

CT-GAN-Based Engineering System for Intelligent Graphic Advertising Design

Shan Gao¹ and Qiuhan Wang²

¹ Lecturer, School of Design and Arts, North Minzu University, Yinchuan, 750000, China,

E-mail: gaoshan801@outlook.com (corresponding author).

² Master's Student, School of Design and Arts, North Minzu University, Yinchuan, 750000, China, E-mail:

qiuhanwang@heu.uu.me

Project Management

Received January 24, 2026; revised March 17, 2026; accepted September 21, 2026

Available online September 27, 2026

Abstract: To address the inefficiency, high cost, and poor scalability of traditional manual graphic advertising design in large-scale digital marketing, where organizations face pressures of rapid content iteration, personalized targeting, and cross-departmental collaboration. This paper proposes an engineering-oriented Conditional Transformation Generative Adversarial Network (CT-GAN) that integrates a Conditional Generative Adversarial Network (CGAN) and Transformer modules for intelligent, automated, high-quality graphic advertising generation. CT-GAN addresses key engineering challenges (slow convergence, insufficient clarity) that hinder the practical deployment of traditional models. Evaluated on FashionAd-5K (5,000 images) and AdDesign-20K (20,000 multi-category images), it achieves mIoU values of 0.8–0.9 and 0.5–0.7, respectively, outperforming baselines in element localization. In practical applications, CT-GAN enables end-to-end automated workflows to generate personalized, high-fidelity ads with rational layouts. It optimizes organizational creative processes and aligns with engineering priorities (cost control, scalability, cross-scenario adaptation), providing robust technical support for intelligent design systems and data-driven digital marketing.

Keywords: Conditional generative adversarial network; production system; digital engineering workflow; decision support; operational efficiency.

Copyright © Journal of Engineering, Project, and Production Management (EPPM-Journal).

DOI 10.32738/JEPPM-2026-106

1. Introduction

Graphic advertising design is crucial for brand communication and marketing, as its visual quality and creativity directly impact commercial outcomes and brand perception. Effective designs must balance aesthetics with clear messaging and audience engagement. However, the growth of e-commerce and fragmented media has increased demand for faster, more personalized, and diverse advertising content, challenging traditional manual design workflows, which are costly and difficult to scale. Current technical approaches also face limitations: template-based tools lack creativity, while deep learning-based image generation often produces flawed layouts, distorted elements, or inconsistent styles, with limited controllability (Yang, 2024). Domestic and foreign scholars have conducted a series of explorations in the field of advertising design and generation. Xu et al. (2024) proposed a layout generation model based on a Conditional Generative Adversarial Network (CGAN) to improve the efficiency of advertising banner generation. The results showed that the model could automatically generate basic layout structures according to product categories, providing a foundation for automated design. This aligns with the advertising design need for rapid prototyping of layout options based on content type. Johri et al. (2025) addressed the problems of loose image-text association and insufficient adaptability of copywriting and products in e-commerce advertising by constructing an encoder-decoder framework that integrates product images and text descriptions and strengthens cross-modal information interaction through attention mechanisms. The results successfully generated simple advertising images with high correlation between copywriting and products, improving the semantic coherence of advertising.

Based on the above research, the key to achieving intelligent advertising generation lies in two core issues. One is how to generate high-quality, high-fidelity, and semantically compliant product images. The second is how to intelligently, harmoniously, and aesthetically integrate the product generated with other advertising elements. CGAN provides a feasible path to controllable image generation by incorporating conditional variables, such as category labels (Zhang et al., 2025). Meanwhile, the Transformer architecture, with its powerful self-attention mechanism, can effectively capture long-range global dependencies in images, demonstrating great potential to improve the quality and coherence of generated images (Li et al., 2025). Given the advantages of both technologies, this study proposes a high-quality automated design method for print advertising based on the Conditional Transformation Generative Adversarial Network (CT-GAN). Aiming to achieve end-to-end advertising automation, from product concept to finished product, through structural innovation in generative

models and rule-based design of layout algorithms, this study provides efficient and intelligent solutions for creative content production in digital marketing. By framing CT-GAN within an engineering systems context, this study helps bridge the gap between generative artificial intelligence and practical production planning, offering insights into how intelligent design tools can be leveraged to control costs, improve scalability, and integrate seamlessly into existing industrial workflows.

2. Graphic Advertising Design Based on CT-GAN Image Generation Network

2.1. Basic Generation Of Graphic Advertising Images Based on CGAN

As the core of the advertisement, the product image must be realistic, detailed, and true to the actual product, as it directly influences purchase intention. The marketing landscape of the digital economy demands both efficiency and personalization in ad creation. AIGC technologies are driving the industry's shift from human-led creation to a human-machine collaborative, intelligent production paradigm. Consumer acceptance now extends beyond information delivery to emphasize perceived authenticity and emotional resonance. However, traditional manual design struggles to balance large-scale output with personalized adaptation, often resulting in emotionally hollow or stylistically homogeneous ads that provoke consumer resistance. As a key AIGC branch, CGAN enhances the original GAN by introducing a conditional variable y^* , enabling precise control over image categories and providing a technical basis for diverse graphic ad generation. Conditional variables can include product category, promotional theme, and brand style. Its network structure is shown in Fig. 1 (Wang, 2024).

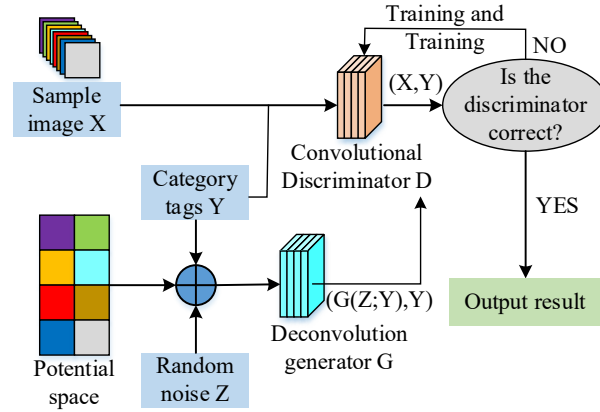


Fig. 1. CGAN structure

As shown in Fig. 1, the core architecture of CGAN consists of a generator G and a discriminator D , which achieve dynamic balance through adversarial training. Unlike traditional GANs, the innovation of CGAN lies in feeding the class label Y as an additional input to both the generator and the discriminator. The generator G is fed into random noise Z and a conditional variable Y , and generates an advertising image corresponding to the label Y by learning the mapping function $G(Z|Y)$. The discriminator D receives the real image X or the generated image $G(Z|Y)$ and its corresponding label Y , and outputs the true probability $D(Z|Y)$ or $D(G(Z|Y))$. This enables the model to generate corresponding product images for the specified graphic advertising category, thereby providing diverse options for displaying advertising products (Leonardi and Alfonsius, 2024; Panhe et al., 2025). Its objective function is shown in Eq. (1).

$$\min_G \max V(D, G) = E_{X \sim p_{data}(X)} [\log D(X|Y)] + E_{X \sim p_z(X)} [\log(1 - D(G(Z|Y)))] \quad (1)$$

Eq. (1) $\min_G \max V(D, G)$ represents the objective function for generating the graphic advertising image. $p_{data}(x)$ represents the data distribution of the real advertising image. The $p_z(X)$ represents the distribution of the random noise vector. The $D(X|Y)$ represents the probability that the discriminator classifies the image as a real sample given a category label. and $G(Z|Y)$ represents the pseudo-sample generated by the generator under noisy and labeled conditions. To meet the requirements for generating graphic advertising images, the specific structures of the generator and discriminator are shown in Fig. 2.

The generator takes a 100-dimensional noise vector Z and a 1-dimensional class label y as input, producing a 28×28 advertising image. The label y is embedded into a 50-dimensional vector, processed by a fully connected layer, reshaped into a $7 \times 7 \times 1$ feature map, and concatenated with Z (after similar fully connected and reshaping steps, producing a $7 \times 7 \times 128$ feature map). The combined $7 \times 7 \times 129$ feature map is upsampled via two transposed convolutions, and a final convolution outputs a single-channel image. LeakyReLU is used in hidden layers, and Tanh normalizes pixel values to $[-1, 1]$ (Wang and Liu, 2024). The discriminator takes a 28×28 image and its label y , and outputs a real/fake probability. The label is embedded, processed and reshaped to $28 \times 28 \times 1$, then concatenated with the image to form a $28 \times 28 \times 2$ input. Two convolutional layers with LeakyReLU extract features, followed by a fully connected layer to produce the probability. Training proceeds with the generator producing realistic, class-consistent images to fool the discriminator, while the discriminator learns to distinguish real from generated samples until the generator stably outputs high-quality images (Sultan and Wani, 2024; Pang and Liu, 2024).

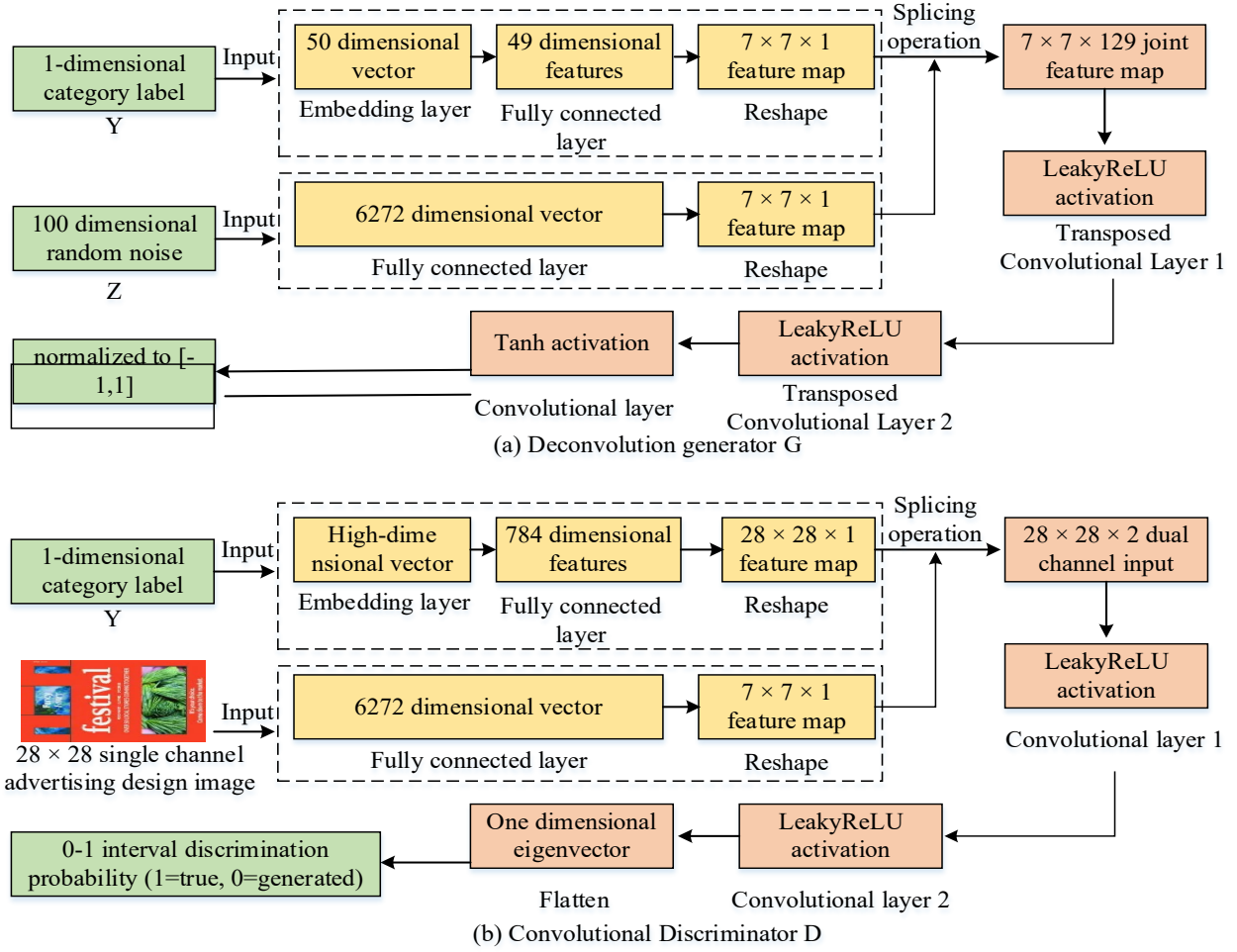


Fig. 2. Generator and discriminator structure

2.2. Construction of CT-GAN Model Based on Transformer

Although the CGAN model can achieve basic generation of graphic advertising images, it has issues such as slow convergence of the loss function, insufficient image clarity and limited stylistic variety. The Transformer module has advantages in feature extraction efficiency and global information capture due to its parallel computing power and multi-head attention mechanism (Wang et al., 2025; Bai, 2025). Therefore, this study integrates the Transformer module into the CGAN generator to construct the CT-GAN model, thereby improving image generation quality and accelerating model convergence. This architectural improvement reduces the computational overhead of iterative training cycles, making the model more suitable for deployment in real-time or near-real-time digital asset management systems, where rapid content iteration and low-latency generation are critical to maintaining production efficiency. The core of the Transformer is the self-attention mechanism, which adaptively captures global dependencies by computing correlation weights for each element in the input sequence, as denoted in Eq. (2) (Li et al., 2024b).

$$Attention(Q, K, V) = \text{soft max} \left(\frac{QK^T}{\sqrt{d_k}} \right) V \quad (2)$$

In Eq. (2), $Attention(Q, K, V)$ represents the self-attention mechanism; Q , K , and V represent the query matrix, key matrix, and value matrix, respectively; d_k is the dimension of K , used to scale the inner product result to avoid gradient vanishing. Due to the diverse characteristics of graphic advertising images, a multi-head attention mechanism is adopted to better adapt to them, as shown in Eq. (3).

$$MultiHead(Q, K, V) = \text{Concat}(head_1, \dots, head_h) W^O \quad (3)$$

In Eq. (3), $MultiHead(Q, K, V)$ represents the multi-head attention mechanism; $head_1, \dots, head_h$ represent multiple parallel attention heads, each independently calculating a set of attention results to capture feature associations in different dimensions. The Concat represents the concatenation operation. The W^O represents the output weight matrix. The structure of the multi-head self-attention mechanism is shown in Fig. 3.

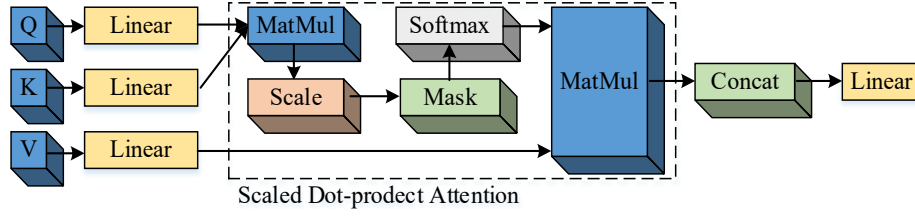


Fig. 3. Schematic diagram of the multi-head self-attention mechanism

As shown in Fig. 3, the visual elements of an advertising image are first mapped into query, key, and value matrices via linear layers, then processed by the Scaled Dot-product Attention module. Attention weights between elements are computed via matrix multiplication, scaling, masking, and Softmax, then combined with the value matrix to produce feature outputs for a single attention head. Multiple parallel heads capture element relationships across different dimensions (e.g., color matching, layout), and the resulting representations are concatenated and linearly transformed to generate the final multi-head attention features (Li et al., 2024a). Based on this mechanism, a Transformer module for flat advertising design is proposed. Input features are embedded and positionally encoded to clarify element representations and spatial relationships. Multi-layer multi-head attention captures global dependencies across color, layout, and semantics, followed by feedforward networks, residual connections, and layer normalization to stabilize feature propagation. The decoder generates design-consistent outputs via masked and encoder-decoder attention, and finally produces element probability distributions via a linear transformation and Softmax, enabling end-to-end intelligent graphic advertising generation with diverse styles and rational layouts (Pan and Liu, 2024). The goal is to output a well-composed key visual for a complete advertisement design. By integrating this Transformer module into a CGAN generator, the CT-GAN model is constructed, as shown in Fig. 4.

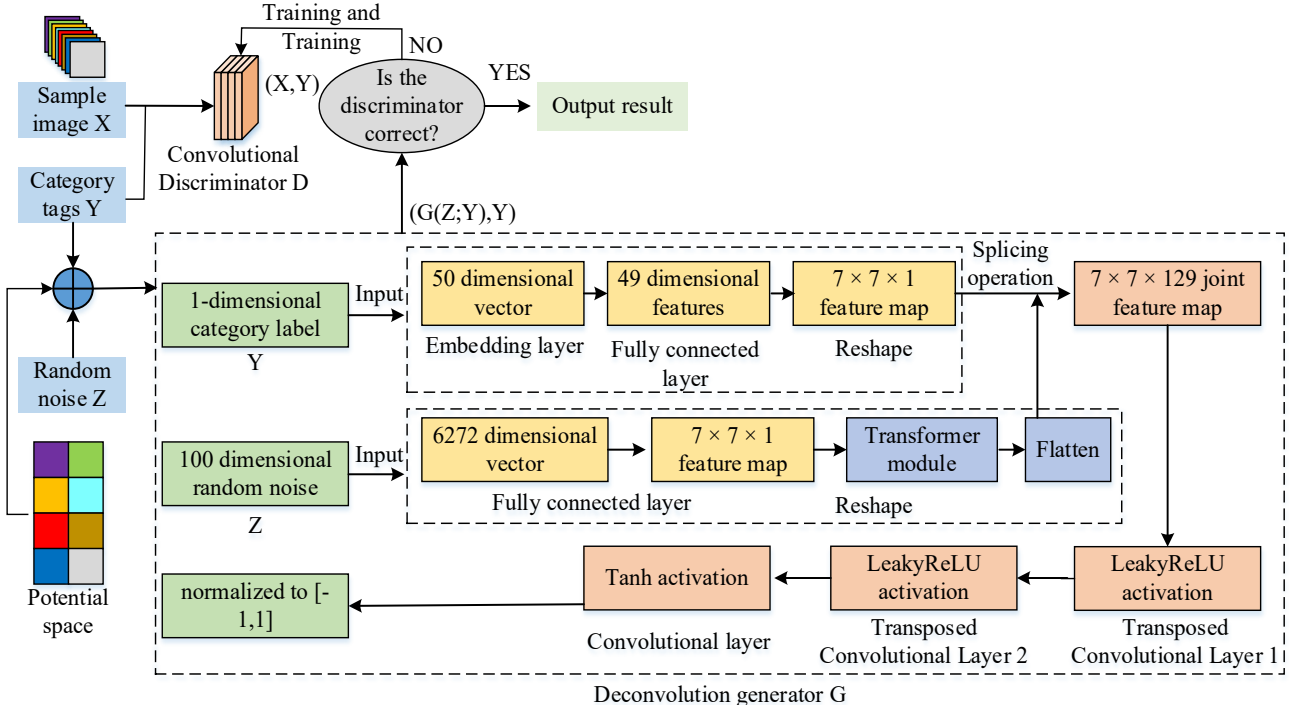


Fig. 4. Schematic of transformer-based CT-GAN model structure

As shown in Fig. 4, the overall architecture of CT-GAN is consistent with that of CGAN, with the core difference being the addition of a Transformer feature-extraction module in the generator. The generator's input remains random noise and class labels. The noise vector is transformed into a multi-dimensional vector via a fully connected layer, then reshaped into image-sequence features, which are fed into the Transformer module for global feature extraction. The Transformer module has two attention heads and a 16-dimensional hidden layer in the feedforward network (Ma, 2024). It captures global dependencies between features through a multi-head attention mechanism. After being flattened, the features are concatenated with the label-processed feature map. Subsequent operations are consistent with CGAN.

2.3. Generation and Integration of Graphic Advertising Design Layout

After generating a high-quality product main image using the CT-GAN model, it is necessary to automatically generate and integrate the graphic advertising layout. The goal of this step is to intelligently arrange the product image, advertising copy (such as the title and promotional text), brand logo and background according to aesthetic principles to form a harmonious, visually appealing, complete advertisement image (Chen et al., 2024). This directly addresses the challenge of automating the labor-intensive compositing and typesetting phase of ad design. This step assembles the original generated product image with the pre-existing background to create the final publishable image. This reduces the need for graphic designers to

manually intervene in each output, streamlines the production process and ensures brand consistency across a large number of advertising creatives. The specific steps are shown in Fig. 5.

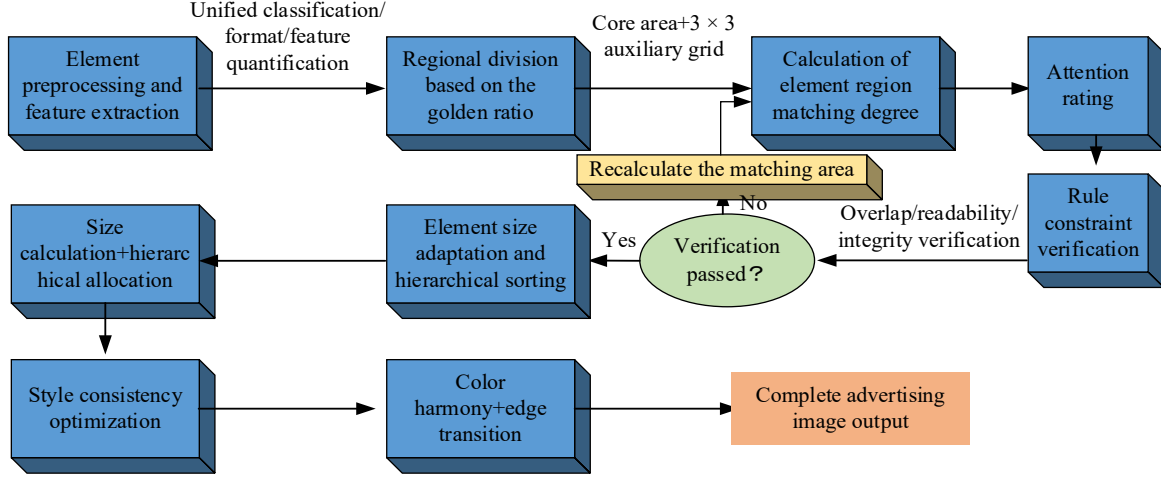


Fig. 5. Graphic advertising design layout generation and integration process

As shown in Fig. 5, before performing layout calculations, all advertising visual elements need to be standardized to ensure the compatibility and rationality of subsequent layout. The study divides advertising element E_i into five categories, including core product images generated by CT-GAN E_1 , title copy E_2 , promotion/explanation copy E_3 , brand logo E_4 , and background layer E_5 . All elements are uniformly set to $H \times W$ resolution and RGBA format to support transparent overlay. The visual weight $W_{vis}(E_i)$ of the advertising image is calculated by the color saturation S , contrast C , and area ratio A of the elements. The formula is shown in Eq. (4). The larger the weight value, the higher the visual priority (Zahra and Parisa, 2024).

$$W_{vis}(E_i) = \alpha \cdot S(E_i) + \beta \cdot C(E_i) + \gamma \cdot A(E_i) \quad (4)$$

In Eq. (4), α , β , and γ these are weight coefficients, which are 0.3, 0.4, and 0.3, respectively, after experimental calibration. Eq. (4) quantifies the principle of visual communication, converting subjective perception into objective values to guide layout decisions. Then, the "golden section + grid layout" fusion algorithm is adopted. First, the advertising canvas is divided into a core area and an auxiliary area, and the elements are then accurately positioned through grid subdivision (Liu, 2023). Different golden section ratios are used depending on the advertising type (horizontal/vertical). The core area of the horizontal advertisement (width-to-height ratio 16:9) is the intersection of the left 2/3 and the top 2/3 of the canvas, which is used to place product images E_1 . The core area of the vertical advertisement (width-to-height ratio 9:16) is the top 1/2 area of the canvas. The remaining area is divided into 3×3 grids, and the coordinates of each grid are calculated using Eq. (5) to place auxiliary elements such as copywriting and the brand logo.

$$\begin{cases} R_{m,n} = [x_{start}, y_{start}, x_{end}, y_{end}] \\ x_{start} = (m-1) \cdot \frac{W}{3}, x_{end} = m \cdot \frac{W}{3} \\ y_{start} = (n-1) \cdot \frac{H}{3}, y_{end} = n \cdot \frac{H}{3} \end{cases} \quad (5)$$

In Eq. (5), $R_{m,n}$ represents the advertising design rule. The m represents the grid column index. The n represents the grid row index. The W represents the canvas width, and H represents the canvas height. Eq. (5) provides a spatial framework for balanced visual communication. Combining the weights of the advertising design elements with the advertising design rules, the optimal placement area for each element is determined, as shown in Eq. (6).

$$Score(E_i, R_{m,n}) = \omega_1 \cdot W_{vis}(E_i) \cdot I(R_{m,n}) + \omega_2 \cdot W_{sem}(E_i) \cdot S(R_{m,n}) \quad (6)$$

In Eq. (6), $Score(E_i, R_{m,n})$ is the matching score between the advertising element and the region. The higher the score, the more suitable the element is for placement in the region. ω_1 and ω_2 are the balance coefficients. $W_{vis}(E_i)$ and $W_{sem}(E_i)$ are the weights of the element and the semantics, respectively. $I(R_{m,n})$ is the importance of the coefficient for the region $R_{m,n}$. The core region takes 1.0, and the auxiliary region takes 0.1-0.8 based on visual priority. $S(R_{m,n})$ is the semantic adaptation coefficient of region $R_{m,n}$, which highlights the semantic adaptation between the region and the element. Eq. (6) operationalizes the visual hierarchy principle, mapping information priority onto spatial placement to ensure logical message delivery.

Next, the element size is dynamically adjusted based on the placement area and the element's characteristics, and the hierarchical relationship is clarified to ensure the visual rationality of the advertising image design. Size adaptation is shown in Eq. (7) (Ma, 2024).

$$\begin{cases} W_i = \min(W_{ori}(E_i) \bullet k, W_{R_{m,n}} 0.9) \\ H_i = W_i \bullet AR(E_i) \end{cases} \quad (7)$$

In Eq. (7), $W_{ori}(E_i)$ is the original width of the element. The k is the scaling factor. The $W_{R_{m,n}} 0.9$ is the width of the placement area. The $AR(E_i)$ is the original aspect ratio of the element. The 0.9 is the margin factor. W_i and H_i are the dynamically adjusted element sizes, respectively. The hierarchical sorting rules are ordered from "background layer to product image, to title copy, to promotional copy, to brand logo" (Bello and Oladipo, 2024).

Finally, through color harmonization and edge transition processing, the style of all elements is kept consistent, creating a harmonious visual effect. Color harmonization is based on the main color tone of the product image, and the color deviation of other elements is adjusted, as shown in Eq. (8).

$$H_{adj}(E_i) = H_{ori}(E_i) - |H_{ori}(E_i) - H_{main}(E_1)|\tau \quad (8)$$

In Eq. (8), $H_{adj}(E_i)$ is the adjusted element color tone. The $H_{ori}(E_i)$ is the original element color tone. $H_{main}(E_1)$ is the main color tone of the product image. The τ is the harmonic coefficient. This ensures text and logo colors complement the product, avoiding clashing palettes that can undermine ad effectiveness. Eq. (8) codifies color harmony theory, algorithmically enforcing visual consistency by anchoring all tones to the dominant product color. By automating the entire pipeline from core image generation to final asset assembly, this system functions as a cohesive unit within a larger digital production ecosystem. It effectively addresses the problems of low efficiency and a lack of personalization in traditional advertising design, providing a complete and scalable solution for the automated generation of graphic advertising in intelligent design and digital marketing, thereby optimizing the digital asset lifecycle from creation to deployment.

3. Results

3.1. Experimental Environment Development

To verify the effectiveness of the proposed CT-GAN model and advertising design generation integration scheme, a dedicated environment was constructed for the experiment. The hardware configuration included an Intel Core i9-13900K processor, an NVIDIA GeForce RTX 4090 graphics card (with 24GB of video memory), and 64GB of DDR5 memory. The software was based on Windows 11 system, and the model was developed and trained using Python 3.9 and PyTorch 1.13.1 framework. The advertising element data was managed through a MySQL database. The experimental dataset includes 5000 clothing advertisement images (FashionAd-5K), crawled and manually screened from mainstream e-commerce platforms, as well as the AdDesign-20K dataset, which covers 8 categories and contains 20000 annotated images. The learning rate during model training was set to 3×10^{-4} , using the AdamW optimizer with a batch size of 32, 100 training epochs, a Dropout rate of 0.1, a hidden layer dimension of 512, and a maximum node count of 50. To verify the superiority of the designed CT-GAN model in graphic advertising image design, CGAN, Deep Convolutional Generative Adversarial Network (DCGAN), and Wasserstein Generative Adversarial Network with Gradient Penalty (WGAN-GP) were selected as the baseline models for comparative testing under the same experimental conditions.

3.2. Algorithm Performance Analysis

3.2.1. Analysis of model's mean intersection over union test results

This study used the Mean Intersection over Union (mIoU) as the core evaluation metric. Semantic segmentation was performed on the advertising images generated by each model on the test set, and the IoU for each element category was calculated and averaged. The results are shown in Fig. 6.

In Fig. 6(a), the CT-GAN model had a stable mIoU value between 0.8 and 0.9 on this dataset, consistently higher than the other three comparison models. This indicates that CT-GAN has the strongest ability to locate and preserve the shapes of key visual elements, such as clothing products and text descriptions, within the clothing advertising domain, and to generate clothing advertising images with a reasonable layout and clear boundaries between elements. Similar to Fig. 6(a), the CT-GAN model also exhibited the best performance, with a stable mIoU value between 0.5 and 0.7, consistently higher than the other comparison models. This result proves that CT-GAN is not only suitable for advertising generation in a specific domain but also demonstrates excellent element localization capabilities across general advertising datasets spanning multiple categories, reflecting the model's strong generalization and versatility.

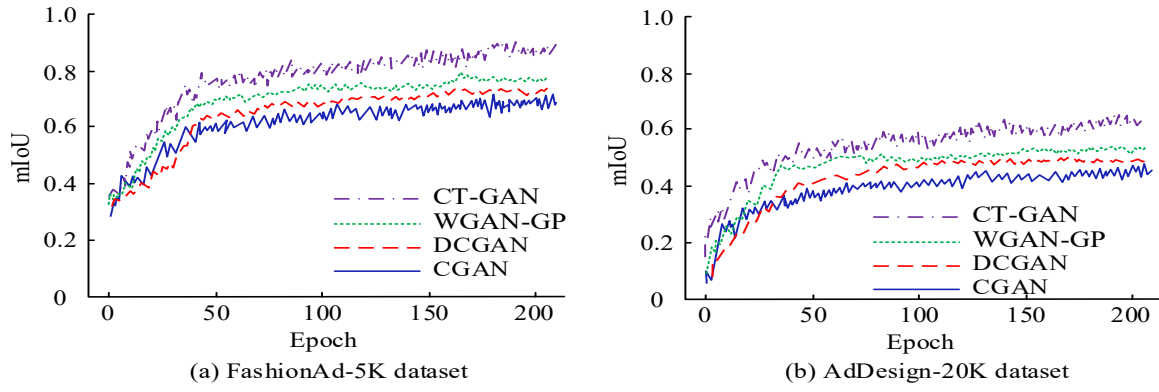


Fig. 6. mIoU test results for the four models

3.2.2. Analysis of test results on model loss function value and convergence changes

To thoroughly evaluate the stability and efficiency of each adversarial training process, the study recorded changes in the loss values of CT-GAN and the baseline model during training and analyzed their convergence characteristics. The results are shown in Fig. 7.

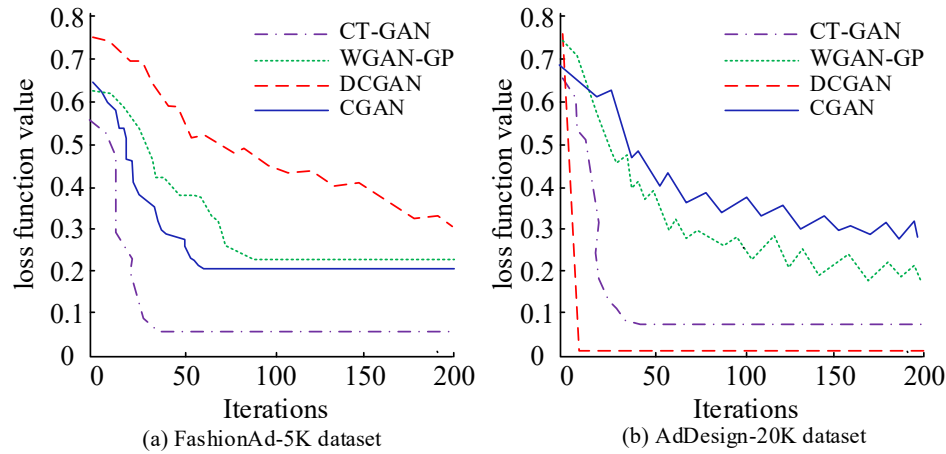


Fig. 7. Test results of loss function values and convergence changes on four models

In Fig. 7(a) of the FashionAd-5K dataset, the CT-GAN loss decreased most rapidly and then stabilized after approximately 40 training epochs, with the final loss remaining around 0.05. In contrast, CGAN and WGAN-GP required approximately 60 and 75 epochs, respectively, to reach stability, with their loss values fluctuating around 0.20 and 0.22, respectively. DCGAN exhibited the slowest convergence and the most pronounced fluctuations, with its loss failing to stabilize by the end of training, indicating significant instability. In Fig. 7(b), in the AdDesign-20K dataset, the discriminator loss of CT-GAN converged rapidly to approximately 0.07 and remained stable, while the discriminator losses of CGAN and WGAN-GP showed significant oscillations in the middle of training, reflecting the difficulty these models faced in finding Nash equilibrium points when dealing with large datasets.

3.3. Analysis of Effects of Graphic Advertising Image Design

3.3.1. Objective indicators for image design quality evaluation

This study used Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) as supplementary evaluation metrics. A higher PSNR value indicates less image distortion, while a higher SSIM value indicates greater structural similarity between the generated and real images. The evaluation results of each model on the two datasets are shown in Fig. 8.

In Figs 8(a) and 8(b), on the FashionAd-5K and AdDesign-20K datasets, CT-GAN achieved PSNR values of 28.7 dB and 26.9 dB, respectively, which are approximately 3.4 dB and 2.8 dB higher than the second-best performing CGAN model. A higher PSNR value indicates smaller pixel-level errors between the generated image and the real image, and lower image noise. This demonstrates that the generated ad images are more closely aligned with the images in the real ad library at the pixel level. As shown in Figs 8(c) and 8(d), CT-GAN achieved SSIM values of 0.892 and 0.865 on the FashionAd-5K and AdDesign-20K datasets, respectively, significantly higher than other models

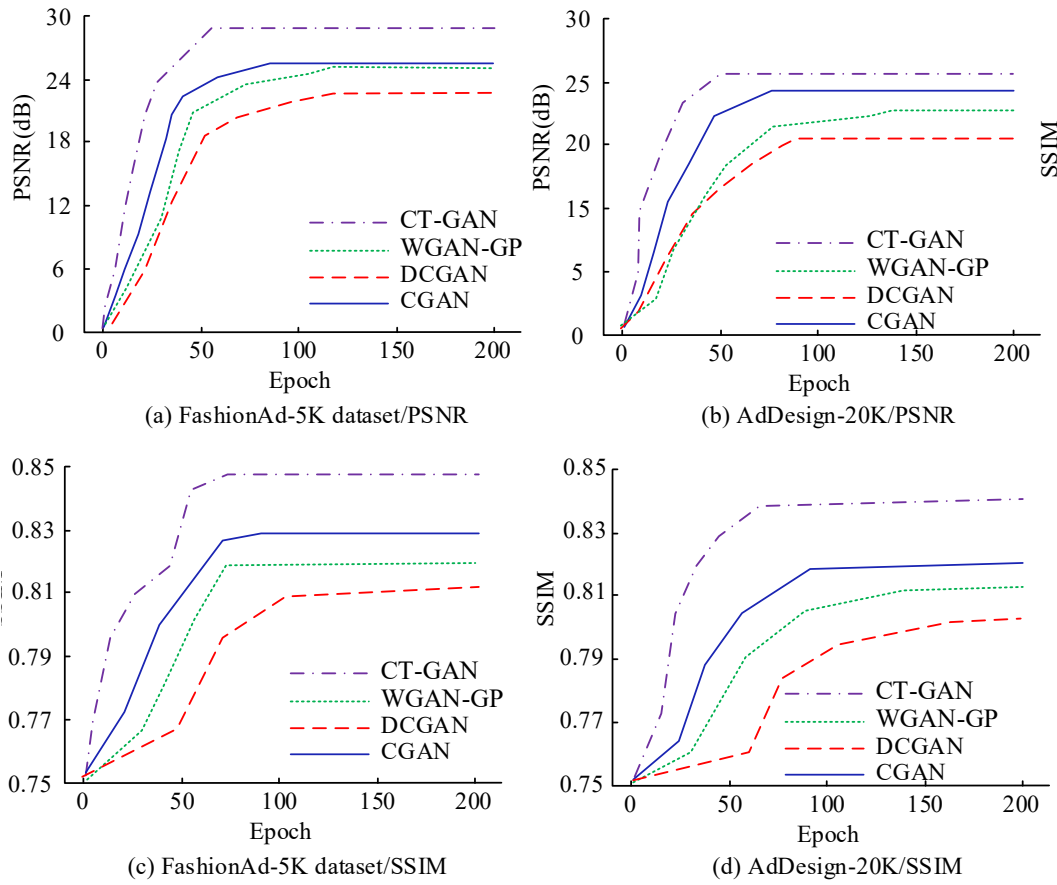


Fig. 8. Test results of objective indicators of image design quality for four models

3.3.2. Comparative Analysis of Personalized Advertising Generation Effects

Taking the architectural design of an "urban cultural complex" as an example, four different models were used to generate corresponding architectural scheme renderings, and then a visual comparative analysis was conducted. The evaluation criteria included diversity of design elements, color application, visual impact, and theme recognizability. 30 professional architects evaluated the results on a 5-point scale. The visualization results generated by the architecture scheme are shown in Fig. 9.



Fig. 9. Comparison of generation effect on architectural design schemes

As shown in Fig. 9, among the four building scheme generation methods, CT-GAN ranked highest in design element diversity, color application, visual impact, and theme recognition. The generated architectural renderings integrated rich cultural and modern design elements, featuring strong, harmonious color combinations. It created a strong visual impact through the contrast between materials and lighting, while clearly conveying the theme of "urban cultural complex" by integrating architectural form and spatial layout. From the visual level to the thematic expression, it showcased the

professional level of architectural design. Although CGAN can reflect the theme of "urban cultural complexes", it is relatively simple in design-element diversity, using only a few decorative columns and simple geometric shapes for embellishment, and lacking multilevel spatial combinations. The diversity of design elements in WGAN-GP was superior to that of CGAN, which attempted to incorporate features such as elevated pedestrian walkways and public art installations. However, the use of color was clearly insufficient, and the overall rendering was quite dark, weakening the vitality that cultural complexes should have. The architectural renderings generated by DCGAN performed the worst in terms of design elements, colors, visual impact, and thematic expression. The architectural elements were disorderly and lacked connection, and the color matching was also abrupt. It failed to produce effective visual effects, making it difficult for the audience to quickly identify the core theme of the "urban cultural complex".

The above automation design scheme is not limited to creative advertising in buildings. Its intelligent layout and visual element optimization methods can also be applied to visual information communication in engineering and infrastructure. For example, when bidding for architectural design schemes or presenting them to the public, the architectural renderings, technical drawings, descriptive text, and project logos generated by CT-GAN can be intelligently typeset based on clarity, visual weight, and logical relationships to generate display images that are rigorous in layout, highlight key points, have a unified style, and conform to professional aesthetics, thereby improving communication efficiency and professionalism. In asset management, this technology can be used to generate standardized diagrams for facility maintenance manuals or operating procedures. The system can automatically integrate elements such as 3D model rendering, operation step text and safety warning signs for the device according to preset formats and hierarchical relationships, generating clear, easy-to-understand guidance images, thereby reducing training costs and minimizing the risk of maloperation caused by unclear information transmission.

4. Conclusion

The research proposes a systematic solution to the core problem of automated generation of print advertisements. The experimental results show that the proposed CT-GAN model achieves mIoUs of 0.8-0.9 and 0.5-0.7 on the FashionAd-5K and AdDesign-20K datasets, respectively, which are significantly better than those of baseline models such as CGAN, DCGAN, and WGAN-GP, demonstrating its accuracy in key-element localization. Meanwhile, CT-GAN exhibits excellent training stability and convergence speed, with its loss function rapidly stabilizing at around 0.05 (FashionAd-5K) and 0.07 (AdDesign-20K) after approximately 40 training cycles. In terms of generated image quality, CT-GAN also leads comprehensively in two objective indicators: Peak Signal-to-Noise Ratio (PSNR, 28.7 dB and 26.9 dB, respectively) and structural similarity (SSIM, 0.892 and 0.865, respectively). Managers will shift from relying on manual graphic advertising design and fragmented template tools to adopting the end-to-end CT-GAN automated generation system, thereby reconfiguring budget allocation, production workflows, cross-department collaboration, and large-scale personalized marketing delivery decisions. While the CT-GAN system demonstrates strong performance on e-commerce datasets, its generalizability across diverse cultural aesthetics and abstract brand campaigns remains to be validated. Future work will focus on incorporating reinforcement learning for adaptive design exploration and developing standardized APIs for seamless integration into industrial marketing automation workflows.

Author Contributions

Shan Gao contributed to conceptualization, methodology, analysis, investigation, draft preparation, and manuscript editing. Qiuhan Wang contributed to methodology, software, validation, data collection, visualization, and manuscript editing.

Funding

The research is supported by: North Minzu University Young Talents Cultivation Project, Supported by the Fundamental Research Funds for the Central Universities, North Minzu University, (NO. 2025QNPY46).

Institutional Review Board Statement

Not applicable.

Declaration of Artificial Intelligence (AI) Tools

The authors used AI tools solely for language editing and to improve readability. The authors reviewed and verified all content and take full responsibility for the accuracy and integrity of the manuscript.

References

- Bai, R. (2025). Multi-scale Query-based Transformer for Image Forgery Localization. *Signal, Image and Video Processing*, 19(13), 24-29. doi: 10.1007/s11760-025-04669-5.
- Bello, R. W. and Oladipo, M. A. (2024). Mask YOLOv7-based Drone Vision System for Automated Cattle Detection and Counting. *Artificial Intelligence and Applications*, 2(2), 115-125. doi: 10.47852/bonviewAIA42021603.
- Chen, L., Jing, Q., Tsang, Y., and Zhou, T. (2024). Iris: A multi-constraint graphic layout generation system. *Frontiers of Information Technology & Electronic Engineering*, 25(7), 968-987. doi: 10.1631/FITEE.2300312.
- di Summa, M., Mosca, N., Colella, L., Bianco, V., Ferraro, P., and Stella, E. (2023). Realistic interactions with micro-objects in a virtual environment. *AIA*, 12621(7), 206-210. doi: 10.1117/12.2678665.
- Johri, A., Singh, R. K., Kushwaha, B. P., Alhumoudi, H., and Alakkas, A. (2025). Leveraging open innovation for e-commerce success: The contingent role of accounting information systems and artificial intelligence. *Journal of Innovation & Knowledge*, 10(4), 100737. doi: 10.1016/j.jik.2025.100737.

- Leonardi, A. C. and Alfonsius, A. (2024). The influence of advertising, brand image, and product quality on purchase decisions of skincare beauty products in Medan. *Business and Entrepreneurial Review*, 24(1), 75-88. doi: 10.25105/v24i1.20661.
- Li, G., Jin, D., Yu, Q., Zheng, Y. and Qi, M. (2024a). MultiIB - TransUNet: Transformer with multiple information bottleneck blocks for CT and ultrasound image segmentation. *Medical Physics*, 51(2), 1-5. doi: 10.1002/mp.16662.
- Li, K., Xue, Y., Zhao, J., Li, H., and Zhang, S. (2024b). A cross-attention integrated shifted window transformer for remote sensing image scene recognition with limited data. *Journal of Applied Remote Sensing*, 18(3), 14-19. doi: 10.1117/1.JRS.18.036506.
- Liu, Y. (2023). Design of graphic design assistant system based on artificial intelligence. *International Journal of Information Technologies and Systems Approach*, 16(3), 1-13. doi: 10.4018/IJITSA.324761.
- Ma, J. (2024). Extraction and reconstruction of traditional art visual elements by graphic design incorporating deep learning. *Applied Mathematics and Nonlinear Sciences*, 9(1). doi: 10.2478/amns-2024-2450.
- Pan, H. and Liu, R. (2024). Mining research on correlation factors of residential electricity stability based on improved FP-growth algorithm. *Management of Environmental Quality: An International Journal*, 35(3), 547-566. doi: 10.1108/MEQ-05-2023-0142.
- Pang, B. and Liu, Y. N. (2024). PRO-SSRGAN: Stable super-resolution generative adversarial network based on parameter reconstructive optimization on Gaofen-5 remote-sensing images. *International Journal of Remote Sensing*, 45(9), 3022-3053. doi: 10.1080/01431161.2024.2339205.
- Panhe, H., Lingfeng, C., Zhiyuan, Z., and Zhen, L. (2025). Feature fusion CGAN based HRRP denoising and reconstruction method. *Chinese Journal of Electronics*, 35(1), 1-14. doi: 10.23919/cje.2025.00.022.
- Sultan, B. and Wani, M. A. (2024). Enhancing steganography capacity through multi-stage generator model in generative adversarial network based image concealment. *Journal of Electronic Imaging*, 33(3), 51-52. doi: 10.1117/1.JEI.33.3.033026.
- Wang, Q., Chen, X., and Su, H. (2025). Application of MLP-mixer-based image style transfer technology in graphic design. *International Journal of Advanced Computer Science & Applications*, 16(1), 112-115. doi: 10.14569/IJACSA.2025.0160159.
- Wang, X. and Liu, J. (2024). Spatially regularized leaky ReLU in dual space for CNN based image segmentation. *Inverse Problems & Imaging*, 18(6), 45-47. doi: 10.3934/ipi.2024016.
- Wang, Y., Wang, Y., and Wang, C. (2025). High-precision inverse design of multi-category grating couplers based on KAN-transformer. *Journal of Lightwave Technology*, 43(15), 7257-7264. doi: 10.1109/JLT.2025.3574780.
- Wang, Z. (2024). Design of pattern image generation technique based on self-attentive residual conditional generative adversarial network. *Journal of Radiation Research and Applied Sciences*, 17(4), 101157. doi: 10.1016/j.jrras.2024.101157.
- Xu, J., Xu, P., Yang, Z., Liu, F., Xu, L., and Lou, J. (2024). Freeform metasurface design with a conditional generative adversarial network. *Applied Physics A: Materials Science & Processing*, 130(8), 45-49. doi: 10.1007/s00339-024-07694-2.
- Yang, Y. (2024). Research on innovative practice and theoretical exploration of visual communication design for future media. *Applied Mathematics and Nonlinear Sciences*, 9(1), 9-11. doi: 10.2478/amns-2024-1669.
- Zahra, S. R. and Parisa, Z. (2024). From text to visual: Analyzing the influence of graphic design on English language storytelling. *i-manager's Journal on Humanities & Social Sciences*, 4(2), 50-51. doi: 10.26634/jhss.4.2.20847.



Shan Gao is dedicated to the education, mentorship, and creative development of students pursuing disciplines in design and the visual arts. His teaching and institutional work align with North Minzu University's commitment to advancing cultural design, minority heritage preservation, and contemporary artistic practices within the Ningxia region. Through his academic role, he contributes to regional design initiatives and helps prepare the next generation of creative professionals for the industry.



Qiuhan Wang is a master's student at the School of Design and Arts, North Minzu University. Her research interests include digital media technology and visual communication design.