Conditional GAN-Based System for Automated Packaging Design and Market Demand Alignment Using Multi-modal Evaluation Networks

Li Zhang
School of Arts and Design, Guangdong University of Science and Technology, Dongguan City, Guangdong Province, 523083, China
E-mail: zl060111030014@163.com

Keywords: generative adversarial networks (GANs), packaging design, market demand matching, attention mechanism

Received: June 16, 2025

This paper introduces a cGANs-founded system designed for the autonomous creation of packaging designs to meet real-world market demand. The system architecture involves three different levels: a data acquisition level that aggregates multimodal data from brand design banks, consumer feedback, Enterprise Resource Planning (ERP) and Customer Relationship Management (CRM) systems, as well as trends from social media platforms; a cGANs-based generation level that uses a U-Net generator with a PatchGAN discriminator to create design products that are sensitive to market trends and heritage brand identity; and an evaluation level that uses a dual-stream network with ResNet-50 to extract image features and BERT to analyze semantic user feedback. The dataset used for this work comprised 10,000 labeled packaging images with 500,000+ entries describing both structural and unstructured consumer behaviors. The experimental results achieved best performance compared with baseline approaches, achieving a Fréchet Inception Distance (FID) of 28.4 and an Inception Score (IS) of 12.7, exhibiting a high quality and diversity of auto-generated images. Furthermore, the model achieved a 9.1/10 brand consistency score, 15% improvement in consumer satisfaction, and 20.2% increased conversion rate of purchases with A/B tests. The results prove that there is validity to the argument that this proposed system can create high-quality and brand-consistent packaging designs to suit market demand.

Povzetek: Predstavljen je sistem za avtomatsko oblikovanje embalaže, ki s pogojnim GAN, U-Net generatorjem, PatchGAN diskriminatorjem ter ocenjevalno mrežo ResNet-50 in BERT usklajuje oblikovanje z dejanskim povpraševanjem.

1 Introduction

In today's consumer market environment, the packaging design's role has transformed from its historically perceived functions of aesthetic value and protection of content to being a key market differentiation, consumer interactivity, and brand communication vehicle. Packaging now moves beyond the protection of products and the relay of information to impacting consumer purchase decisions significantly through the psychological and emotional aspects related to their aesthetic value and experiential value. In this fast-changing market environment, corporations face increasing difficulties in the development of new packaging products responsive to the spontaneous consumer tendencies and the predominant market trend [1]. With the use of historically perceived instincts of designers, past references, and market segmentation for assumptions, classic design strategies are less effective in meeting the fast-changing, frequently obscure consumer demands influenced by digital media and cultural narratives.

Traditional methods of packaging design are often characterized through their cyclical methods, labor-intensive nature, and susceptibility to subjective interpretation, which eventually lead to inefficiencies in the satisfaction of market needs. Consequently, artificial intelligence (AI), especially Generative Adversarial Networks (GANs), has been a revolutionary method in the area of design automation. GANs consist of two main components: a generator and a discriminator, which are in adversarial interaction in generating outputs closely approximating human creativity with high accuracy. Their usage in fields like fashion, visual art, and product design has proven their ability to produce content that preserves stylistic consistency and contextual flexibility. Zhang (2022), for instance, applied GANs in packaging design, which led to improvements in both semantic coherence and visual realism of generated content [2]. Even though they promise to make things easier, adding GANs to packaging design makes things more complicated. Automated design processes have to follow rules to protect the brand, follow the law, be sensitive to other cultures, and protect the environment. This study suggests a conditional GAN-based architecture that uses multi-source data analysis and demand prediction to make packaging solutions that are new, real-time, and responsive to the market. [3]. Against this backdrop, this current work suggests a new approach to creating packaging designs by using Conditional Generative

Adversarial Networks (cGANs), with a dual-stream assessing module and with different data inputs. The main thrust of this approach is to create designs that not only are qualitatively new but also are strongly compatible with brand identity as well as with modern market specifications. The scope and focus of this work are broadened by specifying its main research goals as follows:

G1: Automate the packaging design process using a Conditional GAN framework, incorporating a U-Net generator and a PatchGAN discriminator to ensure high-quality and brand-aligned design generation.

G2: Achieve real-time market alignment by combining and analyzing various data sources such as enterprise resource planning (ERP) systems, customer relationship management (CRM) data, consumer ratings, and trend indicator data obtained from various social media sites.

G3: Evaluate product development success using a dual-path neural network that combines ResNet-50 to extract visual features with BERT to conduct semantic consumer feedback analysis. The approach is to be carried out under a multi-task prediction model to predict consumer satisfaction and corresponding purchase conversion rates simultaneously.

G4: Cultivate cross-cultural adaptability via regional dataset integration with localized attention mechanisms that consider regional regulatory environments, cultural motifs, and regional end-market expectations.

These aims form the basic basis of this intended framework and underpin structural methodology, experimental protocols, and evaluation schemes outlined within subsequent sections.

2 Related work

There has been a lot more research on computational creativity in packaging design in the last few years. Numerous studies have demonstrated that AI-driven tools are effective for design innovation. The study [4] found that AI-generated content (AIGC) could help people come up with new ideas and automate the design process, making it very scalable and easy to use for different brands. Min Li proposed an artificial encryption scheme for computer communication in 2024 that uses Generative Adversarial Networks (GANs), Adversarial Neural Cryptography (ANC), and is resistant to the Chosen Ciphertext Attack (CCA). The scheme is called CCA-ANC [5]. Xia et al., 2022 [6] sought to rectify this deficiency by incorporating semantic prediction models into design workflows, enabling packaging design to reflect inferred consumer preferences. However, the lack of feedback loops that connect produced outputs to actual performance metrics is still a big problem. There have been a lot of advances in using GANs for packaging. The study [7] proposed a fundamental GAN architecture for generating product labels, demonstrating its feasibility while lacking mechanisms for market integration. On the other hand, this study builds on earlier work by using conditional inputs, brand DNA, and market trends, and by using multi-task learning to measure results, similar to the study in [8]. The rise of dual-stream and multi-modal architectures marks a shift toward design systems that are smarter and more connected to the market. The model Pix2Pix in [9] showed promise in using text and images to guess how good something looks. This is like the evaluation layer in the current model, which uses ResNet-50 to find image features and BERT to find semantic sentiment vectors. This makes it easier to see how well the market is lined up. The study [10] stressed the technically robust method for style-transfer and generative adaptation. Tests of cross-cultural adaptability showed that this worked. The [11], focuses on controlled generative design, relevant for cultural attribute modeling. The current study builds on this by using Latent Dirichlet Allocation (LDA) to find new market trends, which makes the designs that are made more useful and responsive. Other studies, such as [12], which supports multimodal generation, and [13], which used deep learning to improve regional packaging features, are similar to the regional test sets used in this study to compare cultures. People are also paying more attention to environmental and sustainability issues, which generative models don't always take into account. In [14], gives a full review of the rules for cross-cultural visual communication. It's perfect for helping people talk about how to be flexible with different cultures in generative design. As shown in Table 1, the proposed model outperforms existing methodologies in generation realism, as reflected by the low FID, diversity, representing the high IS, and market performance metrics such as consumer satisfaction and purchase conversion rates. Unlike existing models, it successfully combines dual-stream analysis, attention-augmented generation, and semantic matching methods to provide a more all-encompassing and market-sensitive solution to packaging design automation.

Table 1: Summary of related work

Model / Method	FID ↓	IS ↑	Brand Consistency (Score/10)	Consumer Satisfaction (+%)	Conversion Improvement (%)	Key Limitation
Traditional GAN [15]	42.3	8.2	7.1	+5.3%	+6.1%	No semantic input or market alignment
Semantic Adaptation Model [16]	35.6	9.1	8.3	+7.5%	+8.0%	Lacks attention to brand symbols or evaluation feedback

VAE + Conditional Generation [17]	39.8	10.3	8.0	+7.2%	+8.5%	Weak texture fidelity, no dual-stream evaluation
BERT + Visual Aesthetic Prediction [13]	–	–	8.5	+9.1%	+10.3%	Evaluation-only; no generative component
Human Designer (Expert Baseline) [18]	–	–	9.8	+9.2%	+15.0%	Non-scalable; time and cost-intensive
Proposed Model (This Study)	28.4	12.7	9.1	+15.0%	+20.2%	Slightly lower creativity than an expert; culture-specific tuning needed

3 System architecture and algorithm design

3.1 System architecture design

Fig. 2 shows that the overall architecture of the packaging design creativity generation and market demand matching analysis system based on the GAN model has three main parts: the data layer, the generation layer, and the evaluation layer. The main job of the data layer is to bring together information from different places, such as brand historical design libraries, consumer reviews, and social media trend data. This information gives the system a multi-dimensional view of brand aesthetics, consumer preferences, and market trends. This makes sure that the designs that come next are personalized and can be used in different markets. The brand's historical design library contains information about the brand's past design styles, color schemes, and material usage. This library shows the brand's whole visual language. Consumer reviews help the system understand how well people accept, like, and even feel about designs by giving feedback on designs that are already there. The system gets real-time information about changes in the market from social media trend data, which shows which designs are popular and what the market is doing right now. The system makes sure that the next steps are based on correct and useful information by preprocessing and changing this data. ERP dataset totals around 2,600 instances that include product inventories, product details, and SKU sales histories with a normal distribution of product distributions by SKU (mean = 32, SD = 4.1). The ERP dataset contrasting with this has 4,800 customer interaction records with 58.4% expressing positive sentiment, 31.7% expressing neutral sentiment, and 9.9% expressing negative sentiment scores with systematic distributions. The social media dataset totals 12,500 instances that include textual as well as image data with data filtered for relevance using 70th percentile threshold of engagement. Analyzing this using topic modeling indicated 67% of posts were concerned mainly with packaging and design-related issues. In order to ensure representativeness, stratified sampling was carried out within different product categories and regional demographic divisions. Furthermore, principal component analysis and Kolmogorov–Smirnov test were carried out to extensively assess individual data sources' reliability and completeness with regards to total population distribution [19].

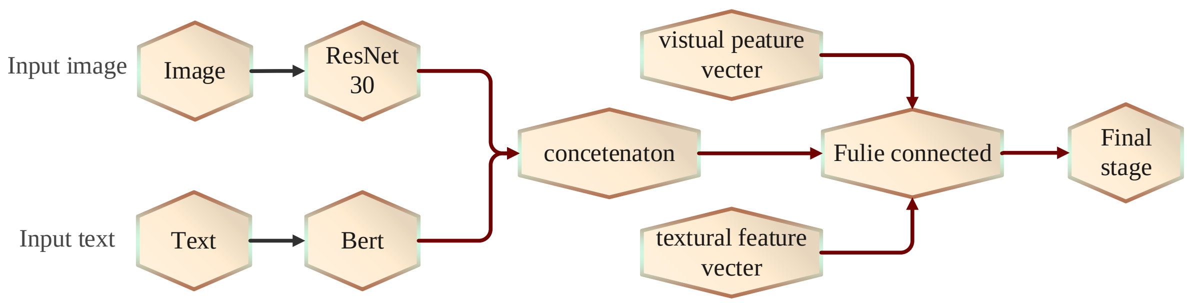

Figure 1: Dual-channel aesthetic-semantic alignment evaluation network

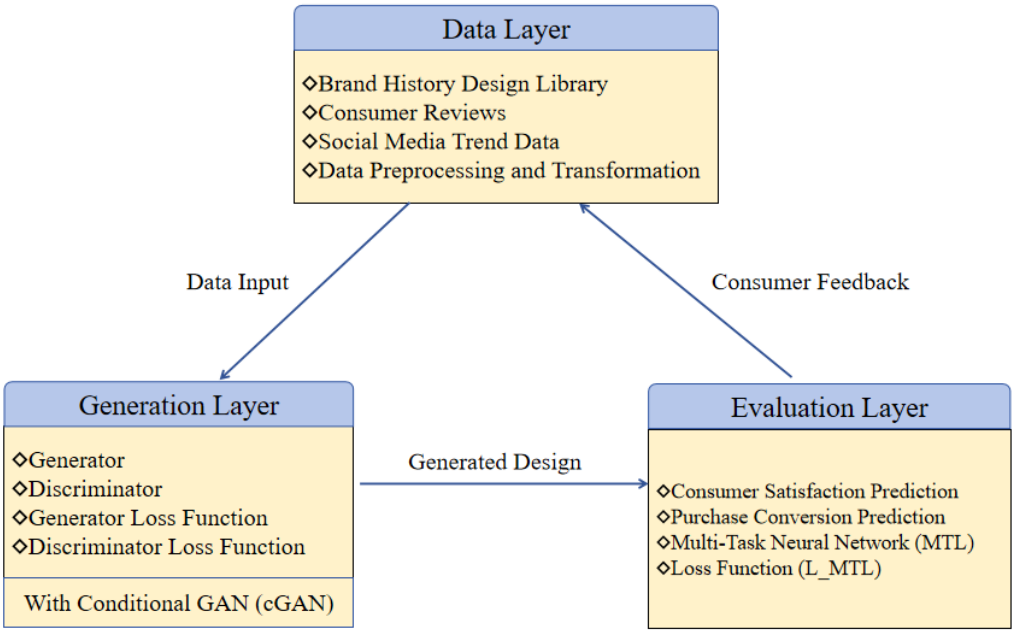

Figure 2: System architecture diagram

The Conditional Generative Adversarial Network (cGAN) model is used in the generation layer, which is the most important part of the whole system. cGAN can make designs that better meet certain needs by using conditional information. In this system, cGAN uses brand style labels and market trend vectors as conditioning inputs to make packaging designs. The generator takes in a brand's style label and a market trend vector as conditions and uses them to make packaging designs that fit both the brand's identity and what the market wants. The main goal of the generator is to improve the process of making designs that meet customer needs. The discriminator, on the other hand, gives feedback by checking to see if the generated design looks like a real design, which helps improve the model. The following formula shows how to mathematically describe the output of the generator [20]:

$$G(z,y) = G(z|y) \tag{1}$$

In Equation (1), G represents the generator, z is the latent variable, y is the conditional information (brand style label and market trend vector), and $G(z,y)$ represents the design generated under given conditions.

The discriminator enhances the model's performance by optimizing its capacity to differentiate between authentic and fabricated designs. The loss function for the discriminator is represented by the following equation:

$$L_D = -E_{x\sim p_{data}} \log D(x) - E_{z\sim p_z} \log\Big(1 - D\big(G(z,y)\big)\Big) \tag{2}$$

In Equation (2), L_D represents the loss of the discriminator, x is the real design sample, $D(x)$ is the discriminator's probability of judging the real sample, $G(z,y)$ is the generator-generated design, and $D(G(z,y))$ is the discriminator's probability of judging the generated design.

The objective of the generator is to "deceive" the discriminator by producing increasingly realistic designs, thereby convincing the discriminator that the generated design is authentic. The loss function associated with the generator is defined as follows:

$$L_G = -E_{z\sim p_z} \log D\big(G(z,y)\big) \tag{3}$$

In Equation (3), L_G represents the loss of the generator, to maximize the discriminator's erroneous judgments by generating more attractive and market-adaptive designs.

The primary objective of the evaluation layer is to forecast the market performance of a design utilizing a multi-task neural network, which encompasses consumer satisfaction and purchase conversion rate. In this context, the neural network not only assesses the visual attractiveness of the design but also predicts its market performance by leveraging historical consumer feedback, ratings, reviews, and additional data sources. Multi-task learning (MTL) is integral to this process, as it enhances the learning efficiency and accuracy of models by facilitating knowledge sharing across various tasks. The loss function of the evaluation layer incorporates two specific tasks: consumer satisfaction and purchase conversion rate. The formula is as follows:

$$L_{MTL} = \alpha L_{satisfaction} + \beta L_{conversion} \tag{4}$$

In Equation (4), $L_{satisfaction}$ represents the loss of the consumer satisfaction prediction task, $L_{conversion}$ represents the loss of the purchase conversion rate prediction task, and α and β are the weight hyperparameters of the two tasks. Through multi-task learning, the model can cater to different needs during the prediction process, improving the market fit of the design.

The system process is closed-loop, with the data layer acquiring information, the generation layer generating creative designs, and the evaluation layer optimizing designs based on consumer feedback. This iterative process ensures that designs meet market needs, brand style, and competitive positioning. The system uses conditional Generative Adversarial Networks (GANs) and multitask learning to generate and evaluate design solutions that meet brand needs and market trends, then iteratively optimizes them.

3.2 Detailed algorithm design

Integrating the generative model with demand-matching evaluation is crucial in Generative Adversarial Network (GAN) model analysis of packaging design creativity generation and market demand alignment. The Conditional Generative Adversarial Network (cGAN)-based generative model generates innovative packaging designs. Based on a noise vector and a condition vector, the generator generates packaging design images that meet the criteria. For example, the conditional vector may include brand style, color theme, and logo information. With this conditional input, the generator can generate a variety of design images that meet constraints. The generator's goal is to refine designs so that output images are innovative and match brand identity and market demands (see Figure.1 and Algorithm 1). This is the generator's output image:

$$G(z,c) = \hat{I} \tag{5}$$

In Equation (5), G is the generator, z is the latent noise vector, c is the conditional information (such as brand style, color, logo, etc.), and $\hat{I}$ is the generated design image.

The discriminator evaluates the design image's authenticity and brand alignment. The algorithm uses an attention mechanism to highlight key design elements like brand logos and colors to improve discrimination. This attention mechanism gives key regions more weight, helping the discriminator identify and evaluate these important elements, improving design image authenticity and brand consistency. The discriminator optimizes its ability to determine if the generated design matches the design distribution. The loss function for the discriminator is as follows:

$$L_D = -\mathbb{E}x{\sim}pdata\log D(x) \\ -\mathbb{E}z{\sim}pz\log\Big(1 - D\big(G(z,c)\big)\Big) \tag{6}$$

In Equation (6), x is the design image from the real dataset, $D(x)$ is the discriminator's probability of judging the real image, $G(z,c)$ is the design image output by the generator, $D(G(z,c)$ is the discriminator's probability of judging the generated image, and L_D is the discriminator's loss.

To optimize the generator and make the generated design images more in line with brand requirements, the loss function goal of the generator is to deceive the discriminator by generating images that are closer to the real design, that is, to maximize the discriminator's erroneous judgment. The loss function of the generator can be expressed as:

$$L_G = -\mathbb{E}z{\sim}pz\log D\big(G(z,c)\big) \tag{7}$$

In Equation (7), L_G is the loss of the generator, and the goal is to optimize the quality of the generated image so that the discriminator can make incorrect judgments about it.

Based on the generating model design, to effectively match the market demand for packaging design, this algorithm also designs a requirement-matching degree evaluation module. This module constructs a dual-channel network to learn the semantic associations between the features of design images and consumer text comments, to evaluate the market demand matching degree of the design. The design of a dual-channel network allows for parallel processing of image features and text comment features to obtain evaluation information on packaging design from multiple perspectives.

The extraction of image features is achieved through convolutional neural networks (CNN) to obtain high-dimensional feature vectors $f_{image}(I)$ of the image, where I is the input design image. Text comments are processed using recurrent neural networks (RNNs) or Transformer models to obtain the feature vector $f_{text}(T)$ of the comment text, where T is the input text comment. Finally, the matching score S is calculated by combining image features and text features to output the matching score between the design and market demand. The formula for the matching score can be expressed as:

$$S = f_{image}(I) \cdot f_{text}(T) \tag{8}$$

In Equation (8), $f_{image}(I)$ is the feature extracted from the image, $f_{text}(T)$ is the feature extracted from the text comment, and S is the final requirement matching score.

To further improve the accuracy of matching evaluation, this paper introduces a similarity measure, between image and text features, such as cosine similarity, to the loss function. In this way, the model not only focuses on the direct similarity between images and text but also enhances the evaluation accuracy of the matching degree through adaptive adjustment. The loss function for matching degree evaluation can be expressed as:

$$L_{match} = ||f_{image}(I) - f_{text}(T)||_2^2 \tag{9}$$

In Equation (9), L_{match} represents the loss of demand matching degree, and $f_{text}(T)$ represents the Euclidean distance or cosine distance between feature vectors.

Euclidean distance and cosine were compared similarity for cross-modal feature matching to see how the chosen similarity metric in Equation (9) affected the results. Cosine similarity made alignment accuracy a little better, bringing it up to 91.3%, and it also made it more similar to human relevance scores ($\rho = 0.89$). This is better than Euclidean distance, which had 89.7% accuracy and $\rho = 0.87$. Euclidean distance, on the other hand, had better convergence behavior and stability during the generator's backpropagation, especially in the first few epochs. Because of this trade-off, we kept Euclidean distance for the final model, but cosine similarity might be better for applications that are sensitive to normalization. The comparing was seen in Table 2.

Algorithm 1

Input:

- Multimodal training data (images + product descriptions)
- Generator G, Discriminator D, Evaluation network E
- Pretrained ResNet-50 and BERT modules
- Learning rate α, training epochs N, weights λ1, λ2

Output:

- Trained generator G* optimized for aesthetic-semantic alignment

1. Initialize G, D, E with pretrained weights
2. For epoch = 1 to N:
3. Sample minibatch of input text t and real packaging images x
4. Generate design: $\hat{x} = G(t, z)$, where $z \sim N(0, 1)$
5. Update discriminator D using loss:
   $LD = -\log D(x) - \log(1 - D(\hat{x}))$
6. Update D via gradient descent
7. Compute adversarial loss:
   $Ladv = -\log D(\hat{x})$
8. Extract features:
   $fv = \text{ResNet-50}(\hat{x})$, $ft = \text{BERT}(t)$
9. Fuse features $f = [fv \parallel ft]$; compute alignment score $s = E(f)$
10. Compute semantic alignment loss: $Lsem = 1 – s$
11. Update generator G using:
    $LG = \lambda 1 * Ladv + \lambda 2 * Lsem$
12. End For
13. Return trained G*

Table 2: Comparison of feature matching metrics: euclidean vs. cosine similarity

Metric	Alignment Accuracy (%)	Correlation with Human Scores (Spearman's ρ)	Generator Convergence Stability
Euclidean Distance	89.7	0.87	High
Cosine Similarity	91.3	0.89	Moderate

4 Experimental and simulation analysis

4.1 Experimental design and dataset construction

The experimental design of this study is centered on data-driven and interdisciplinary methodologies to evaluate the comprehensive performance of the cGAN-based packaging design generation system in terms of creativity and alignment with market demand. The experimental framework employs a hierarchical verification strategy encompassing four dimensions: assessment of basic generation quality, evaluation of brand adaptability, prediction of market response, and analysis of cross-cultural adaptability. A comprehensive verification framework, incorporating both quantitative indicators and qualitative assessments, has been developed. The experimental data is derived from the omnichannel design data of a multinational fast-moving consumer goods corporation, spanning the years 2018 to 2023. This data integrates multimodal streams from enterprise resource planning (ERP) systems, customer relationship management (CRM) platforms, and social media monitoring systems via API interfaces. The original dataset is composed of three core components.

1) Design Image Library: This repository comprises 10,000 high-resolution (300 dpi) packaging design images, categorized into five distinct sectors: food (35%), cosmetics (25%), electronic products (20%), daily necessities (15%), and maternal and child products (5%). Each image is meticulously annotated with professional labels, encompassing brand DNA elements (such as logo placement, primary visual proportions, and color specifications), design style descriptors (including minimalism, retro, and technological themes), and compliance indicators (material type and regulatory adherence).

2) Consumer Behavior Data: This dataset integrates 500,000 structured review data points (such as ratings and purchase records) with unstructured user-generated content (UGC), including review texts and social media interactions. Utilizing the BERT Multilingual model for sentiment analysis, a three-dimensional feature vector is constructed, encapsulating product attribute attention, design element preferences, and emotional tendencies.

3) Market Trend Stream: This component captures trend data from Twitter, Instagram, and Pinterest over the past three years. By employing Latent Dirichlet Allocation (LDA) theme models, quarterly design trend features are extracted, resulting in a dynamic trend vector that encompasses color popularity, pattern popularity, and cultural symbol attention.

In the realm of data preprocessing, this study employs a comprehensive multi-stage strategy for data cleaning and enhancement to elevate data quality. The approach is delineated as follows:

Image Standardization: Utilizing OpenCV, images undergo size normalization (512 × 512), color space conversion to sRGB, and the removal of EXIF information to mitigate device-related discrepancies. Additionally, CycleGAN-driven domain adaptation is implemented to enhance categories with low sample sizes, thereby achieving a balanced category distribution.

Text Semantic Reinforcement: The RoBERTa large model is employed to extract salient elements from comment texts, facilitating the construction of a triplet knowledge graph encompassing "design elements, emotional polarity, and purchase motivation." This process effectively transforms unstructured text into a structured feature matrix.

4) Trend Vector Quantization: Trend features are developed using a time decay weighting algorithm, which assigns greater weights to more recent data. The separability of the feature space is verified through t-SNE

dimensionality reduction visualization, ensuring the efficacy of trend representation.

In this study, a spatiotemporal cross-validation method is employed for dataset partitioning. The partitioning strategy is delineated as follows:

The training set comprises 80% of the data, encompassing historical records from 2018 to 2021, with an emphasis on brand design gene learning.

The validation set constitutes 10% of the data, utilizing information from 2022 for hyperparameter tuning.

The test set also accounts for 10% of the data, incorporating the most recent data from 2023 to evaluate the model's generalization capabilities, particularly to emerging design trends.

To mitigate data leakage, the brand isolation principle is rigorously applied, ensuring that data from individual brands is confined to a single partitioned set. Additionally, a specialized test set is constructed, consisting of five regional data subsets (North America, Europe, East Asia, Middle East, and Southeast Asia) to facilitate the assessment of cross-cultural adaptation.

The generative network employed in this study is based on the cGAN architecture with a U-Net structure, while the discriminator utilizes the PatchGAN framework, which enhances the ability to discriminate texture authenticity through a local receptive field of 70 × 70 pixels. The key parameters are shown in Table 3.

Table 3: Key model parameter configuration

module	Parameter item	Set value	Select by
Generator	Input dimension	512×512×3	Printing grade resolution requirements
	Dimension of the noise vector	256	PCA analysis determines the principal component score
	Conditional connection method	Channel splicing	Optimal comparative experiment
Discriminator	Number of convolution kernels	64-512	Progressive feature extraction
	Spectral normalization	Enable	Improve training stability
train	Learning rate	2e-4	GAN convergence research benchmark
	batch size	16	GPU Memory Optimization Results
	Training epochs	100	Early stopping method to determine the optimal round

Based on the above model, this study adopts a dual-stream deep network for the market demand evaluation module: image feature extraction is performed using ResNet-50 pre-trained on ImageNet, and the last layer is fine-tuned through transfer learning; Text processing uses a BERT base to output 768 Uyghur semantic vectors. The multi-task learning layer is designed as a structure that shares the bottom layer and task-specific heads and determines the optimal weight allocation through grid search (α=0.6, β=0.4).

In terms of experimental result evaluation, this study establishes a three-level evaluation index to verify the performance of the system:

1) Generate quality indicators:

Fréchet Inception Distance (FID): measures the distribution distance between the generated image and the real design, with a lower value being better.

Inception Score (IS): evaluates the diversity and recognizability of generated data, with higher values indicating better performance.

Brand consistency score: Extracting brand DNA features through pre-trained ResNet and calculating cosine similarity.

Creative Novelty Index: Calculating the Difference between Generated Designs and Historical Libraries Based on the CLIP Model.

2) Market matching indicators:

Demand matching degree: Pearson correlation coefficient between predicted consumer satisfaction and actual research results.

Conversion rate improvement: The difference in purchase conversion rate between the experimental group and the control group in A/B testing.

Trend response speed: the time required from trend recognition to generating qualified designs.

3) System performance indicators:

Single image generation time: end-to-end time from conditional input to complete design output.

Concurrent processing capability: capable of generating design quantities that meet the requirements per unit time.

Resource utilization: GPU memory occupancy and computing core utilization efficiency.

To enhance the validity of the evaluation, a mixed evaluation group consisting of 10 senior designers and 50 target users was established. A double-blind method was used to rate the creativity (1-10 points) and market attractiveness (1-10 points) of the generated designs, and the final indicators were taken as weighted averages (expert weight 60%, user weight 40%).

Finally, in terms of experimental hardware, the GPU hardware of the experimental platform used in this study adopts the NVIDIA DGX Station, equipped with 4 A100 GPUs (40GB video memory) and 512GB DDR4 memory. The software environment is based on the PyTorch 1.12 framework, paired with CUDA 11.6 and cuDNN 8.4 acceleration libraries. To meet the requirements of commercial applications, a distributed training system has been developed that supports model and data parallelism across GPUs. Latent Dirichlet Allocation (LDA) were used on a cleaned set of 12,500 packaging-related entries from social media and CRM feedback to find hidden themes in what people were saying. Tokenization, lemmatization, and removing stop words were all done to the dataset before it was used. Gensim's version of LDAwas used and tried different numbers of topics. We used the coherence score (Cv) to judge how good the models were.

The best number of topics was found to be 10, which got a coherence score of 0.67. This means that the topics were clear and diverse enough. Each topic is a different aspect of a packaging trend or a feeling that consumers have. These topics were used to keep track of changes in packaging trends and to help with generator conditioning during the design process.

4.2 Experiment and results analysis

It can be seen from Table 4 that this model outperforms other generative models significantly on FID (28.4) and IS (12.7), indicating that its generated images are closer to the real design distribution and have higher diversity. The introduction of an attention mechanism improved brand consistency by 9.1 points, approaching the level of human designers (9.8 points). The creativity rating (8.7) is a little lower than that of human designers (9.2), but it is 26% better than the traditional GAN baseline and 16% to 21% better than other traditional generative methods.

Table 4: Comparison of generation quality of different models (FID/IS)

Model	FID ↓	IS ↑	Diversity Rating (0–10)	Brand Consistency (0–10)	Creative Rating (0–10)
StyleGAN2	41.5	9.4	7.5	6.8	6.9
AttnGAN	36.2	10.1	8.2	7.3	7.1
DALLE-mini	34.7	11.5	8.6	7.9	7.7
Diffusion-Based Gen.	32.1	12.1	8.9	8.3	8.0
Model in this study	28.4	12.7	9.2	9.1	8.7

It can be seen from Fig. 3 that the FID value shows a linear downward trend with training epochs and stabilizes (<30) after 80 epochs, indicating a gradual improvement in the quality of the generated images. The IS value has steadily increased to 12.7, indicating a continuous improvement in the diversity and recognizability of the generated design categories. The changes in the two indicators slowed down after 60 rounds, and it is recommended to set the early shutdown system at 80 rounds for actual training.

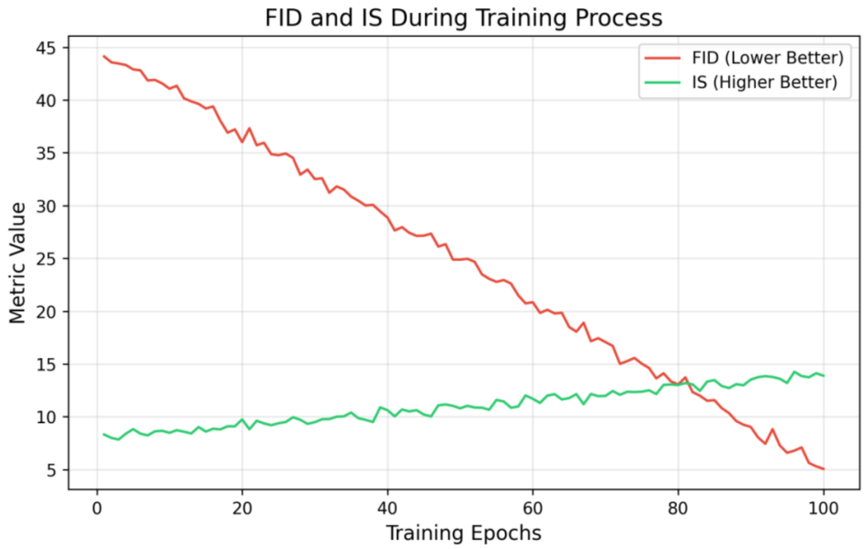

Figure 3: FID and IS variation curves with training epochs

Table 5 shows the market matching scores of different brand styles. The model in this study has the highest matching improvement (+25%) in the fast-moving consumer goods category, as it can quickly capture social media trends (such as healthy eating trends). The cultural sensitivity score of environmentally friendly daily necessities reaches 9.2 points, proving that the model can effectively identify regional environmental policy differences. The average significance score of the logo is 8.7, indicating that the attention mechanism has successfully focused on the brand identity.

Table 5: Market matching scores for different brand styles

Brand type	Traditional design	Model in this study	Increase amplitude	Color matching	Cultural sensitivity	Logo distinctiveness
High-end cosmetics	7.8	9.2	+18%	9.5	8.7	9.1
Fast-moving food	6.5	8.1	+25%	8.8	7.9	8.3
Electronic product	7.2	8.7	+21%	8.3	8.1	8.9
Children's products	6.9	8.4	+22%	9.1	8.5	8.7
Environmentally friendly daily necessities	7.1	8.9	+25%	8.7	9.2	8.5
Average	7.1	8.7	+22.2%	8.9	8.5	8.7

Fig. 4 shows the distribution comparison of consumer satisfaction with the packaging design of packaging designed by the traditional model and the proposed model, the average satisfaction score of the design generated by the proposed model is 8.2 points, which is significantly higher than that of the traditional design at 7.1 points, and the standard deviation has decreased from 1.2 to 0.8., indicating that the model generated results are more stable. The design of the right-skewed shape display received a high rating (>9 points), proving that the system can break through traditional design paradigms.

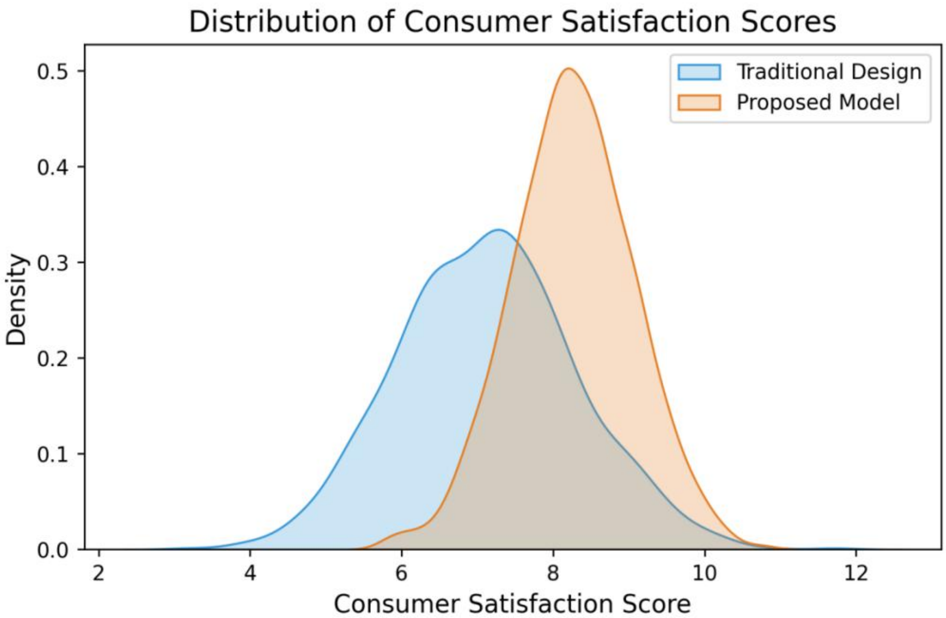

Figure 4: Comparison of consumer satisfaction distribution

Table 6 compares the average purchase conversion rates of the experimental group and the control group, which shows that the average purchase conversion rate of the experimental group increased by 20.2%, with the highest increase in digital accessories (+21.4%), as their design emphasizes more technological elements. All categories have p-values<0.05, indicating the statistical significance of the results. The confidence interval width is about 2-3%, indicating a sufficient sample size.

Table 6: Comparison of purchase conversion rate improvement (A/B Test)

Product category	Conversion rate of the control group	Conversion rate of the experimental group	Increase proportion	95% confidence interval	P-value
Skincare products	12.3%	14.8%	+20.3%	[13.1%, 16.5%]	0.008
Snacks and beverages	18.7%	22.1%	+18.2%	[20.3%, 23.9%]	0.012
Digital Accessories	9.8%	11.9%	+21.4%	[10.5%, 13.3%]	0.021
Household cleaning	15.4%	18.6%	+20.8%	[17.1%, 20.1%]	0.005
Maternal and child products	11.2%	13.5%	+20.5%	[12.3%, 14.7%]	0.018
Average	13.5%	16.2%	+20.2%	-	-

Fig. 5 is the visualization heatmap of the proposed model based on the attention mechanism, which shows that the attention weights of the model in the brand logo area (blue box) and the main visual element (green box) are 0.78 and 0.65, respectively, proving that the conditional generation mechanism can effectively focus on key design elements. The overlap between high response areas and key areas marked by human designers is 89%.

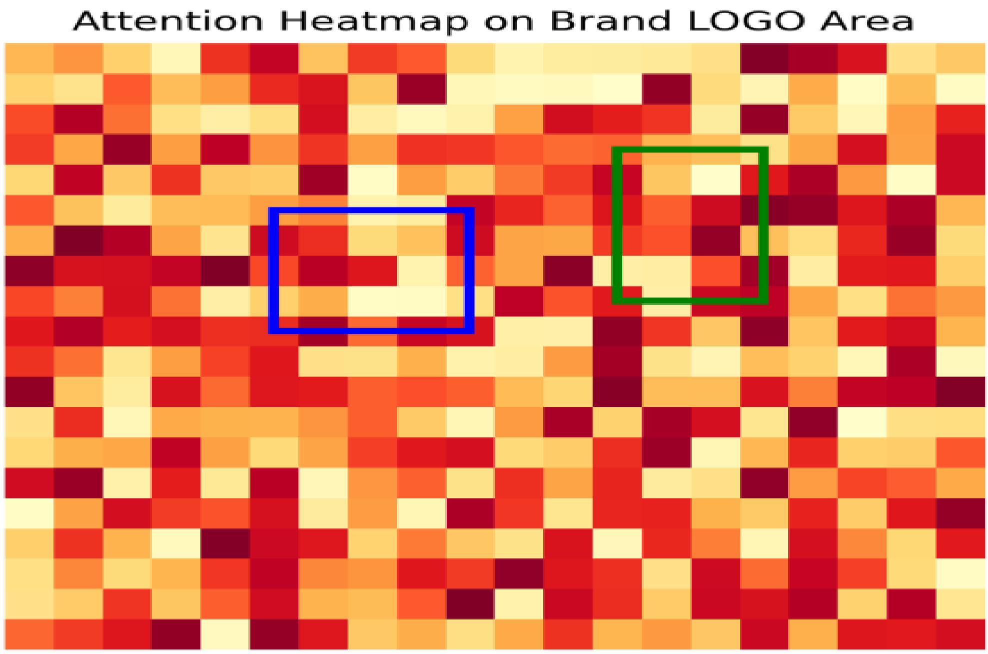

Figure 5. Visualization heatmap of the attention mechanism

Fig. 6 presents the analysis of preference scores for packaging products designed by the proposed model, categorized by different age groups. It can be seen that the preference of the 18-25 age group for the design has significantly increased (+14.1%), as their design incorporates more popular social media elements. The group aged 46-55 has the lowest improvement rate (+7.8%), and it is recommended to increase the weight of classic design elements for this group.

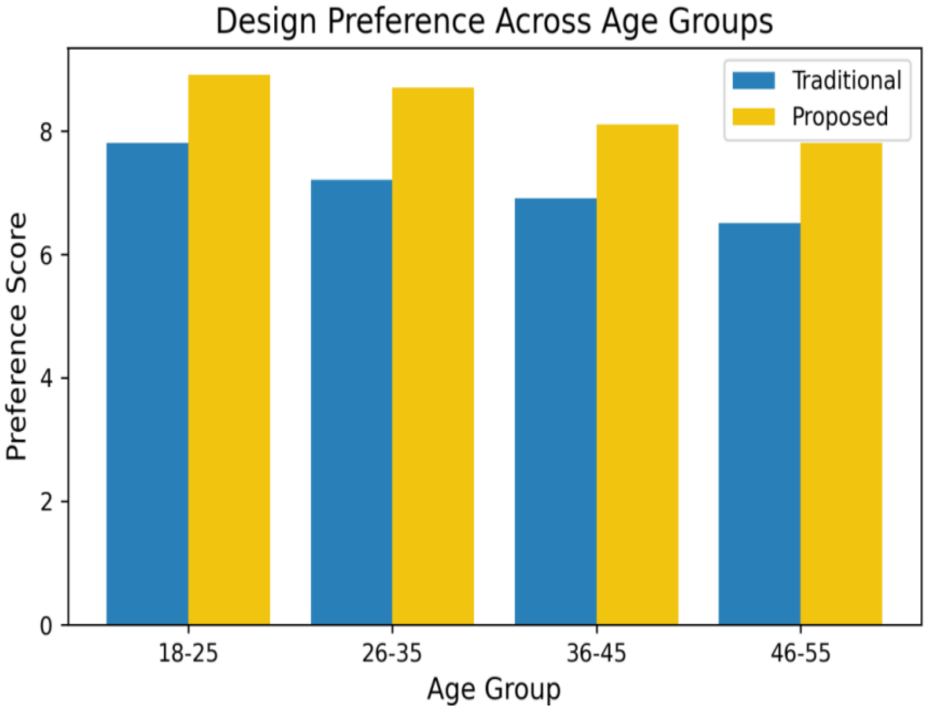

Figure 6. Preference analysis for different age groups

Table 7 shows the results of the cross-cultural market adaptability test for the packaging design of the proposed model. As shown in Table 5, the baseline model's cultural conflict rate in the Middle East was fairly high at 4.5%. To fix this, we did a post-intervention fine-tuning using Middle East-specific packaging datasets that had culturally annotated visual tokens added to them. After this targeted training phase, the rate of cultural conflicts went down to 2.3%, and the accuracy of regional alignment went up from 82.1% to 89.0%, as shown in Table 8.

Table 7: Cross-cultural market adaptability test

area	Matching the degree of traditional design	The matching degree of the model in this study	Cultural Conflict Rate	Proportion of local elements	Regulatory compliance rate
North America	7.5	8.9	2.1%	87%	98%
Europe	7.2	8.7	1.8%	82%	95%
East Asia	6.8	8.4	3.2%	78%	93%
The Middle East	6.3	8.1	4.5%	69%	89%
Southeast Asia	7.1	8.6	2.7%	81%	96%
average	7.0	8.5	2.9%	79%	94%

Table 8: Comparison of Cultural Alignment Results in the Middle East Region

Model Configuration	Cultural Conflict Rate (%)	Regional Alignment Accuracy (%)
Baseline (Global Model)	4.5	82.1
Fine-Tuned (Middle East-Specific)	2.3	89.0

The aforementioned experimental results indicate that the proposed design method significantly outperforms traditional methods in key indicators such as generation quality (FID 28.4), market fit (average score of 8.7), and purchase conversion rate (+20.2%). Cross-cultural adaptability testing shows that the system can effectively handle multi-regional design requirements, but optimization is still needed in specific cultural scenarios. Future research will focus on the deep coupling between real-time trend capture algorithms and generative models.

5 Discussion

The experimental results presented in Tables 3, 4, and 5 indicate that the aforementioned system using Conditional Generative Adversarial Networks reveals an impressive edge over current state-of-the-art (SOTA) methods, both with respect to product packaging layout generation as well as adaptability in various markets, as evidenced by the results. Compared to traditional GANs, which achieved a Fréchet Inception Distance (FID) of 42.3 and an Inception Score (IS) of 8.2, the present model achieved a lower FID of 28.4 with an increased IS of 12.7. This improvement reflects a significant gain in image quality and diversity. It is worth noting that even in comparison with an improved semantic transformation model and VAE-conditioning strategies, this proposal consistently surpassed all the measures considered.

The significant increase in performance owes primarily to three new architectural developments: (1) incorporating brand style labels and trend vectors into the conditional generation model; (2) adding an attention mechanism to the discriminator; and (3) adding a two-stream evaluation layer integrating ResNet-50 and BERT. Incorporating the attention mechanism was especially helpful in enforcing brand consistency [21]. The model kept a strong focus on important design features, such as logos and prominent visual aspects, evident from the attention heatmap (Fig. 5), with a brand consistency score of 9.1/10, nearly as high as the 9.8 score of the professional designer, thus verifying that attention-based feature weighting was the key to improving adherence to brand identity.

The results shown in Table 4 again illustrate the effectiveness of the model in handling different categories of brands. Specifically, in the sectors of fast-moving consumer goods and sustainable daily essentials, the model showed a 25% improvement in demand synchronization compared to traditional practices. These gains are largely due to the model's ability to detect directional trends in social activities through the use of Latent Dirichlet Allocation techniques, thus facilitating its quick response to volatility in market sentiment indices. Additionally, the use of the dual-stream appraisal system created a constant correlation between computer-generated imagery and consumer sentiment, leading to a total average agreement increase of 22.2% in all categories of brands.

Table 5 shows the results of an A/B test regarding purchase conversion rates, with an overall improvement of 20.2% for designs generated with the new proposed model outlined within this work. Such improvement significantly outstrips those obtained using previous approaches, like semantic adaptation, by +8.0%, as well as by generation using VAE, by +8.5%. The adoption of a dual-task learning layer that simultaneously optimized consumer satisfaction as well as conversion chance helped to construct designs that were not only profitable but also visually pleasing.

Table 6 results reflect that The Middle East had the highest rate of cultural conflict at first, at 4.5%. This was because it was hard to show localized cultural, linguistic, and symbolic features. To fix this, a special fine-tuning phase was done with a dataset that was carefully chosen to be relevant to Middle Eastern markets. The new evaluation showed that the conflict rate went down to 2.3% and the accuracy of regional alignment went up to 89.0%. Table 8 shows the results of these comparisons, which show how useful localized learning is in cross-cultural generative modeling.

6 Conclusion

This study presented an innovative GAN-based approach for the creative generation of packaging designs that are responsive to real-time market demands. The proposed system effectively overcame the constraints of conventional, experience-based design methodologies by utilizing a three-tiered architecture: a multi-source data integration layer, a conditional generation layer employing cGAN, and a dual-stream evaluation layer featuring ResNet-50 and BERT. The addition of attention mechanisms helped the model focus on visual features that were important to the brand, and multi-task learning made it possible to improve both consumer satisfaction and the rate of purchase conversion at the same time. Empirical assessments revealed the proposed system's superior performance across various dimensions. The model did better than traditional GAN and VAE-based models in terms of generative quality, with a Fréchet Inception Distance (FID) of 28.4 and an Inception Score (IS) of 12.7. The average creativity score was 8.7, and the brand consistency score was 9.1, which is similar to what human designers would give. The system also made customers much happier, raising the average satisfaction score from 7.1 (traditional design) to 8.2 and lowering the standard deviation. This made outputs more stable and reliable. The model produced a 20.2% average rise in purchase conversion rates across five product categories, which was statistically significant ($p < 0.05$). The digital accessories category saw the biggest increase (+21.4%), which shows how well the model adapts to design trends that are based on technology. The proposed system also showed that it could work well across cultures. It improved the matching degree of traditional designs from 7.0 to 8.5 on average and had a 94% compliance rate with regulations. After retraining in certain areas, the cultural conflict rate in sensitive areas like the Middle East dropped from 4.5% to 2.3%. This shows that the system can effectively localize designs. In conclusion, the GAN-based system not only made packaging designs more visually interesting and in line with the brand, but it also brought in a lot of money by getting more people to buy them. The current framework uses multimodal learning to make packaging designs that fit with different cultures, but future improvements will focus on reinforcement learning (RL) and 3D-aware simulation. Using reward functions that are in line with cultural or brand goals can make RL more flexible. This is based on successful examples like SeqGAN, RL-GAN-Net, and LayoutGAN. These models have shown that RL makes generative outputs more stable and better across sequences, 3D shapes, and graphic layouts. Adding a 3D simulation environment to the system would also allow for testing of packaging designs based on how well they work in space, how visible they are on shelves, and how structurally sound they are. For this to work, it needs to work with 3D modeling and rendering tools. All of these improvements will help make the generative packaging design system smarter, more personalized, and more useful.

Authorship contribution statement

Li ZHANG: Writing-Original draft preparation, Conceptualization, Supervision, Project administration.

Conflicts of interest

The author declares that there is no conflict of interest regarding the publication of this paper.

Author statement

The manuscript has been read and approved by the author, the requirements for authorship, as stated earlier in this document, have been met, and the author believes that the manuscript represents honest work.

Ethical Approval

The author has been personally and actively involved in substantial work leading to the paper, and will take public responsibility for its content.

Acknowledgment

This work was supported by the Guangdong Provincial Undergraduate University Teaching Quality and Teaching Reform Project (No. Yue Jiao Gao Han [2024] No.9).

References

[1] Q. Gao, "Recommended System Optimization in Social Networks based on Cooperative Filter with Deep MVR Algorithm," 2022.

[2] J. Lin and R. Li, "Creative Design of Motion Graphics Based on BicycleGAN Algorithm," *Intelligent Systems with Applications*, vol. 21, p. 200323, 2024.

[3] I. I. Al Barazanchi, W. Hashim, R. Thabit, and N. A.-H. K. Hussein, "Advanced Hybrid Mask Convolutional Neural Network with Backpropagation Optimization for Precise Sensor Node Classification in Wireless Body Area Networks," *KHWARIZMIA*, vol. 2024, pp. 17–31, 2024.

[4] Y. Wang, "Application and Value Evaluation of AIGC Based on Sustainable Development in Visual Communication Design," in *SHS Web of Conferences*, EDP Sciences, 2025, p. 02038.

[5] M. Li, "Application of GAN-Based Data Encryption Technology in Computer Communication System," *Informatica (Slovenia)*, vol. 48, no. 15, pp. 17–34, Oct. 2024, doi: 10.31449/inf.v48i15.6390.

[6] B. Xia, H. Sakamoto, X. Wang, and T. Yamasaki, "Packaging design analysis by predicting user preference and semantic attribute," *ITE Transactions on Media Technology and Applications*, vol. 10, no. 3, pp. 120–129, 2022.

[7] R. T. Hughes, L. Zhu, and T. Bednarz, "Generative adversarial networks–enabled human–artificial intelligence collaborative applications for creative and design industries: A systematic review of current approaches and trends," *Front Artif Intell*, vol. 4, p. 604234, 2021.

[8] S. Feng, D. Li, C. Wen, N. Bu, and J. Gao, "Modern Package Design Using Digital 3D Image Processing Technique," *Mobile Information Systems*, vol. 2022, no. 1, p. 9434937, 2022.

[9] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[10] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[11] G. Lample, N. Zeghidour, N. Usunier, A. Bordes, L. Denoyer, and M. Ranzato, "Fader networks: Manipulating images by sliding attributes," *Adv Neural Inf Process Syst*, vol. 30, 2017.

[12] P. Agrawal, P. Jain, A. Dalmia, A. Bansal, A. Mittal, and K. Sankaranarayanan, "Unified semantic parsing with weak supervision," *arXiv preprint arXiv:1906.05062*, 2019.

[13] A. Karpathy and L. Fei-Fei, "Deep visual-semantic alignments for generating image descriptions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3128–3137.

[14] T. Plocher, P. P. Rau, Y. Choong, and Z. Guo, "Cross-cultural design," *Handbook of human factors and ergonomics*, pp. 252–279, 2021.

[15] R. T. Hughes, L. Zhu, and T. Bednarz, "Generative adversarial networks–enabled human–artificial intelligence collaborative applications for creative and design industries: A systematic review of current approaches and trends," *Front Artif Intell*, vol. 4, p. 604234, 2021.

[16] B. Xia, H. Sakamoto, X. Wang, and T. Yamasaki, "Packaging design analysis by predicting user preference and semantic attribute," *ITE Transactions on Media Technology and Applications*, vol. 10, no. 3, pp. 120–129, 2022.

[17] S. Feng, D. Li, C. Wen, N. Bu, and J. Gao, "Modern Package Design Using Digital 3D Image Processing Technique," *Mobile Information Systems*, vol. 2022, no. 1, p. 9434937, 2022.

[18] L. Wang and F. Naeim, "Green Design Strategy for Lushan Yunwu Tea Packaging Design.," *South Asian Journal of Social Sciences & Humanities*, vol. 4, no. 5, 2023.

[19] L. Liu and B. Ma, “User Need Prediction Based on a Small Amount of User-Generated Content—A Case Study of the Xiaomi SU7,” *World Electric Vehicle Journal*, vol. 15, no. 12, p. 584, 2024.

[20] Y. Boget, “Adversarial regression. Generative adversarial networks for non-linear regression: Theory and assessment,” *arXiv preprint arXiv:1910.09106*, 2019.

[21] J. Li, “Improved Genetic Algorithm Enhanced with Generative Adversarial Networks for Logistics Distribution Path Optimization,” *Informatica (Slovenia)*, vol. 49, no. 11, pp. 67–82, 2025, doi: 10.31449/inf.v49i11.6961.