and efficient deep context embedding model. Through word embedding factorisation, cross-layer parameter sharing, and sentence-level coherence loss, it significantly reduces parameter size and computational costs while improving the efficiency of semantic modelling and the ability to capture long-distance dependencies. This is its core inherent advantage compared to the former two models. As a key module for semantic modelling, the feature extraction layer is responsible for further extracting high-level semantic features from the initial embedding vectors generated by the ALBERT word embedding layer. This paper adopts a dilated convolutional neural network (DCNN) (Basha et al., 2025) for lexical-level feature extraction.

Figure 1 Product emotional needs identification process (see online version for colours)



Given an input sequence representation $x = [x_1, x_2, \ldots, x_n]$, traditional convolution operations calculate the output feature at position i, denoted as y_i, as shown in equation (4). The dilated convolution introduces a dilation factor d, and its calculation is shown in equation (5). When $d = 1$, dilated convolution degenerates into ordinary convolution. As d increases, the receptive field of the convolution kernel grows exponentially, capturing longer-range contextual dependencies.

$$y_i = \sum_{k=0}^{K-1} x_{i+k} \cdot w_k \tag{4}$$

$$y_i = \sum_{k=0}^{K-1} x_{i+k\cdot d} \cdot w_k \tag{5}$$

where x_i represents the word vector at position i in the user review text sequence, and w_k is the kth parameter of the convolution kernel.

DCNN can accurately and efficiently extract lexical-level features by expanding the receptive field of the convolution kernel, thereby laying a solid foundation for subsequent product sentiment demand analysis.

3.2 Product emotional needs recognition based on sparse attention mechanism

Aiming to address issues such as sharply increased computational complexity and reduced feature generalisation capability caused by traditional self-attention mechanisms, this paper designs SMSAN to dynamically filter global key information. SMSAN consists of two submodules: the sliding window local attention mechanism (SWLAM) and global semantic significance screening (GSSA).

SWLAM divides product review text into multiple local windows, where each word only calculates attention weights within a fixed-size window range, thus greatly reducing computational complexity. First input the text features $H = (h_1, h_2, \ldots, h_n)$ extracted in the previous section. For the word feature at position i h_i, the local attention weight is as follows.

$$\alpha_{ij}^{local} = \frac{exp\left(\left(W_q h_i\right)^{\mathrm{T}}\left(W_k h_j\right)\right)}{\sum_{k=i-w}^{i+w} exp\left(\left(W_q h_i\right)^{\mathrm{T}}\left(W_k h_k\right)\right)} \tag{6}$$

where h_i is the feature vector of each token obtained through DCNN. W_q and W_k are the projection matrices for query and key in the attention mechanism, respectively, and w is the window size.

Although SWLAM can effectively capture local semantics, it may miss key information that is dispersed and spans windows in review text. To address this issue, this paper further designs a GSSA mechanism to dynamically filter out the most semantically significant critical information from the text.

In GSSA, the significance score of each token is calculated through a semantic scoring function, as shown in equation (7), where W_s and b_s are learnable parameters, V is the context vector parameter, and σ is the sigmoid activation function. The output scores between 0 and 1 represent the importance degree of each word. All words are ranked based on semantic significance scores, and only the top $rate \times n$ words with the highest scores are selected to construct global sparse attention.

$$s_i = \sigma\left(V^T \tanh\left(W_s h_i + b_s\right)\right) \tag{7}$$

As product review text varies in length and semantic complexity, a fixed sparsity rate may not meet requirements across different text scenarios. Therefore, an adaptive sparsity rate control strategy is introduced in the global semantics module, as shown in equation (8), where h_{cls} is the DCNN output vector. The sparsity rate ‘rate’ dynamically adjusts based on text length and complexity, achieving an adaptive balance between computational efficiency and feature capture accuracy.

$$rate = sigmoid\left(\left(W_r h_{cls} + b_r\right)\right) \tag{8}$$

During SMSAN training, to more accurately capture fine-grained emotional needs in product review texts, this paper adopts a cross-entropy loss function to optimise the output product sentiment demand results. It dynamically adjusts category weights to alleviate data imbalance issues and combines L2 regularisation to suppress overfitting, thereby improving accuracy and robustness for product sentiment demand classification.

In product review text sentiment demand classification, negative reviews are fewer, leading to imbalanced label distribution. If the standard cross-entropy function is used directly, the model tends to favour high-frequency categories. Therefore, this model employs inverse frequency-based dynamic category weight adjustment. The weighted cross-entropy loss function is defined in equation (9), where N is the number of samples, C represents product sentiment demand categorised into positive, negative, and neutral classes, $y_{i,c}$ is the encoding of true labels, $p_{i,c}$ is the predicted probability, f_c is the weight coefficient, and β is the global scaling factor.

$$L_{CE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C}\frac{\beta}{\sqrt{f_c+\varepsilon}}\cdot y_{i,c}\log\left(p_{i,c}\right) \tag{9}$$

4 Innovative product design based on emotional needs and deep learning

4.1 Overview of innovative product design methods

With increasingly diverse consumer demands, traditional product design methods struggle to meet emotional needs across different user groups. Existing product design approaches integrating sentiment demand typically classify text features into word-level and sentence-level features for separate training, resulting in insufficient utilisation of relevant information about user sentiment demand. Additionally, both the semantic representation of text and visual characteristics of products exhibit sparse mapping patterns spatially, which negatively affects the correlation between images and semantic information. To generate products that meet user sentiment demands, this paper proposes an innovative product design method based on product sentiment demand and deep learning algorithms. The spatial cross module separately reconstructs the feature maps of products to enhance the precision of spatial visual features. A cross-modal semantic correlation attention module (SCAM) utilises attribute encoding and word encoding as embedding vectors, guiding the model to focus on sub-regions of product images most relevant to current sentiment demand, effectively enhancing the match between generated product images and user sentiment needs. Figure 2 shows the overall structure of the proposed method.

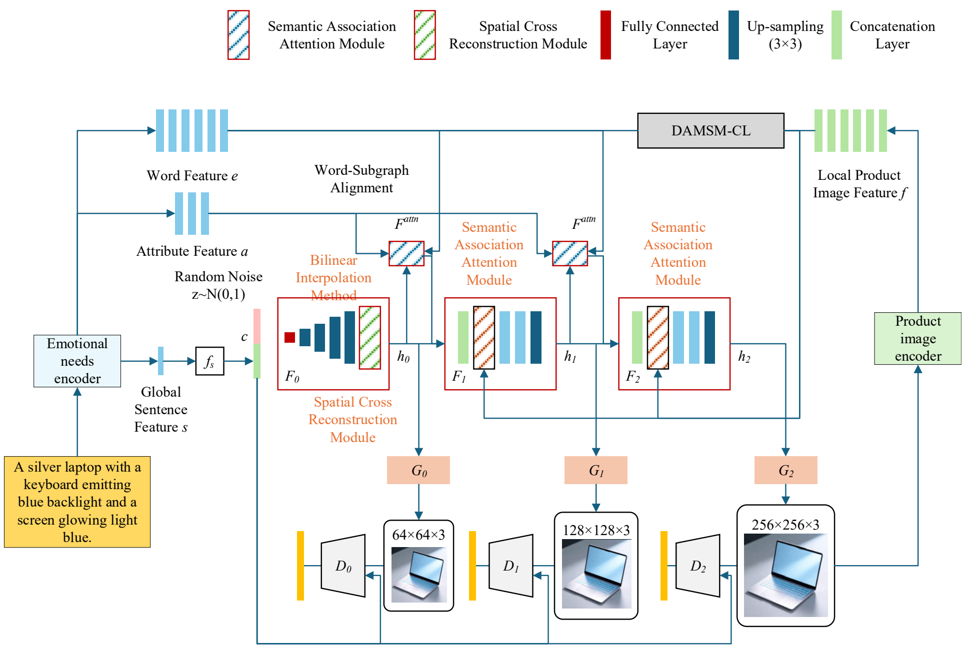
4.2 Cross-reconstruction of product images

In GAN-based product image design, if initial generated image quality is poor or structural anomalies occur, these issues will accumulate during subsequent refinement stages, ultimately leading to a decrease in visual fidelity and semantic consistency of the output products. The network architecture presented in this paper consists of multiple generators and discriminators, following a tree-like structure. This paper first generates low-resolution images at the initial stage by using sentence embeddings enhanced with conditional semantic augmentation (CSA) and random noise vector z to generate an initial image h_0, containing accurate colours and coarse structures of products; subsequent branches focus on generating high-resolution product images emphasising details. h_0 represents the product image features corresponding to this initial stage, as shown in equation (10).

$$h_0 = F_0\left(s_{ca}, z\right) \tag{10}$$

where F_0 is the generator at the initial generation stage, whose original structure consists of 1 fully connected layer and 4 upsampling layers. This paper adopts bilinear interpolation during upsampling, and after the last upsampling operation, it goes through a designed spatial cross reconstruction module with improvements.

Figure 2 The overall structure of the proposed innovative product design method (see online version for colours)



The core purpose of feature separation is to effectively distinguish product feature maps with dense information from those with sparse spatial content. This paper uses scale factors in group normalisation (GN) layers to evaluate the information content of different feature maps, as shown below.

$$GN(X) = \gamma \frac{X - \mu}{\sqrt{\sigma^2 + \xi}} + \eta \tag{11}$$

where μ and σ are the mean and standard deviation of X, respectively, while ξ is an extremely small constant introduced to ensure numerical stability. γ and η are learnable affine transformation parameters.

In GN layers, γ measures the variance of spatial pixels across each batch and channel. The randomly initialised parameter γ is normalised to obtain relevant weights W_γ, which calculates the importance of different feature maps.

$$W_\gamma = \{w_i\} = \frac{\gamma_i}{\sum_{j=1}^{C} \gamma_j} \tag{12}$$

After applying the sigmoid function, the adjusted product feature map weight values are normalised into a preset interval with a threshold set for gating in order to achieve precise guidance via W_1 and W_2. Weights above the threshold represent information-rich key features while those below represent redundant features.

$$W_n = Gate\Big(sigmoid\big(W_\gamma\,(GN(X))\big)\Big),\ n = 1, 2 \tag{13}$$

Finally, X is multiplied by W_1 and W_2 respectively, resulting in two weighted feature maps: one with rich informational content and strong expressive power denoted as X_1^w, and the other with less information that can be considered redundant, denoted as X_2^w.

$$\begin{cases} X_1^w = W_1^* X \\ X_2^w = W_2^* X \end{cases} \tag{14}$$

To further reduce the impact of spatial redundant features, we choose to use reconstruction operations by adding information-dense and sparse features to generate more comprehensive product features. Group convolution and pointwise convolution are applied before cross-reconstruction to enhance local structures and texture details in the input images. The cross-constructed features X_1^w and X_2^w are concatenated to obtain a spatially refined feature map X^w.

4.3 Emotion-driven innovative product design

4.3.1 Conditions for enhancing emotional needs in integrated products

Conditional enhancement is employed to convert sentence features of product emotional requirements into conditional vectors and reduce their dimensionality. The mismatch between product emotional requirements and limited sample sizes may lead to discontinuous latent data manifolds, which could affect the training stability of neural networks. However, the conditional enhancement module can smooth text condition distributions in the latent space, enrich semantic information for product emotion requirement features, enhance diversity in generated product images, as shown in Figure 3.

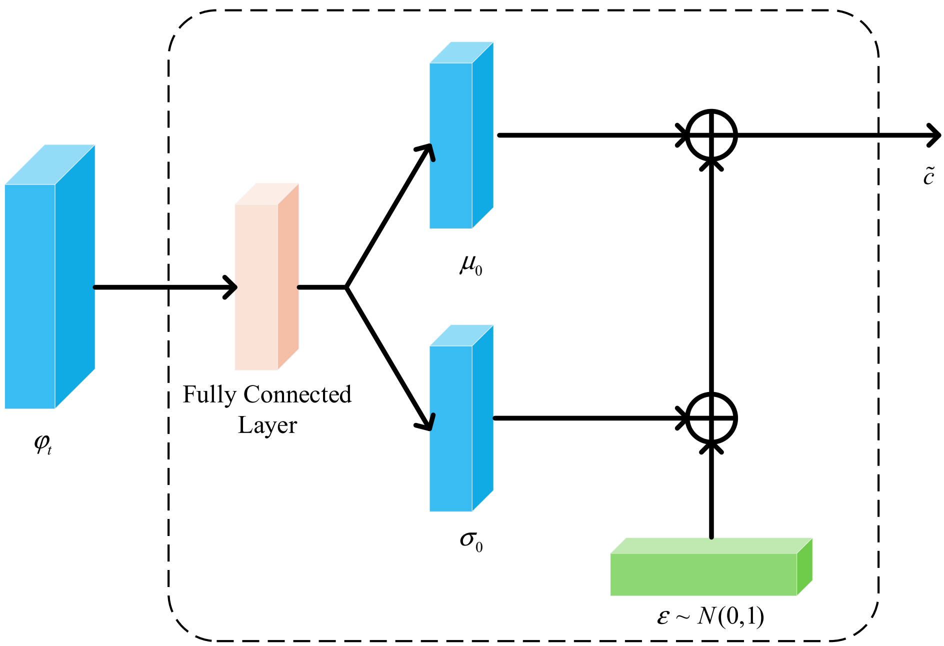
To obtain the enhanced product emotion requirement text features $\tilde{c}$, first, the initial text features φ_t are mapped to distribution parameters in the latent space through a fully connected layer, generating $N\left(\mu_0\left(\varphi_t\right), \sum_0\left(\varphi_t\right)\right)$ and μ_0 in Gaussian distribution σ_0, then using reparameterisation trick to randomly sample conditional latent variables from this Gaussian distribution. The mean function $\mu_0(\varphi_t)$ and diagonal covariance function

$\sum_0(\varphi_t)$ are functions of product emotion requirement features φ_t, as shown in equation (15), where ⊗ is element-wise multiplication, and ζ follows a normal distribution with mathematical expectation μ and variance σ^2.

$$\tilde{c} = \mu_0 + \sigma_0 \otimes \zeta \tag{15}$$

By concatenating product emotion requirements with random noise. Inputting the same requirement, differences in initial noise will lead to different generated product images, suppressing overfitting. The conditional enhancement mechanism allows the model generating more diverse visual outputs corresponding to user's emotion requirements when generating products images, including variations in shape, structure and posture, thus significantly enhancing the diversity of generated samples.

Figure 3 Condition enhancement module (see online version for colours)



4.3.2 The semantic connection between emotional needs and product design

When generating product image sub-regions, attention is guided to focus on different emotional requirements; however, relying solely on emotional requirements cannot ensure global semantic consistency. When selecting matrix multiplication to calculate the correlation between product review text words and product image features, attribute features are introduced based on word features. When matching the most relevant word vectors, attribute information related to product image content can be considered, dynamically fusing user emotional requirement information into product image features. SCAM has three inputs: word features e, attribute features a, and product image features h from the hidden layer.

Firstly, in the attention module, different dimensional attribute encodings and word encodings are transformed through newly added perception layer convolutions to a common vector space of image features. Secondly, attribute encodings are combined with word encodings by stacking along the row direction of word vectors e to reinforce language features on the word level.

$$w = Ue + Va_m \tag{16}$$

where U and V are the perception layers that transform word features and attribute features during computation. w is the processed word encoding, n_r is the dimensionality of product image feature vectors, n_t is the dimensionality of text embedding, and T is the number of words.

Finally, enhanced word encodings are used to refine fine-grained details in product images. The contextual vector c_j of a word is a dynamic representation of all words with respect to the jth sub-region h_j of the current image, constraining its generation by matching the most relevant emotional word vectors as shown in equation (17).

$$c_j = \sum_{i=0}^{T-1} \beta_{j,i} w_i \tag{17}$$

where $\beta_{j,i} = \exp\left(s_{j,i}\right) \Big/ \sum_{k=0}^{T-1} \exp\left(s_{j,k}\right)$ and $s_{j,i} = h_j^T w_i$ are the inner product similarities between the ith word of product emotional demand and the jth sub-region of the product image, N is the size of the image, and $\beta_{j,i}$ represents the importance of the ith word when generating the jth sub-region of the product image.

Contextual relations are crucial for generating high-resolution details by linking words to spatial positions in product images, focusing on the most relevant sub-regions to form a multi-modal context vector.

$$F^{atm}(e, a, h) = \left(c_0, c_1, ..., c_{N-1}\right) \tag{18}$$

SCAM improves semantic bias caused by pre-trained text encodings during product image generation through attribute encoding and enhances visual features of products using encoded rich word contextual information.

5 Experimental results analysis

5.1 Quantification of emotional features in products

This paper uses crawler software to collect online reviews for computer products on social platforms. By crawling review data from mainstream social platforms, a total of 36,245 computer product reviews were obtained. After data cleaning and removing duplicates and invalid information, finally, 23,281 computer product review data entries formed the dataset. Crawled information includes reviewer usernames, names of reviewed computer products, content of reviews, as well as likes counts and purchase times for each review. The dataset was divided into training set, test set, and validation set in a ratio of 7:2:1. All experiments were conducted on an Intel I7-7700HQ CPU, 4G Nvidia GTX1050 GPU, Win10 64-bit computer based on the

Tensorflow1.12.0+Keras2.2.4 deep learning framework using the constructed dataset for training. The hyperparameters used in the experiments are shown in Table 1.

Table 1 Experimental parameter settings

Parameter	*Value*
Hidden size	768
Embedding size	128
Number of layers	12
Attention heads	12
Vocabulary size	30,000
Max sequence length	512
Learning rate	0.0002
Image feature dimension	2,048
Loss function hyperparameters	5

After calculation, the emotional quantification results of computer products in the dataset are shown in Table 2. As can be seen from Table 2, sorted by attention level from highest to lowest, the order of computer product features is performance, appearance, service conditions, screen, hardware configuration, system, keyboard, hard disk, sound effects, and network.

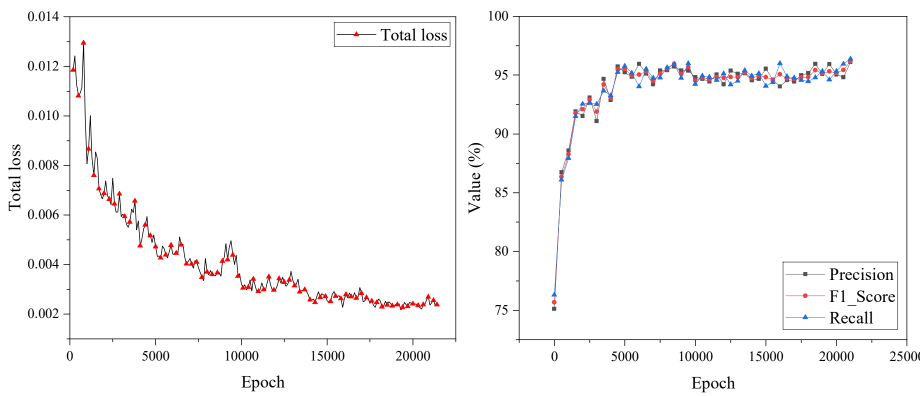
Table 2 The emotional quantification results of various product features of computers

Product features	*User sentiment index*		*User attention*	*User satisfaction*
	Positive	*Negative*		
Performance	0.645	0.075	0.700	79.17
Appearance	0.450	0.018	0.468	92.31
Service condition	0.442	0.038	0.500	84.17
Screen	0.260	0.020	0.280	85.71
Hardware configuration	0.198	0.015	0.232	85.92
System	0.121	0.010	0.143	84.73
Keyboard	0.091	0.018	0.099	66.97
Hard disk	0.086	0.018	0.095	65.38
Sound effect	0.062	0.005	0.074	85.07
Network	0.053	0.005	0.064	82.76
Average value	0.241	0.022	0.263	81.22

The product characteristic of performance has the highest level of attention, but its satisfaction rate is relatively low; such characteristics should be given priority consideration. From the search results, consumers are dissatisfied with program execution and standby time in the performance features of computers: programs run sluggishly and take longer to execute, while battery capacity is too small to meet consumer demands. In practical terms, computer hardware has an aggressive power consumption unlock policy, which can easily cause processor throttling during use, leading to sluggish program operation and extended running times. Therefore, from a design perspective, products with different battery capacities could be provided according to needs to avoid performance shortages caused by insufficient battery capacity.

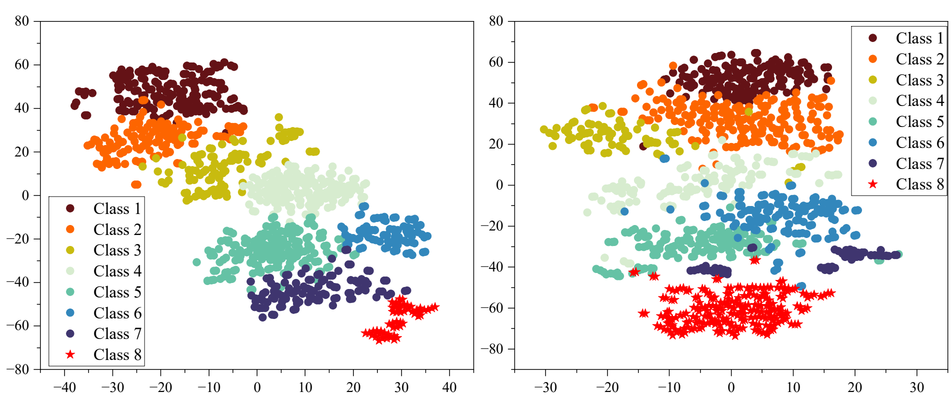
This paper analyses the performance of the product sentiment demand classification model SMSAN. As shown in Figure 4, the total loss of the proposed model SMSAN is depicted in Figure 4(a), and the model's performance on the validation set is presented in Figure 4(b). From the figure, it can be seen that the total loss tends to stabilise and reaches its minimum threshold around 10,000 steps. Meanwhile, although there are fluctuations in the model's performance on the validation set, both the overall accuracy metric and F1_Score exceed 90%. This indicates not only does the model efficiently extract product characteristics from social media platform comment texts but also operates stably.

Figure 4 Performance analysis of the proposed model (a) total loss (b) emotional need recognition performance (see online version for colours)



(a) (b)

Figure 5 The sentiment demand classification results of SMSAN model (a) the classification results of emotional needs of the SMSAN model (b) the classification results of the original product review data (see online version for colours)



(a) (b)

The sentiment demand classification results of SMSAN model on 8 categories of mobile phone products are shown in Figure 5. Figure 5(a) shows the classification results of the proposed SMSAN model, while Figure 5(b) presents the classification results of the original product review data. From the classification effect in Figure 5(a), it can be seen that the distribution of classification centre points of the proposed SMSAN model is more uniform and reasonable. However, in Figure 5(b), there are many mixed regions with different coloured dots, and the clusters where black dots locate are relatively dispersed; the unreasonable distribution of classification centre points leads to a poor final classification effect.

To further verify the sentiment demand classification performance of SMSAN model, this paper uses objective evaluation indicators such as mean average precision (mAP), average recall per class (CR), average F1 score per class (CF1), overall average recall (OR), and overall average F1 score (OF1) to compare SMSAN with PIOSVM (Song, 2024), VAECNN (Luo et al., 2025), DCTRAN (Guo and Ye, 2023), BCGAN (Deng et al., 2023), PSR-GAN (Tang et al., 2025), RIGAN (Niu and Zhou, 2025), as shown in Table 3. The mAP of SMSAN is 92.71%, which represents an improvement of 4.48%–19.12% over the baseline models. The CR and OR of SMSAN are 90.82% and 93.57%, respectively, representing at least a 4.77% and 3.54% increase compared to the baseline models. The CF1 and OF1 of SMSAN are 93.06% and 95.62%, representing at least an improvement of 4.12% and 3.96% over the baseline models. SMSAN captures short-range semantic dependencies through local sliding windows, while effectively combining local and global semantic information by integrating a global semantic saliency screening module. The introduced adaptive sparse rate regulation module and syntactic distance attenuation factor further reduce redundant information interference and improve model accuracy.

Table 3 Product-specific emotional needs recognition performance

Model	*mAP/%*	*CR/%*	*CF1/%*	*OR/%*	*OF1/%*
PIOSVM	73.59	74.71	77.84	78.92	82.01
VAECNN	76.04	79.42	80.55	83.46	84.53
DCTRAN	80.53	82.66	84.52	85.88	87.04
BCGAN	81.42	81.97	83.61	83.69	86.48
PSR-GAN	85.95	84.51	85.73	87.27	88.75
RIGAN	88.23	86.05	88.94	90.03	91.66
SMSAN	92.71	90.82	93.06	93.57	95.62

5.2 Product image generation performance analysis

This paper uses peak signal-to-noise ratio (PSNR), inception score (IS), FID, and R-precision (RP) to analyse the product image generation quality of PIOSVM, VAECNN, DCTRAN, BCGAN, PSR-GAN, RIGAN, and SMSAN models, as shown in Figure 6. The higher the values of PSNR, IS, and RP are, the better the product image generation effect is. FID has an inverse relationship with the product image generation effect. The PSNR of SMSAN is 59.86, which represents at least a 17.42% improvement over baseline models, indicating that SMSAN achieves higher-quality product image generation. IS evaluates both the clarity of images and the diversity within generated batches. Compared to PIOSVM, VAECNN, DCTRAN, BCGAN, PSR-GAN, and RIGAN, the IS of SMSAN improved by 44.58%, 26.71%, 18.35%, 8.92%, 6.10%, and 1.23%, respectively. The RP metric evaluates whether the generated product images meet user preferences through the matching degree between product images and user emotional needs (Zhou et al., 2021). The RP of SMSAN is 75.24, representing an improvement of 8.46%-75.26% over baseline models. FID is a metric used to quantify the similarity in statistical distribution between generated images and real images. A lower value indicates that the quality and diversity of generated images are closer to those of real images overall. The FID of SMSAN is 11.88, which represents at least a 28.82% reduction compared to baseline models.

Figure 6 Performance comparison of different product design methods (see online version for colours)

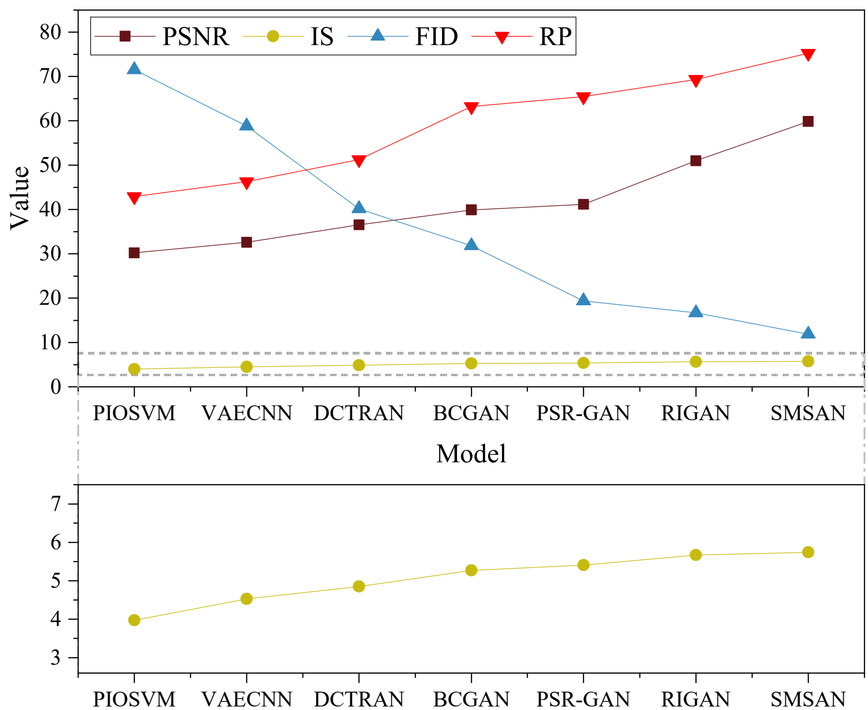


Table 3 Statistical significance test results

The suggested method	*Comparative method*	*PSNR*	*IS*	*FID*	*RP*
SMSAN	PIOSVM	0.041	0.035	0.044	0.035
	VAECNN	0.026	0.021	0.032	0.029
	DCTRAN	0.021	0.014	0.025	0.020
	BCGAN	0.017	0.011	0.018	0.016
	PSR-GAN	0.015	0.007	0.019	0.013
	RIGAN	0.012	0.003	0.015	0.009

To further verify the effectiveness of SMSAN, in a 5-fold cross-validation, this paper conducts statistical significance tests on SMSAN to assess the reliability of the above experimental results. The significance level α is set at 0.05, and the P-value test results for PSNR, IS, FID, and RP between SMSAN and the other six models are shown in Table 4. All P-values for each performance metric between SMSAN and the other six models are less than 0.05, indicating that there is a significant difference in performance between SMSAN and the other six models. The innovative product image generation performance of SMSAN surpasses comparative methods, achieving high-quality product images aligned with user emotional needs.

6 Conclusions

A vast amount of comment data from users on social networks continues to emerge, providing rich decision-making basis for product design. To address the issue that existing research fails to fully mine users' emotional needs in social networks, resulting in designed products deviating from users' preferences, this paper proposes an innovative product design method based on deep learning driven by social network data. Firstly, analyse the emotional needs of the product based on the multi-scale attention network. The short-range dependency relationship is captured by using the local attention of the sliding window, and combined with the global semantic saliency screening at the same time, the collaborative fusion of local and global information is effectively achieved. In addition, the model also incorporates an adaptive sparsity rate control module to capture the rich local short-range and global long-range semantic information in product user comments in social networks, and dynamically screen the core and key information.

Declarations

All authors declare that they have no conflicts of interest.

References

Alaniz, T. and Biazzo, S. (2019) 'Emotional design: the development of a process to envision emotion-centric new product ideas', *Procedia Computer Science*, Vol. 158, pp.474–484.

Areshey, A. and Mathkour, H. (2024) 'Exploring transformer models for sentiment classification: a comparison of BERT, RoBERTa, ALBERT, DistilBERT, and XLNet', *Expert Systems*, Vol. 41, No. 11, pp.13–25.

Baek, F., Kim, D., Park, S., Kim, H. and Lee, S. (2022) 'Conditional generative adversarial networks with adversarial attack and defense for generative data augmentation', *Journal of Computing in Civil Engineering*, Vol. 36, No. 3, pp.44–57.

Basha, R., Pathak, P., Sudha, M., Soumya, K. and Arockia Venice, J. (2025) 'Optimization of quantum dilated convolutional neural networks: image recognition with quantum computing', *Internet Technology Letters*, Vol. 8, No. 3, pp.70–81.

Bhad, P.U. and Buktar, R.B. (2021) 'Integrated approach for an artificial intelligence-based generative product design', *International Journal of Design Engineering*, Vol. 10, No. 2, pp.110v135.

Bordas, A., Le Masson, P., Thomas, M. and Weil, B. (2024) 'What is generative in generative artificial intelligence? A design-based perspective', *Research in Engineering Design*, Vol. 35, No. 4, pp.427v443.

Buker, T., Schmitt, T., Miehling, J. and Wartzack, S. (2023) 'Exploring the importance of a usable and emotional product design from the user's perspective', *Ergonomics*, Vol. 66, No. 5, pp.580v591.

Chan, K.Y., Kwong, C., Wongthongtham, P., Jiang, H., Fung, C.K., Abu-Salih, B., Liu, Z., Wong, T. and Jain, P. (2020) 'Affective design using machine learning: a survey and its prospect of conjoining big data', *International Journal of Computer Integrated Manufacturing*, Vol. 33, No. 7, pp.645–669.

Chen, Y., Xia, R., Yang, K. and Zou, K. (2024) 'DNNAM: image inpainting algorithm via deep neural networks and attention mechanism', *Applied Soft Computing*, Vol. 154, pp.32–44.

Deng, Z., Lv, J., Liu, X. and Hou, Y. (2023) 'Bionic design model for co-creative product innovation based on deep generative and BID', *International Journal of Computational Intelligence Systems*, Vol. 16, No. 1, pp.8–15.

Gai, T., Wu, J., Liang, C., Cao, M. and Zhang, Z. (2024) 'A quality function deployment model by social network and group decision making: application to product design of e-commerce platforms', *Engineering Applications of Artificial Intelligence*, Vol. 133, pp.108–119.

Gozuacik, N., Sakar, C.O. and Ozcan, S. (2021) 'Social media-based opinion retrieval for product analysis using multi-task deep neural networks', *Expert Systems with Applications*, Vol. 183, pp.53–68.

Gui, J., Sun, Z., Wen, Y., Tao, D. and Ye, J. (2021) 'A review on generative adversarial networks: algorithms, theory, and applications', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 35, No. 4, pp.3313–3332.

Guo, W. and Ye, H. (2023) 'Emotional analysis and expression in advertising art design based on deep learning algorithms', *Computer-Aided Design and Applications*, Vol. 21, pp.147-162.

Jin, J., Jia, D. and Chen, K. (2022) 'Mining online reviews with a Kansei-integrated Kano model for innovative product design', *International Journal of Production Research*, Vol. 60, No. 22, pp.6708–6727.

Kang, X., Nagasawa, S.y., Wu, Y. and Xiong, X. (2023) 'RETRACTED: emotional design of bamboo chair based on deep convolution neural network and deep convolution generative adversarial network', *Journal of Intelligent & Fuzzy Systems*, Vol. 44, No. 2, pp.1977–1989.

Lee, K. (2021) 'A systematic review on social sustainability of artificial intelligence in product design', *Sustainability*, Vol. 13, No. 5, pp.26–38.

Li, S., Zhang, Y., Li, Y. and Yu, Z. (2021) 'The user preference identification for product improvement based on online comment patch', *Electronic Commerce Research*, Vol. 21, No. 2, pp.42–57.

Li, Z., Tian, Z., Wang, J., Wang, W. and Huang, G. (2018) 'Dynamic mapping of design elements and affective responses: a machine learning based method for affective design', *Journal of Engineering Design*, Vol. 29, No. 7, pp.358–380.

Luo, S., Zhang, X. and Wang, N. (2025) 'Integrating Kansei engineering and GA-CNN-Attention for emotionally driven design: a methodological approach to aligning product features with user preferences', *Journal of Engineering Design*, Vol. 4, pp.1–47.

Ma, J., Yu, W., Liang, P., Li, C. and Jiang, J. (2019) 'FusionGAN: a generative adversarial network for infrared and visible image fusion', *Information Fusion*, Vol. 48, pp.11-26.

Niu, M. and Zhou, Y. (2025) 'The optimization of illustration design in cultural and creative products for liaoning region under intelligent generative adversarial network', *IEEE Access*, Vol. 13, pp.114746–114755.

Rathore, A.K., Das, S. and Ilavarasan, P.V. (2018) 'Social media data inputs in product design: case of a smartphone', *Global Journal of Flexible Systems Management*, Vol. 19, No. 3, pp.255–272.

Song, H. (2024) 'Product image modelling optimisation design method based on improved support vector machine', *International Journal of Product Development*, Vol. 28, No. 4, pp.227–240.

Tang, W.-Y., Xiang, Z.-R., Yu, S.-L., Zhi, J.-Y. and Yang, Z. (2025) 'PSR-GAN: a product concept sketch rendering method based on generative adversarial network and colour tags', *Journal of Engineering Design*, Vol. 6, pp.1–23.

Wang, S., Liu, Y., Sun, L. and Chen, G. (2025) 'Automobile exterior emotional design method based on deep learning and multiple views imagery integrating calculation', *Expert Systems with Applications*, Vol. 262, pp.125–137.

Zhang, Y., Tu, Z., Zhao, W. and He, L. (2022) 'Design of emotional branding communication model based on system dynamics in social media environment and its influence on new product sales', *Frontiers in Psychology*, Vol. 13, pp.86–93.

Zhou, A., Liu, H., Zhang, S. and Ouyang, J. (2021) 'Evaluation and design method for product form aesthetics based on deep learning', *IEEE Access*, Vol. 9, pp.108992–109003.