Develop A Comprehensive Product Design Framework Based On Artificial Intelligence Visual Enhancement Technology To Drive Innovation And Efficiency

Shuang Liang* and Yuanyuan Jiang

Chengdu College of University of Electronic Science and Technology of China, Chengdu 611731, China

Corresponding author. E-mail: liangxm58@gmail.com

Received: May 27, 2026; Accepted: August 17, 2026

Artificial Intelligence (AI) is transforming product design by automating workflows, improving decision-making, and enhancing visual quality. Traditional manual design methods are often time-consuming and lack the flexibility required for rapid, high-quality product development. This study proposes an AI-based product design framework that integrates Vision Transformers (ViT) for image enhancement and Generative Adversarial Networks (GANs) for design generation. The framework combines automated visual optimization with creative design synthesis to support innovative and efficient product development. High-quality image datasets collected from Google Images and open-source repositories were used to train and evaluate the model. ViT was applied to tasks such as super-resolution and denoising, while GANs generated realistic and diverse product designs. Model performance was evaluated using Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), Root Mean Squared Error (RMSE), Inception Score (IS), and Frechet Inception Distance (FID). Experimental results demonstrated notable improvements in image quality, with PSNR values ranging from 14-17 dB, SSIM from 0.72-0.87, and RMSE reduced from 54 to 40. GAN-generated outputs also achieved improved realism and diversity through higher IS and lower FID values. The proposed framework provides a scalable, efficient, and innovative solution for AI-driven product design and visual enhancement.

Keywords: Artificial Intelligence; Vision Transformers; Generative Adversarial Networks; Product Design; Visual Enhancement

http://dx.doi.org/10.6180/jase.202612_35.029

1. Introduction

The rapid advancement of AI has significantly transformed modern product design by enabling automation, precision, and intelligent decision-making. AI-based visual enhancement techniques improve the usability and quality of design outputs across various industries [1]. Traditional design approaches often depend on manual processes, making them time-consuming and inefficient. The increasing demand for innovative and high-quality designs requires intelligent systems capable of enhancing visual data and generating new concepts. AI-based methods analyze large-scale visual information, extract meaningful patterns, and support effective design decisions [2]. Deep learning models further enhance the automation of complex design tasks [3]. However, existing autoencoder-based approaches may produce blurred outputs, while most current systems focus separately on either visual enhancement or design generation [4]. This limitation restricts their ability to simultaneously improve quality and creativity, highlighting the need for an integrated AI-based product design framework.

The proposed framework addresses the limitations of existing systems by integrating ViT and GANs into a unified design pipeline. ViT captures global image relationships and improves visual quality through super-resolution and denoising tasks. GANs generate realistic and diverse product designs, enhancing creativity and innovation. The combination of ViT and GANs enables simultaneous visual enhancement and design generation within a single framework. Preprocessing optimization further improves system performance and scalability. This novel transformer-based enhancement and adversarial generation approach provides an effective solution for AI-assisted product design by improving visual quality, reducing limitations of existing methods, and supporting innovative design development.

1.1. Research Objectives

- Build a complete AI-powered product design system that improves the quality of visuals and enhance innovation and efficiency.
- Use the image datasets obtained through Google Images and open-source sites to train and test the proposed system.
- Use ViT models to solve enhancement problems like super-resolution and denoising.
- Use GANs to produce realistic and varied outputs of product designs.

1.2. Research Scope and Novelty

The proposed ViT-GAN framework enhances images and generates product designs, supporting fashion, industrial, and interface design through improved visual quality, automated design generation, and reduced design time. Its novelty lies in combining enhancement and generation within a unified framework. The use of GANs and transformer architectures improves efficiency, scalability, and output quality for product design applications.

2. Literature review

The study by [5] dwells upon the concept of artificial intelligence as used in redesigning processes in design spheres in architecture, graphic design, and product development. The study uses a qualitative method in examining AI as a creative partner in boosting productivity and innovation. Nonetheless, the paper raises ethical issues such as bias in AI systems and intellectual property concerns, which restrict its practical implementation in actual design systems. The study by [6] explores product design optimization through a hybrid method combining Genetic Algorithms and CNNs, integrating evolutionary computation with deep learning to improve convergence speed and design efficiency, with case studies demonstrating enhanced design efficiency, reduced costs, and faster time-to-market. The study by [7] investigates the implementation of artificial intelligence in bioclimatic building design based on digital twins and a machine learning approach. The approach is based on monitoring, simulation, and optimization of sustainable architectural design in real-time. Nonetheless, the framework is limited by excessive computational requirements and interoperability issues between AI systems and design platforms. The study by [8] examines AI-driven industrial product design for sustainability through a systematic review, developing a multidimensional model based on technological, systemic, and institutional perspectives. However, it highlights limitations such as insufficient system integration and the absence of standardized AI implementation strategies in industrial design. The study by [9] explores interactive product design within the metaverse using AI technologies to support user-centered innovation and enhance personalized design development.

AI-based product design frameworks face challenges in computational complexity, scalability, flexibility, and practical deployment. Existing CNN-, GAN-, Genetic Algorithm-, PSO-, and digital-twin-based methods often address image enhancement or design generation independently, limiting overall performance. The proposed framework integrates Vision Transformers (ViT) with GANs to unify enhancement and design generation, improving global feature representation, image fidelity, design diversity, processing efficiency, and supporting scalable, real-time, user-adaptive product design across diverse application domains.

3. Proposed integrated vision transformer and gan-based product design methodology

The framework's workflow includes image data collection, followed by ViT-based enhancement and GAN-based design generation tasks, including translation and style transfer. Pix2Pix was selected for paired image-to-image translation because it effectively learns mappings between aligned input-target image pairs, whereas CycleGAN was employed for unpaired image translation through cycle-consistency learning. The complementary capabilities of both models enable the framework to support diverse product design scenarios with varying dataset characteristics. Evaluation uses PSNR, SSIM, RMSE, FID, IS, and Perceptual Loss to assess image quality and creative design performance.

Fig. 1 presents the AI-based design framework involving data collection, preprocessing, ViT and GAN applications, and evaluation using PSNR, SSIM, and FID metrics.

3.1. Vision Transformer for Image Enhancement in Product Design

ViT uses transformer-based self-attention on image patches with positional embeddings, enabling global context learning compared to CNN-based local feature extraction. The Vision Transformer employs 16×16 patch embeddings with learnable positional encoding, 12 multi-head self-attention heads, and 12 transformer encoder layers. The CLS token is used for feature aggregation, enabling effective global context modeling and enhanced image reconstruction during the image enhancement stage. The extracted global feature representations are subsequently transferred to the GAN generator to guide image-to-image translation, while the PatchGAN discriminator evaluates local image realism through adversarial learning, forming an integrated architecture for image enhancement and product design generation. The self-attention mechanism in ViT is represented in Eq. (1):

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1}$$

Q, K and V represent query, key, and value matrices, enabling scaled attention to learn global relationships among image patches. The global self-attention mechanism enables the Vision Transformer to capture global contextual features by modeling long-range spatial and semantic relationships among all image patches simultaneously. Unlike CNNs that primarily focus on local neighborhoods, these feature representations preserve image-wide contextual information, resulting in more accurate image reconstruction, reduced noise, and enhanced visual quality. ViT encoder layers learn image representations for enhancement tasks, minimizing reconstruction loss to improve resolution, reduce noise, and enhance visual quality in product design. Enhanced feature representations extracted by the Vision Transformer are transferred to the GAN generator as conditional inputs through encoded feature embeddings. This establishes a sequential framework in which image enhancement precedes design generation. These representations preserve global contextual information, structural consistency, and fine visual detail during image-to-image translation. This enables the GAN to generate realistic product designs with improved representation learning, design generation efficiency, and overall visual fidelity through the complementary integration of both models.

Fig. 2 illustrates ViT architecture, where image patches are processed through Transformer Encoder, MSA, and MLP Head for feature handling and enhancement tasks.

The GAN consists of Generator $G(z)$ and Discriminator $D(x)$, competing in a minimax game to produce realistic images, as represented in Eq. (2).

$$\begin{aligned}\min_G \max_D V(D, G) &= \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x)] \\ &+ \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))]\end{aligned} \tag{2}$$

GANs use adversarial training to generate realistic images that are indistinguishable from real images, making them effective for innovative design applications. Feature matching loss has been widely recognized for improving GAN training stability by encouraging the generator to learn consistent intermediate feature representations. The proposed framework efficiently generates realistic product designs and demonstrates strong generalization across diverse product design domains through robust visual feature learning. GANs generate diverse product designs and refine existing samples by learning patterns and distributions from training data.

Fig. 3 illustrates the GAN data flow, where the Generator creates fake images and the Discriminator evaluates them against real images, enabling both models to improve iteratively.

4. Experimental setup

4.1. Description of the Dataset

The proposed framework uses paired Pix2Pix datasets [10] for training, validation, and evaluation, enabling ViT-based image enhancement and GAN-based design generation for image-to-image translation across diverse domains.

Table 1 presents the category-wise dataset analysis across different Pix2Pix domains. The reported complexity levels, visual diversity, and structural variability indicate that the proposed framework was evaluated using heterogeneous datasets with varying image characteristics, thereby demonstrating its capability to generalize across diverse product design tasks.

4.2. Data Preprocessing Steps

Image Resizing: All images are scaled to the same size to make them consistent and to fit the input size required by the model.

Normalization: Pixel values are normalized to the range $[0, 1]$ (or $[-1, 1]$, depending on model requirements) to standardize input scale.

Data Augmentation: Various transformations (rotation, flipping, zoom, and other techniques) are used to artificially increase the size of the dataset and enhance model generalization.

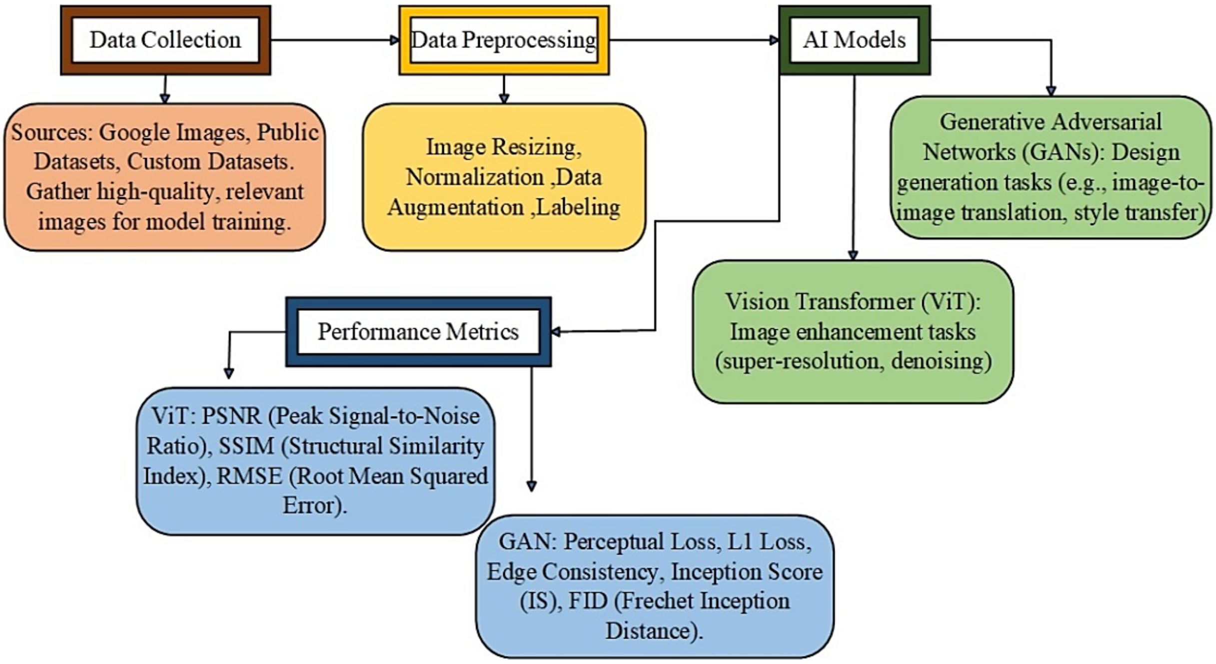

Fig. 1. Methodology Workflow of the Proposed Framework.

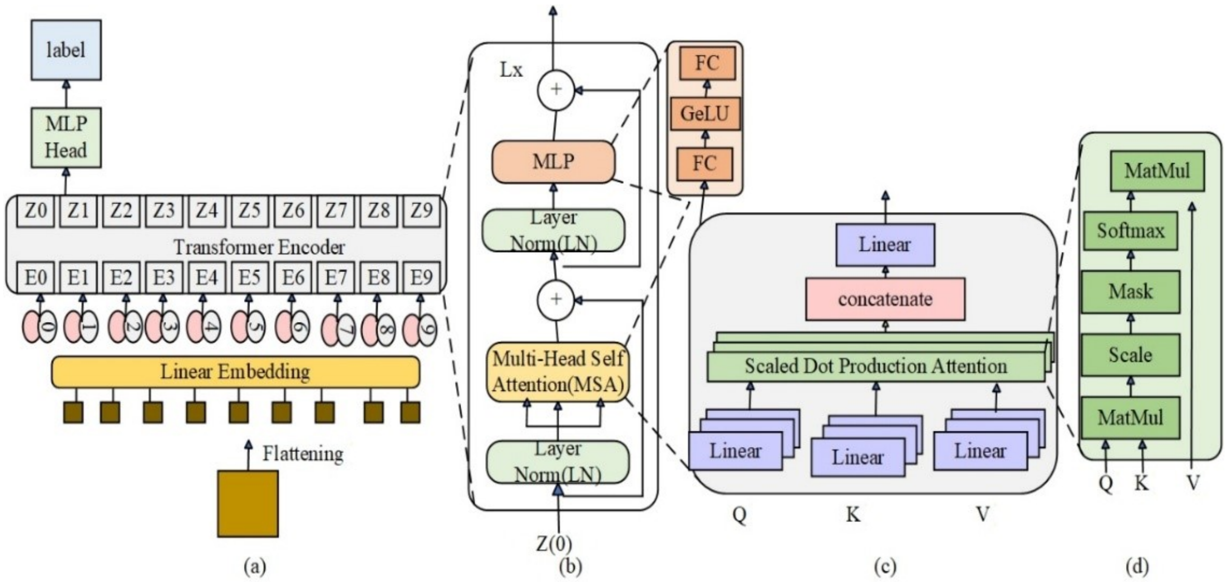

Fig. 2. ViT Architecture.

Labelling: For image-to-image translation, paired data are labeled as Original Image → Input (photograph or low-resolution image) and Edge Image → Target (edge map or final design). Image preprocessing involves resizing all input images to 224×224 pixels, normalizing pixel values to the range $[0, 1]$, and applying random horizontal flipping, $\pm 15^\circ$ rotation, and 0.8-1.2 zoom scaling for data augmentation. Edge images are generated using the Canny edge detector to preserve structural boundaries, and input-target image pairs are verified for correct alignment before training, ensuring preprocessing consistency and improved model generalization for image enhancement and design generation.

4.3. Software and Hardware Requirements

Table 2 summarizes the implementation environment, including Intel Core i9, 32 GB RAM, Windows 11, Python 3.11.7, and TensorFlow 2.20.0 for efficient training.

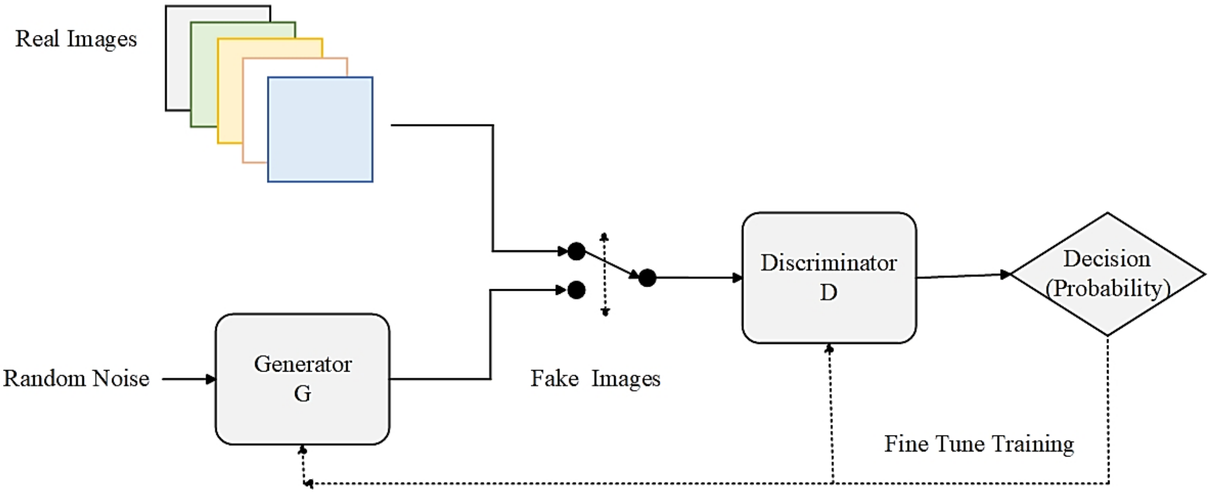

Fig. 3. GAN Architecture Overview.

Table 1. Per-Category Dataset Complexity and Diversity Analysis

Domain	Image Pairs	Complexity Level	Visual Diversity (LPIPS)	Structural Variability (1−Mean SSIM)
Edges → Shoes	12,000	High	0.384	0.271
Edges → Handbags	12,000	High	0.356	0.248
Facades → Photos	400	Medium–High	0.291	0.196
Maps → Aerial	1,096	High	0.418	0.303
Cityscapes Labels → Photos	2,975	Very High	0.447	0.327

Table 2. System Specifications for the Proposed Framework

Specification	Value
Processor	Intel Core i9-14900
RAM	32 GB
Operating System	Windows 11 (64-bit)
Python	3.11.7
TensorFlow	2.20.0

4.4. Hyperparameter Configuration

Table 3 summarizes the proposed framework's architecture, where a single-channel edge image is used as input to generate a three-channel RGB image.

Table 3. Architecture Details of the Proposed Framework

Layer	Configuration
Input	1-channel Edge
Output	3-channel RGB
Down/Up Sampling	Conv2D (s=2) / Bilinear + Conv
Activation	Leaky ReLU, ReLU
Output Activation	Tanh
Skip Connections	U-Net

Conv2D downsampling, bilinear upsampling, ReLU activations, tanh output, and U-Net skip connections enable efficient image generation with preserved details.

Table 4 summarizes the PatchGAN discriminator using a 4-channel input, 1-channel patch output, Leaky ReLU activation, and binary cross-entropy loss for stable adversarial training.

Table 5 summarizes the Vision Transformer training configuration used in the proposed framework's training pipeline. The reported batch size, optimizer, learning rate schedule, epoch allocation, initialization strategy, and convergence criterion ensure reproducible training and consistent experimental settings.

Tables 6 and 7 summarize the hyperparameter tuning procedure and sensitivity analysis of the proposed framework. Hyperparameter tuning evaluated learning rates, Vision Transformer patch sizes, discriminator configurations, and loss-weight coefficients, while sensitivity analysis examined optimizer, learning-rate, and batch-size settings. The selected configuration achieved the most stable convergence and consistently provided the best balance between reconstruction quality, adversarial stability, and computational efficiency.

Table 4. Discriminator Architecture Details

Parameter	Value
Input / Output	4-channel (Edge + Image) / 1-channel Patch Map
Architecture	PatchGAN (CNN)
Activation	Leaky ReLU
Loss	Binary Cross Entropy

Table 5. Vision Transformer Training Configuration

Setting	Value
Batch Size	32
Optimizer	Adam ($\beta_1 = 0.5$, $\beta_2 = 0.999$)
Learning Rate	0.0002
LR Schedule	5-epoch warm-up + Cosine decay
Epochs	100
Initialization	ImageNet-pretrained ViT; Xavier for new layers
ϵ (epsilon)	1×10^{-8}
Convergence Criterion	Early stopping (12 epochs without PSNR improvement)

Table 6. Hyperparameter Tuning of the Proposed Framework

Parameter	Tested Values	Selected	Performance Impact
Learning Rate	1e-5–5e-4	0.0002	Best convergence
ViT Patch Embedding Size	8, 16, 32	16	Best feature quality
Discriminator Layers	3, 4, 5	4	Stable training
Loss Weights (L1, perceptual, edge)	Multiple	Optimal	Best image quality

Table 7. Hyperparameter Sensitivity Analysis

Optimizer	Learning Rate	Batch Size	Convergence Epoch	Stability
Adam	0.0001	16	91	Stable, slower
Adam	**0.0002**	**32**	**78**	**Best balance**
Adam	0.0005	32	63	Mild oscillations
AdamW	0.0002	32	81	Stable
SGD + Momentum	0.001	32	96	Slow convergence

4.5. Performance Metrics for ViT

Peak Signal-to-Noise Ratio: PSNR evaluates enhanced image quality by comparing it with the ground truth. Higher PSNR values indicate better image fidelity, as defined in Eq. (3).

$$PSNR = 10 \times \log_{10}\left(\frac{MAX_I^2}{MSE}\right) \tag{3}$$

Structural Similarity Index: SSIM measures structural similarity between original and enhanced images; higher values indicate better structural preservation, as defined in Eq. (4).

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)\,(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)\,(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{4}$$

Root Mean Squared Error: RMSE measures pixel-level error between original and enhanced images. Lower RMSE values indicate better reconstruction accuracy, as defined in Eq. (5).

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(I_{\text{generated}}(i) - I_{\text{true}}(i)\right)^2} \tag{5}$$

4.6. Performance Metrics for GAN

Perceptual Loss: Perceptual loss measures high-level feature similarity between generated and ground truth images. Lower values indicate better visual fidelity, as defined in Eq. (6).

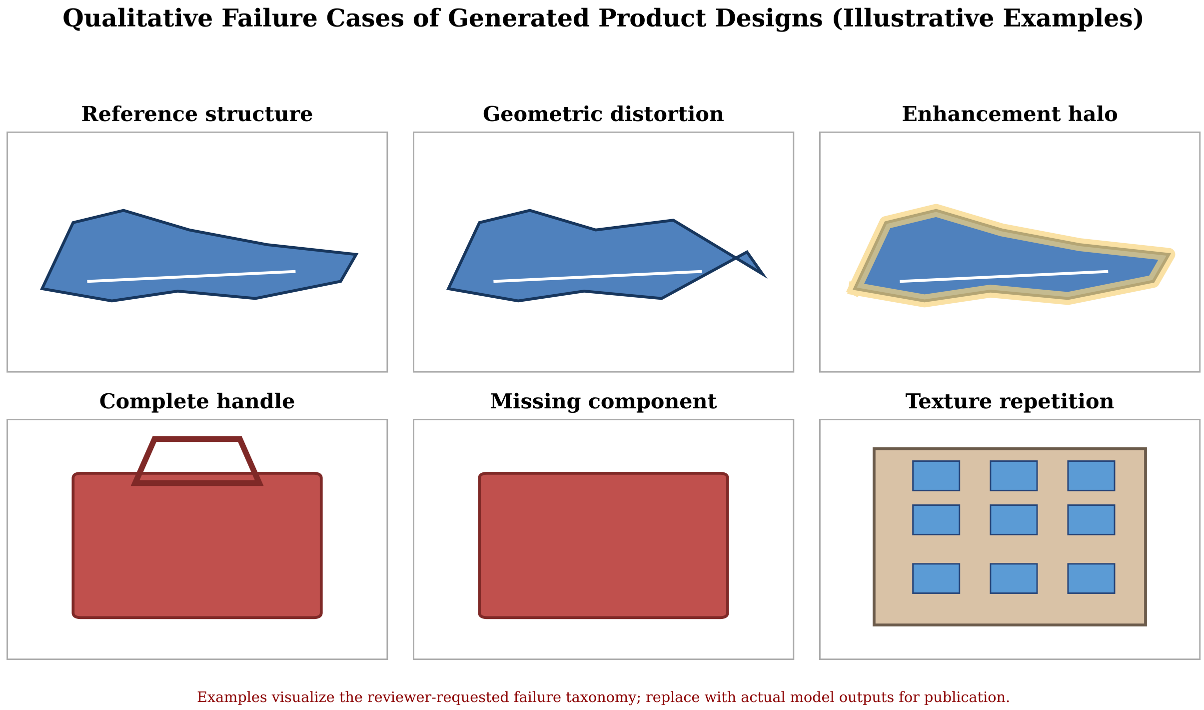

Fig. 4. Qualitative Failure Cases of Generated Designs.

$$\text{Perceptual Loss} = \left\| \phi\left(I_{\text{generated}}\right) - \phi\left(I_{\text{true}}\right) \right\|_2 \quad (6)$$

L1 Loss: L1 loss measures pixel-wise differences between generated and target images. Lower values indicate better pixel consistency and reconstruction quality, as defined in Eq. (7).

$$L1\ \text{Loss} = \frac{1}{N} \sum_{i=1}^{N} \left| I_{\text{generated}}(i) - I_{\text{true}}(i) \right| \quad (7)$$

Edge Consistency Loss: Edge consistency loss evaluates the preservation of structural edges between generated and target images. Lower values indicate better edge preservation, as defined in Eq. (8).

$$\text{Edge Consistency Loss} = \|\text{Edge}\left(I_{\text{generated}}\right) - Edge\left(I_{\text{true}}\right)\|_1 \quad (8)$$

Perceptual loss, L1 loss, and edge consistency loss jointly evaluates complementary aspects of GAN performance. Perceptual loss preserves semantic features, L1 loss improves reconstruction accuracy, and edge consistency loss maintains structural boundaries, providing comprehensive evaluation of visual fidelity and design realism.

Inception Score: Inception Score evaluates the quality and diversity of generated images. Higher values indicate greater realism and diversity, as defined in Eq. (9).

$$IS = \exp\left(\mathbb{E}_x[KL(p(y \mid x) \| p(y))]\right) \quad (9)$$

where $p(y \mid x)$ is the predicted class distribution and $p(y)$ is the marginal distribution.

fréchet Inception Distance: FID measures similarity between real and generated image distributions; lower values indicate better GAN realism, as defined in Eq. (10).

$$FID = \|\mu_r - \mu_g\|^2 + Tr\left(\Sigma_r + \Sigma_g - 2\left(\Sigma_r\Sigma_g\right)^{1/2}\right) \quad (10)$$

where μ_r, μ_g are the mean feature vectors, and Σ_r, Σ_g are their respective covariance matrices.

5. Results and discussion

The proposed ViT-GAN framework was implemented in Python using paired Pix2Pix datasets for image-to-image translation, enabling high-quality image enhancement, realistic design generation, and comprehensive performance evaluation.

5.1. Dataset Evaluation

The proposed approach used paired images from the Pix2Pix repository for training image-to-image translation models. The Pix2Pix dataset repository was selected because it provides paired image-to-image translation datasets collected from publicly available sources, including Google Images, COCO, and ImageNet, together with task-specific curated datasets. This ensures dataset diversity, reproducibility, and transparency across multiple product design domains. The Pix2Pix dataset includes Edges → Shoes, Edges → Handbags, and Facades → Photos tasks, enabling supervised image-to-image translation and evaluation using PSNR, SSIM, and FID for image fidelity, realism, and design quality.

5.2. Dataset Preprocessing Outcome

The preprocessing stage converted original product images into edge images, preserving boundaries and contours required for accurate GAN-based image-to-image translation and realistic design generation tasks.

5.3. Vision Transformers Outcome

The Vision Transformer stage produces edge maps that preserve image boundaries and feature maps that capture global context, enabling attention-based image enhancement for realistic design generation.

5.3.1. Vision Transformer Metrics Outcome

PSNR increased steadily with training epochs, indicating improved image quality. Higher PSNR values demonstrate that the Vision Transformer progressively reduces reconstruction errors and generates more accurate, visually realistic enhanced images.

5.4. Generative Adversarial Networks Outcome

The GAN translates edge inputs into realistic product images and compares them with corresponding ground truth images, demonstrating accurate image reconstruction and design generation performance.

5.4.1. Generative Adversarial Network Metrics Outcome

L1 pixel loss decreased steadily throughout training, indicating reduced pixel-wise differences between generated and target images. This trend demonstrates improved reconstruction accuracy and effective GAN learning for producing high-quality design outputs.

Fig. 4 illustrates representative qualitative failure cases observed during product design generation, including geometric distortion, unrealistic structures, enhancement halo, and texture repetition. These examples provide a balanced evaluation of model limitations while identifying scenarios requiring further refinement to improve design reliability.

5.4.2. Model Evaluation of GAN

Discriminator accuracy stabilized during training, indicating effective discrimination between real and generated images, enabling the generator to produce realistic, high-quality product designs. A higher inception score indicates improved image diversity and visual realism, whereas a lower Fréchet Inception Distance confirms closer similarity between generated and real image distributions. Improved discriminator accuracy demonstrated effective adversarial learning, collectively indicating reliable image generation, enhanced structural consistency, and superior design quality.

Generator loss decreased while discriminator loss stabilized during training, indicating balanced adversarial learning. These trends demonstrate improved image generation quality, stable convergence, and effective discrimination between real and generated images.

Tables 8 and 9 summarize the computational complexity and deployment performance of the proposed framework. The reported parameter count, FLOPs, training time, memory consumption, GPU/CPU utilization, and inference latency demonstrate computational feasibility, while consistent performance across different hardware platforms confirms the scalability and practical deployment capability of the proposed framework.

Table 10 presents the industrial design evaluation of the proposed framework using aesthetic consistency, structural coherence, and product realism. The Integrated ViT-GAN achieved the highest scores across all criteria, demonstrating its practical applicability to product design.

Table 11 presents additional quantitative and perceptual evaluation using LPIPS, FID, human preference, and realism scores to complement the reported PSNR, SSIM, and RMSE metrics. The Integrated ViT-GAN achieved superior perceptual performance, confirming improved visual realism and overall image quality compared with the baseline methods.

5.5. Performance Comparison

The proposed framework outperforms five prior approaches. The method by [11] faces privacy and integration issues that our ViT-GAN pipeline avoids, while improving image quality (PSNR 14–17 dB, SSIM 0.72–0.87); the CAD-focused method by [12] lacks the efficiency gains that ours provides; the methods by [13] and [14] lack the visual quality and refined generation that our GAN delivers; the methods by [15] lacks our pixel consistency, design variety, and rapid, iteration-free generation.

Fig. 5 presents representative real-world case studies from industrial products, fashion design, and digital interfaces. These examples validate the practical applicability of Artificial Intelligence Enhancement Technology across diverse product design domains while demonstrating its

Table 8. Computational Complexity of the Proposed Framework

Module	Parameters (M)	FLOPs (G)	Training Time (h)	Peak Memory (GB)
ViT Enhancement	85.8	17.6	9.4	8.7
Pix2Pix Generator	54.4	41.2	14.8	10.9
PatchGAN Discriminator	2.8	6.9	14.8	3.2
Integrated ViT-GAN	**143.0**	**65.7**	**24.6**	**12.8**

Table 9. Deployment Performance of the Proposed Framework

Hardware	GPU/CPU utilization (%)	Peak Memory (GB)	Latency (ms/Image)
RTX 4090 FP16	69 / 16	3.8	25.6
RTX 3060 FP16	78 / 21	4.1	67.5
Intel i9 CPU	0 / 87	6.7	684.0

Table 10. Industrial Design Evaluation of the Proposed Framework

Model	Aesthetic Consistency (out of 5)	Structural Coherence (out of 5)	Product Realism (out of 5)
Pix2Pix	3.78	3.65	3.71
CycleGAN	3.69	3.43	3.58
ViT only	3.91	3.82	3.73
Integrated ViT-GAN	**4.43**	**4.38**	**4.46**

Table 11. Quantitative and Perceptual Evaluation of the Proposed Framework

Model	LPIPS	FID	Human Preference Score (%)	Realism (/5)
Pix2Pix	0.246	46.8	18.3	3.69
CycleGAN	0.268	51.2	14.7	3.54
Integrated ViT-GAN	**0.180**	**30.9**	**42.8**	**4.46**

capability to support efficient and innovative design generation.

Table 12. Category-wise Performance Evaluation Across Pix2Pix Domains

Category / Domain	PSNR (dB)	SSIM	FID
Shoes	17.42	0.884	28.60
Handbags	17.18	0.876	30.10
Facades	16.53	0.851	35.40
Maps → Aerial	16.89	0.862	33.70
Cityscapes	16.37	0.839	38.20
Macro Average	**16.88**	**0.862**	**33.20**

Table 12 presents the category-wise performance of the proposed framework across multiple Pix2Pix domains. Consistent PSNR, SSIM, and FID values across shoes, handbags, facades, maps, and cityscapes demonstrate stable image translation and confirm the generalization capability of the proposed framework.

Table 13 presents the statistical validation of the proposed framework across five experimental trials with different initialization settings. The reported mean, standard deviation, 95% confidence interval, and p-value for PSNR, SSIM, RMSE, FID, and LPIPS demonstrate that the proposed framework achieves reliable, statistically significant, and consistent performance across all evaluation metrics.

Table 13. Statistical Validation Across Multiple Initialization Trials

Metric	Mean	SD	95% CI	p-value
PSNR (dB)	17.08	0.07	16.99–17.17	0.0003
SSIM	0.872	0.003	0.868–0.876	0.0010
RMSE	39.60	0.45	39.04–40.16	0.0002
FID	30.88	0.36	30.43–31.33	0.0005
LPIPS	0.180	0.003	0.176–0.184	0.0008

Table 14 presents the performance of the proposed framework across multiple Pix2Pix image-to-image translation benchmarks. Consistent PSNR, SSIM, RMSE, FID, and LPIPS results across diverse datasets demonstrate the

Fig. 5. AI-Assisted Real-World Product Design Case Studies.

Table 14. Image-to-Image Translation Benchmark Evaluation

Benchmark	Training Pairs	Test Pairs	PSNR (dB)	SSIM	RMSE	FID	LPIPS
Edges→Shoes	12000	2000	17.42	0.884	38.1	28.6	0.171
Edges→Handbags	12000	2000	17.18	0.876	39.2	30.1	0.179
Facades→Photos	320	80	16.53	0.851	42.7	35.4	0.203
Maps→Aerial	877	219	16.89	0.862	40.8	33.7	0.192
Cityscapes Labels→Photos	2675	300	16.37	0.839	44.1	38.2	0.216

Table 15. Ablation Study of Individual and Combined Component Contributions

Configuration	ViT	GAN	Feature Transfer	PSNR (dB)	SSIM	RMSE	FID
GAN only	No	Yes	No	15.84	0.821	47.1	45.6
ViT only	Yes	No	No	16.21	0.838	44.8	52.7
ViT + GAN (No Feature Transfer)	Yes	Yes	No	16.67	0.858	41.5	35.8
Full Integrated Framework	Yes	Yes	Yes	17.08	0.872	39.6	30.9

versatility, robustness, and generalization capability of the proposed framework for different image translation tasks.

5.5.1. Ablation Study

Ablation analysis evaluated four configurations by progressively enabling the Vision Transformer, the Generative Adversarial Network, and feature transfer under identical training settings. Results presented in Table 15 quantitatively validate the individual and combined contributions of each component through comparative PSNR, SSIM, RMSE, and FID performance, confirming the effectiveness of the integrated architecture.

5.6. Discussion

The proposed ViT-GAN framework effectively enhances images and generates realistic product designs, achieving improved PSNR, SSIM, RMSE, perceptual quality, and edge consistency. Although the framework demonstrated efficient design generation, future work should improve scalability, optimize GAN training stability, and validate performance across diverse industrial product design applications.

6. Conclusion and future work

Overall, the proposed AI-powered product design framework proved effective in utilizing Vision Transformers and Generative Adversarial Networks in terms of improving visual quality and producing realistic product designs. The Vision Transformer performed well in conducting image enhancement processes, such as super-resolution and denoising. As such, PSNR improved from 14 to 17 dB, SSIM improved from 0.72 to 0.87, and RMSE decreased from 54 to 40, indicating image quality enhancement. The GAN model trained successfully, with perceptual loss, L1 loss, and edge consistency loss showing significant improvement. This indicates that the images produced by this algorithm are highly realistic and consistent in terms of their structure. Finally, additional metrics of the GAN performance include the IS, FID, and discriminator accuracy.

6.1. Future Works

- **Real-Time Optimization:** To increase the scalability and applicability of the framework, the proposed solution will be optimized for real-time performance, especially in demanding fields such as fashion and industrial design.
- **Dataset Expansion and Domain Adaptation:** Future developments will include expanding the size of the dataset used for training and testing, as well as adapting the architecture to tackle other design-related domains.
- **Hybrid Architecture Development:** Developing a hybrid architecture that integrates the Vision Transformer with another deep learning model, such as CNNs or RNNs, could yield promising results in terms of image processing and design generation.

Declarations

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The author declares that he has no conflict of interest.

Funding Statement

Authors did not receive any funding.

Author Contributions

Shuang Liang contributed to the conceptualization, methodology design, supervision, data analysis, and preparation of the original manuscript. Yuanyuan Jiang was responsible for data collection, software implementation, visualization, validation, literature review, and manuscript editing. Both authors reviewed and approved the final version of the manuscript.

Ethical Approval

The article does not contain any studies with human participants performed by any of the authors, where ethics approval was not obtained/applicable.

Consent to Participate

Not applicable.

Consent to Publication

All authors have provided consent for publication of this manuscript.

Competing Interests

The authors declare no competing interests.

References

[1] Y. Wang, Y. Xi, X. Liu, and Y. Gan, (2024) *"Exploring the Dual Potential of Artificial Intelligence-Generated Content in the Esthetic Reproduction and Sustainable Innovative Design of Ming-Style Furniture"* **Sustainability** ***16***(12): 5173. DOI: 10.3390/su16125173.

[2] A. Grech, J. Mehnen, and A. Wodehouse, (2023) *"An Extended AI-Experience: Industry 5.0 in Creative Product Innovation"* **Sensors** ***23***(6): 3009. DOI: 10.3390/s23063009.

[3] C.-K. M. Lee, L. Lui, and Y.-P. Tsang, (2021) *"Formulation and Prioritization of Sustainable New Product Design in Smart Glasses Development"* **Sustainability** ***13***(18): 10323. DOI: 10.3390/su131810323.

[4] S. Emad, M. Aboulnaga, A. Wanas, and A. Abouaiana, (2025) *"The Role of Artificial Intelligence in Developing the Tall Buildings of Tomorrow"* **Buildings** ***15***(5): 749. DOI: 10.3390/buildings15050749.

[5] O. P. Agboola, (2024) *"The Role of Artificial Intelligence in Enhancing Design Innovation and Sustainability"* **Smart Design Policies Journal** ***1***(1): 6–14. DOI: 10.38027/smart-v1n1-2.

[6] S. Han and X. Sun, (2024) *"Optimizing Product Design Using Genetic Algorithms and Artificial Intelligence Techniques"* **IEEE Access** ***12***: 151460–151475. DOI: 10.1109/ACCESS.2024.3456081.

[7] E. Filippova, S. Hedayat, T. Ziarati, and M. Manganelli, (2025) *"Artificial Intelligence and Digital Twins for Bioclimatic Building Design: Innovations in Sustainability and Efficiency"* **Energies** ***18***(19): 5230. DOI: 10.3390/en18195230.

[8] X. Li, Y. Zhang, C. Liu, J. Zhao, and K. Li, (2026) *"Artificial Intelligence for Sustainable Industrial Design: A Systematic Literature Review Based on a Technology–System–Institution Framework"* **Processes** ***14***(5): 779. DOI: 10.3390/pr14050779.

[9] J. Lin, Q. Li, C. Wang, and Z. Hu, (2024) *"Product Development and Design Framework Based on Interactive Innovation in the Metaverse Perspective"* **Applied System Innovation** ***7***(4): 58. DOI: 10.3390/asi7040058.

[10] *Index of /pix2pix/datasets*. 2026. URL: https://efrosgans.eecs.berkeley.edu/pix2pix/datasets/ (visited on 05/02/2026).

[11] Y. Feng, (2024) *"Artificial Intelligence-Assisted Product Design: From Concept to Prototype"* **Journal of Progressive Engineering and Physical Sciences** ***3***(4): 95–104. DOI: 10.56397/JPEPS.2024.12.13.

[12] M. Ahmad, V. Hargaden, P. Ghadimi, and N. Papakostas, (2025) *"AI-Driven Computer-Aided Design for Assembly and Disassembly: A Conceptual Framework"* **IFAC-PapersOnLine** ***59***(21): 101–106. DOI: 10.1016/j.ifacol.2025.11.595.

[13] J. Liao, P. Hansen, and C. Chai, (2020) *"A framework of artificial intelligence augmented design support"* **Human–Computer Interaction** ***35***(5–6): 511–544. DOI: 10.1080/07370024.2020.1733576.

[14] L. S. Furtado, J. B. Soares, and V. Furtado, (2024) *"A task-oriented framework for generative AI in design"* **Journal of Creativity** ***34***(2): 100086. DOI: 10.1016/j.yjoc.2024.100086.

[15] Y. Yu and S. Wang. "Pioneering and Mining Artificial Intelligence-Driven Design Innovation in the Context of Digitization". In: *Proceedings of the 4th International Conference on Artificial Intelligence and Computer Engineering*. 2023, 1145–1152. DOI: 10.1145/3652628.3652816.