International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 7343-7355

https://doi.org/10.70517/ijhsa464636

Housing Brand Image Design in Digital Transformation: Visual Communication Innovation in Virtual Exhibition Platforms

Bing Shen[1,*]
[1] School of Fine Arts and Art Design, Luoyang Normal University, Luoyang, Henan, 471000, China
Corresponding authors: (e-mail: shenbing_lynu@163.com).

Abstract With the widespread application of digital intelligent technologies, rebranding has become both a challenge and an opportunity for businesses. This paper first conducts a questionnaire survey to understand the relationship between housing brand image and home-buying decisions. Based on housing brand image design in the context of digital transformation, it innovatively proposes a text-to-image generation method that utilizes a single-stage generative adversarial network (GAN) structure combined with a deep attention mechanism, starting from the application of AIGC technology in brand image design. Analysis reveals that housing brand image significantly influences consumers' home-buying choices, accounting for over 70% of decisions, while over 80% of consumers perceive a connection between housing brand image and construction quality, underscoring the importance of housing brand image design. The experimental results of the designed text-to-image model, including IS, Accuracy, and FID, outperform the comparison methods. Specifically, the IS value and Accuracy value improved by 13.79% to 28.12% and 3.44% to 62.66%, respectively, while the FID value decreased by 2.63% to 36.52%, demonstrating its excellent performance in generating housing brand image design logo images. Therefore, this method can be applied to housing brand image design to promote the intelligence of brand image design and its display and dissemination on virtual platforms.

Index Terms digitalization, generative adversarial networks, image generation, brand image design

I. Introduction

As the housing industry has become increasingly market-oriented, competition within the sector has intensified, and consumers have grown more discerning in the face of market competition. With the maturation of the market economy, a product's brand image has become the primary factor influencing consumer perception. The next phase of competition will be a battle of brands, and housing, as a unique commodity, is no exception [1], [2]. After more than a decade of trials and tribulations, the housing industry is currently undergoing a transformative shift in its marketing model. The most significant manifestation of this shift is the growing emphasis on the quality, taste, and service functions of real estate projects, with a brand marketing approach that focuses on shaping the company's own brand to achieve market differentiation quietly emerging in the market [3]-[5]. Literature [6] indicates that the marketing performance of branded apartments significantly outperforms that of non-branded apartments, and constructs a dynamic model for apartment brand management, providing guidance for brand loyalty maintenance and strategic decision-making. The image shaping of housing brands is crucial for corporate development.

A successful real estate brand image can help companies attract more customers and enhance market competitiveness [7], [8]. Literature [9] focuses on middle-class consumers, analyzing their purchasing intentions under factors such as housing word-of-mouth, price perception, brand image, and brand trust, which significantly influence their purchasing intentions. Literature [10] reports that among the primary factors influencing consumers' housing purchase decisions in Indonesia, consumers prioritize property location and brand image over price factors. Literature [11] uses structural equation modeling to find that the Murabahah contract and property location influence consumers' housing purchase decisions through the mediating effect of brand image. Faced with increasingly intense market competition, housing enterprises must seriously consider how to design their brand image, establish their dominant position in the real estate market, and become winners in the future housing market.

With the digital transformation of various industries, real estate companies have also entered the wave of digital marketing. However, online marketing revenue accounts for less than 10% of total revenue, indicating that online conversion rates are insufficient. This is because most real estate companies' digital marketing of housing properties is presented solely through simple videos, images, and textual descriptions. Consumers may feel fatigued during the browsing process and lack an immersive experience [12]-[14]. The data analysis in Reference [15] shows that social media marketing, brand image, and the combined effect of the two can positively influence consumers' willingness to purchase residential apartments. Reference [16] applied partial least squares analysis and found that digital marketing and word of mouth influence consumers' willingness to purchase through the mediating effect of brand image. Reference [17] confirmed that low-immersion virtual reality technology improves the online profitability of real estate by enhancing the consumer shopping experience, with differences observed across different property prices. Reference [18] explores the application of augmented reality and virtual reality in online housing displays, using a panoramic camera to stitch together the entire three-dimensional space of a residence, enabling consumers to view rooms from various angles and thereby enhancing their purchasing intent. Reference [19] found in its exploration of a family home-buying decision-making model that virtual reality technology-based online real estate marketing can promote offline sales, reduce marketing efforts, narrow transaction gaps, and thereby enhance housing purchase decisions. Literature [20] introduces the effectiveness of virtual reality technology in enhancing brand experience and designing brand imagery. By breaking through spatial and temporal boundaries to enhance consumer experience and combining sensory dimensions in brand design, it constructs a brand imagery recognition and evaluation system to achieve brand imagery construction. Therefore, there is an urgent need for virtual housing display platforms to shape brand imagery and enhance corporate brand competitiveness.

This paper conducts a questionnaire survey of potential homebuyers to explore consumers' perceptions of housing brand image and the relationship between brand image and home purchase decisions. Against this backdrop, digital technology is introduced into housing brand image design, with a focus on the application of AIGC technology in housing brand image design, and an image model is designed based on this text output. Addressing the current issue where generative adversarial networks (GANs) struggle to fully integrate textual information, leading to generated images that lack clarity and diversity, this study uses the DF-GAN algorithm as a baseline model. It introduces channel attention and pixel attention mechanisms into the generator, employing a pair of generators and discriminators based on residual network structures for adversarial training, thereby achieving the DA-GAN model. Finally, experiments were conducted on the COCO-stuff dataset, comparing the proposed method with other image output methods in terms of IS, FID, and accuracy test results at different resolutions. The IoU, R@0.5 and R@0.3 results of each model were analyzed to evaluate the image generation performance of the proposed method.

II. Housing brand image and consumer survey

II. A. Visual Image in Housing Brands

With the deepening development of the market economy, the real estate market has undergone significant changes, transitioning from a situation where properties were in high demand to one of fierce market competition. Developers must adopt a comprehensive brand image design system to leverage their brand advantages and implement marketing strategies in the market.

Brand building in the real estate industry has its own unique characteristics compared to other industries, but fundamentally, it still aims to pursue its own brand proposition as the primary objective. The significance of developers creating real estate brands lies in forming a distinct architectural philosophy to differentiate themselves from other projects. For example, Vanke's “family-oriented,” SOHO's “fashionable,” and Country Garden's “elegant” have all established their own brand positioning. Visual identity design is closely intertwined with the creation of housing brands. Visual identity design begins before a development is completed, showcasing a distinctive project image and infusing the project with a concept to capture the attention of ordinary homebuyers. To stimulate homebuyers' imagination and purchasing motivation, housing brands must establish a comprehensive visual identity. The taste and spiritual attributes of a development are largely conveyed through its visual identity.

II. B. Survey Analysis of Potential Home Buyers

In order to provide a reference for housing brand image design, this paper has developed a survey questionnaire on housing brand visual image. The target audience for the survey is primarily potential homebuyers, with a total of 200 participants, ensuring a certain degree of representativeness. After providing thorough explanations of the professional terminology mentioned in the questionnaire, the respondents completed the survey, resulting in the collection of 182 valid responses. The survey results of potential homebuyers were then compiled and analyzed.

II. B. 1) Key considerations when purchasing a home

As shown in Figure 1, consumers' priorities when purchasing a home are as follows. Among the given options, cost-effectiveness remains the most important factor for homebuyers, with 39.01% of consumers prioritizing cost-effectiveness when purchasing property. Consumers aim to maximize the value of their investment. The percentage of respondents prioritizing visual imagery is 4.95%. On the surface, this suggests that real estate visual imagery is not highly valued by consumers. However, this masks a significant untapped market opportunity in visual imagery design. There is substantial room for growth in the visual imagery of housing brands. As consumers become more educated and home purchases increasingly rational, the importance of housing brand imagery as an added value of real estate products will continue to rise.

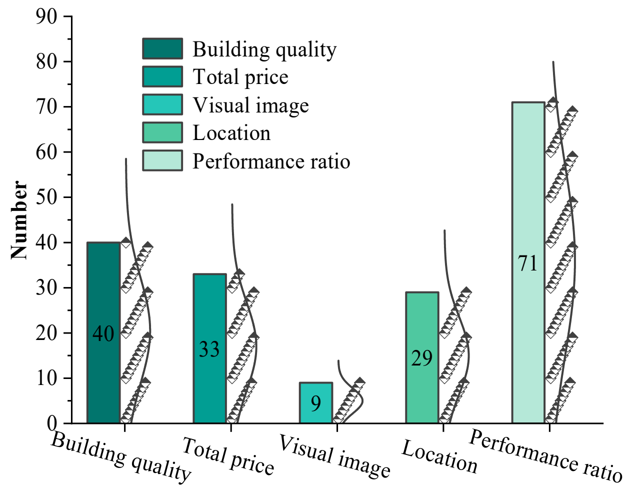


Figure 1: Consumer concerns when buying a house

II. B. 2) Brand Image and Home Purchase Choices

The survey results on the impact of housing brand image on home purchasing decisions are shown in Figure 2. Most consumers still recognize the significance of housing brand image. 73.08% of consumers believe that housing brand image will influence their home purchasing decisions, and 18.68% of consumers believe that it is very important, indicating that housing brand image has begun to drive real estate sales.

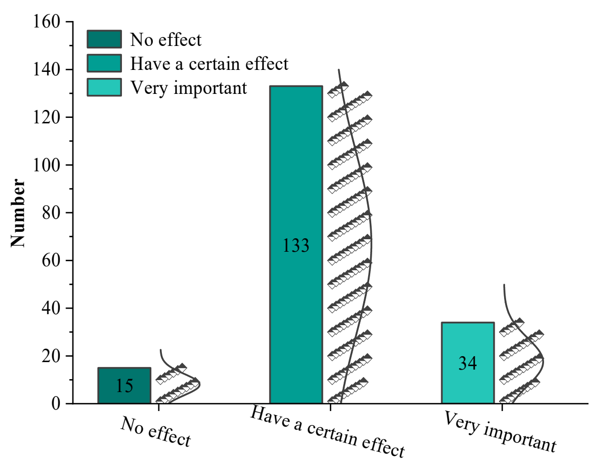


Figure 2: The influence of the housing brand image on home purchase selection

II. B. 3) Brand Image and Construction Quality

The relationship between housing brand image and construction quality is shown in Figure 3. 62.09% of consumers believe that there is a certain correlation between housing brand image and construction quality, while 20.33% of consumers believe there is a significant correlation. Most people believe that the quality of a housing project is related to the quality of its brand visual image. This is a crucial piece of information that developers should take seriously. They should abandon their previous low-key promotional strategies and, while ensuring construction quality, invest more effort into enhancing their brand visual image.

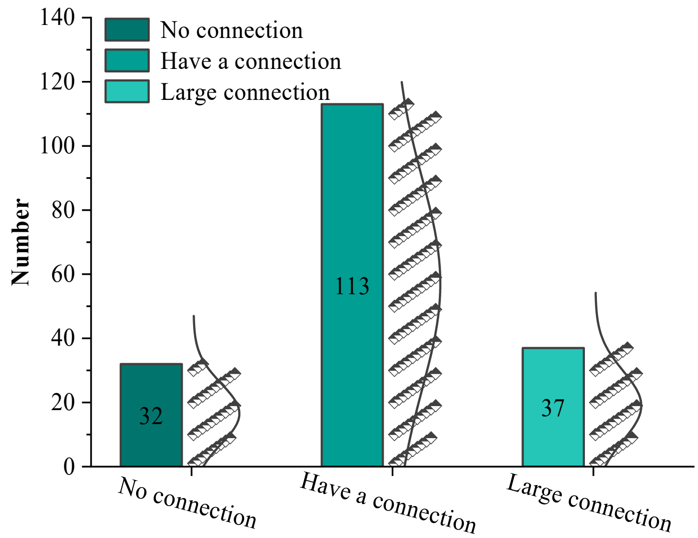


Figure 3: The relationship between the home brand image and the building quality

II. B. 4) Visual Image Focus Points

Homebuyers' personal preferences for visual imagery are shown in Figure 4. Among the four options of color, pattern, font, and application system, the application system received the most attention, accounting for 40.11%, which is closely related to the context in which the visual imagery appears. Color came in second, accounting for 28.02%. In terms of logo design, color is the most attention-grabbing element for consumers, so designers should give more consideration to consumers' psychological acceptance of color schemes.

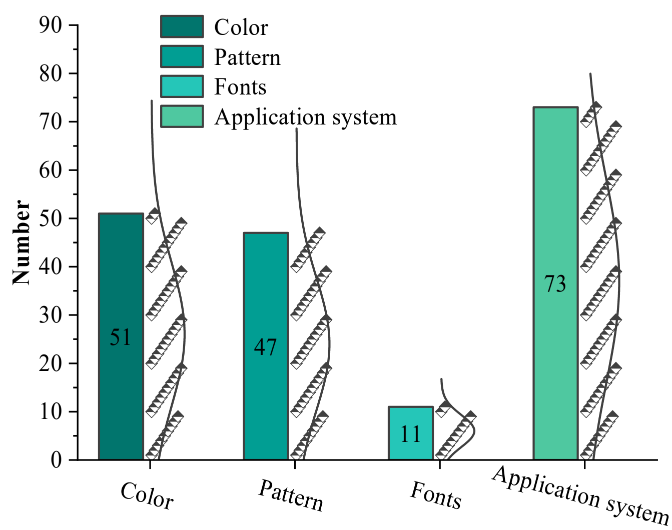


Figure 4: Homebuyers' preferences for visual image

II. B. 5) Visual Image Style

Consumers' preferences for various housing brand visual image styles are shown in Figure 5. Fresh and elegant and luxurious and magnificent are the two most popular housing brand visual image styles among consumers, accounting for 20.88% and 18.68% respectively. In contrast, consumers have lower preferences for creative and nostalgic retro visual image styles, accounting for 13.74% and 9.89% respectively.

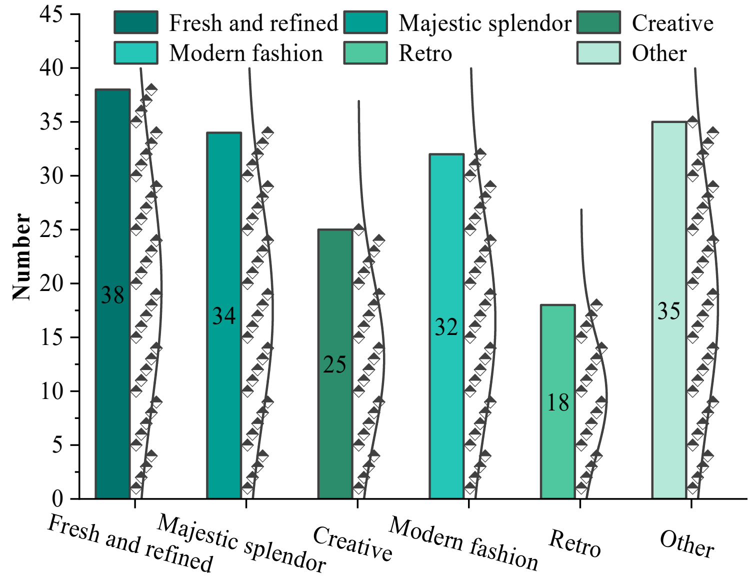


Figure 5: Consumer preference for visual image style

III. Digital-based housing brand image design

The immediacy and global reach of the internet have accelerated the speed and expanded the scope of information dissemination, presenting brands with both greater opportunities and challenges in their communication efforts. People increasingly prefer digital, dynamic, and real-time interactive methods for accessing and exchanging information. The widespread use of the internet and mobile devices has led to information overload, causing audience attention to become fragmented. Brand communication now faces the challenge of information fragmentation, and brand image design must adapt to a diverse communication environment. Based on the survey in the previous chapter, housing brand design is closely related to consumers' home-buying choices. This chapter explores housing brand image design using digital technology.

III. A. Innovative Applications of Digital Technology

With the rapid rise of the digital age, technological innovation has left its mark on the design industry. By adopting the latest digital technologies, companies can continuously innovate their brand image and product services, achieve brand differentiation, and maintain a competitive edge. Housing brand image design is the quality, image, and lifeblood of a real estate company. By closely integrating their strengths with digital technology, real estate companies will undergo transformation, transition, and upgrading during the shift from an industrial society to a digital intelligent society. Their brands will embody intelligent quality, digital imagery, and robust vitality. Current design trends often develop in tandem with technological advancements and market changes. The emergence of technologies like generative AI has disrupted everything. Tools like DALL-E, Midjourney, and StableDiffusion are redefining design rules.

AIGC technology can intervene in the housing brand identity system in many aspects, such as logo design. As we all know, the design style of the logo is diverse, seemingly simple, but not easy to grasp, concise graphics should not only reflect the spirit of enterprise, but also beautiful and generous, easy to identify. In the past, although some online logo generation websites claimed to be able to produce logos with one click, the essence was still to randomly combine the graphics and text in the material library, and the repetition frequency was high, which was prone to infringement incidents. The image function of AIGC technology can truly realize the mass logo design scheme for housing brands, and you can get the corresponding logo design by entering prompt words in the Wensheng diagram function of AIGC. In addition to logo design, AIGC technology can also generate auxiliary patterns for housing brands, which can be widely used in posters, business cards, packaging, flags, handbags and other application systems, which greatly enriches the visual language of housing brand image design and strengthens the brand personality characteristics. Based on this, this paper applies AIGC technology to the design of housing brand image, and constructs a text-generated image model based on generative adversarial network, so as to facilitate the intelligent design of housing brand image and its display on the virtual platform.

In models that use stacked generative adversarial networks for text-to-image generation, the final generated images tend to resemble a combination of various text attributes, lacking visual realism and making model training difficult and prone to divergence. This paper improves upon the DF-GAN model and proposes a text-to-image generation method called DA-GAN that combines a channel-attention mechanism.

III. B. DF-GAN Model

This paper conducts a series of research studies based on the single-level structure of the deep fusion generative adversarial network DF-GAN, which consists of a generator G and a discriminator D.

III. B. 1) Generators and Discriminators

The deep fusion module in generator G uses a residual structure to avoid performance degradation caused by excessive network depth, represented by function $F(\cdot)$, as shown in Equation (1):

$$F(h) = ResG(h) + CNN(h) \tag{1}$$

The residual function $ResG(\cdot)$ consists of two fusion modules and two convolutional layers. The convolutional layers are used to change the channel dimension of the input feature h while keeping the spatial size unchanged, as the affine transformations in the fusion modules cannot alter the channel dimension of the features. $CNN(\cdot)$ is a convolutional layer (with a kernel size of 1×1), ensuring that the channel dimension of the input feature h matches that of the output feature $ResG(h)$, enabling their addition. Each fusion module requires two affine transformations. Each affine transformation requires two multi-layer perceptrons ($\omega - MLP(\cdot)$ and $\beta - MLP(\cdot)$) to learn the affine parameters ω and β from the text encoding s expanded by Gaussian noise z.

The affine transformation algorithm in DF-GAN can be expressed as follows:

$$\omega_i = \omega - MLP(s \oplus z) \tag{2}$$

$$\beta_i = \beta - MLP(s \oplus z) \tag{3}$$

$$AFF(h_i \mid s, z) = \omega_i \cdot h_i + \beta_i \tag{4}$$

Among them, h_i is the hidden image encoding of the current input, i represents the i th affine module, ω_i and β_i are the affine parameters to be learned by the current affine module, and $\oplus$ is the encoding concatenation operation. $\omega - MLP(\cdot)$ and $\beta - MLP(\cdot)$ are multilayer perceptron functions, and $AFF(\cdot)$ is the affine transformation function.

The discriminator D consists of a convolutional layer and several downsampling residual layers. First, the real or generated image is transformed into an image encoding through the convolutional layer, then feature extraction is performed through six downsampling layers, each of which is a residual module. After feature extraction, the features are concatenated with the text encoding s, and finally, the concatenated result is input into a convolutional layer for semantic alignment. The larger the output value, the more similar the semantics. The downsampling layers in the discriminator D use residual structures to effectively avoid performance degradation caused by excessive network depth. The downsampling residual layer is represented by the function $H(\cdot)$, as shown in Equation (5):

$$H(v) = ResD(v) + Avg(CNN(v)) \tag{5}$$

The residual function $ResD(\cdot)$ consists of a convolution layer with a kernel size of 4×4 and a convolution layer with a kernel size of 3×3. $CNN(\cdot)$ is a convolution layer (with a kernel size of 1×1), and $Avg(\cdot)$ is an average pooling downsampling layer (downsampling by halving the side length).

III. B. 2) Network Loss Function

The loss function of DF-GAN includes the generator loss function L_G and the discriminator loss function L_D. The generator loss function L_G is solely composed of the discriminator D's score for the generated fake image g. The higher the score, the smaller the loss. The discriminator loss function L_D consists of the conditional loss function L_{cond} and the matching-aware gradient penalty loss function L_{MAGP}. The matching-aware gradient penalty loss function L_{MAGP} applied in DF-GAN significantly enhances the stability of GAN, minimizing the occurrence of mode collapse. It also enables the discriminator D to perform better semantic alignment, thereby allowing the generator G to output higher-quality generated images g. This eliminates the generator G's reliance on additional fixed networks, and compared to additional fixed networks, the matching-aware gradient penalty loss function L_{MAGP} requires less computational effort. In DF-GAN, analysis revealed that the unconditional loss function L_{umcond} slows down text-image semantic alignment, so the unconditional loss function L_{uncond} is not used here. The conditional loss function L_{cond} is composed of scores assigned by the discriminator D to real images r, generated image g, and mismatched image m, with corresponding loss functions L_r, L_g, and L_m. The higher the score for the real image r, the lower the scores for the generated image g and mismatched image m, resulting in a smaller loss. The output value of the hinge loss function Hinge_loss is restricted to [-1, 1]. For the real image r, when the score is greater than or equal to 1, the loss is 0. For the generated image g and the mismatched image m, when the score is less than or equal to -1, the loss is 0. This allows the network to discard some unnecessary experience and learn more valuable experience.

The generator loss function L_G is defined as follows:

$$L_G = -E_{G(z,s)\sim P_g}\left[D(G(z,s),s)\right] \tag{6}$$

where $G(z,s)$ denotes the image generated by the generator G using random noise z following a standard normal distribution and text encoding s, P_g denotes the set of generated images, and $D(G(z,s),s)$ denotes the discriminator D performing semantic alignment between the generated image and its corresponding text encoding s.

The discriminator loss function L_D is given by the following equation:

$$L_D = L_{cond} + L_{MA-GP} \tag{7}$$

$$L_{cond} = L_r + L_g + L_m \tag{8}$$

$$L_r = -E_{x\sim P_r}[\min(0, -1 + D(x,s))] \tag{9}$$

$$L_g = -\frac{1}{2}\cdot E_{G(z,s)\sim P_g}[\min(0, -1 - D(G(z,s),s))] \tag{10}$$

$$L_m = -\frac{1}{2}\cdot E_{x\sim P_m}[\min(0, -1 - D(x,s))] \tag{11}$$

$$L_{MA-GP} = k\cdot E_{x\sim P_r}[(\|\nabla_x D(x,s)\| + \|\nabla_s D(x,s)\|)^p] \tag{12}$$

where x denotes the real image, P_r denotes the set of real images that correctly match the text encoding s, and P_m denotes the set of real images that incorrectly match the text encoding s. $D(x,s)$ denotes the discriminator D performing semantic alignment between the real image x and its corresponding text encoding s, ∇_x and ∇_s denote the partial derivatives of the discriminator D with respect to x and s, respectively, k and p are used to balance the gradients in the matching-aware gradient penalty L_{MAGP}, and $\min(\cdot)$ denotes the minimum value function.

III. C. DA-GAN Model

III. C. 1) Generator based on channel-pixel attention mechanism

The generator takes text representations and random noise as input and includes seven upsampling modules to generate visual feature maps of different sizes. Each upsampling module contains two convolutional layers, two affine transformation layers, and two channel-space attention layers.

The generator progressively generates high-resolution images from the initial input text representation and random noise through each upsampling module. This process can be expressed by Equations (13) to (16):

$$h_0 = F_0(z) \tag{13}$$

$$h_1 = F_1^{UP}(h_0, s) \tag{14}$$

$$h_i = F_i^{UP}(h_{i-1}, s), \quad i = 2,3,\ldots,7 \tag{15}$$

$$o = G_c(h_7) \tag{16}$$

Among these, z is a noise vector sampled from a standard normal distribution, and F_0 is a fully connected layer whose purpose is to map the random noise into a three-dimensional tensor that can be concatenated with the text representation. F_i^{UP} is the upsampling module proposed in this paper, and G_c is the final sampling layer, which generates a 256-resolution image from the final hidden state h_7. h_0 is the hidden representation output by the fully connected layer F_0, h_1-h_7 are the hidden features output by each upsampling module, and o refers to the final generated 256×256 resolution image.

(1) Affine layer

Before implementing the two attention operations, the feature maps are first fused with text and image information through affine transformations. Two fully connected layers MLP with one hidden layer are used to predict the channel scaling parameter γ and the channel scaling offset parameter β from the sentence vector s:

$$\gamma = MLP_1(s), \quad \beta = MLP_2(s) \tag{17}$$

Assume that the input feature map is represented as $X \in \mathbb{R}^{B\times C\times H\times W}$, where B,C,H,W represent the batch size, number of channels, and width and height of each feature map, respectively. Then, the fully connected layer MLP outputs a vector of length C. First, a channel-wise scaling operation is performed on X using the scaling parameter γ, followed by a shift operation on χ using the shift parameter β, as shown in Equation (18):

$$AF(x_i \mid s) = \gamma_i \dot{x}_i + \beta_i \tag{18}$$

Among them, AF is an affine transformation, x_l is the ith channel in the visual feature map, s is the sentence representation vector, and γ_i and β_i are the scaling and shift parameters of the ith visual channel feature in the feature map, respectively.

(2) Channel attention

The channel attention layer has two inputs: feature map h and text sentence representation s. First, perform global average pooling and global max pooling on feature map h. Let GAP and GMP denote average pooling and max pooling, respectively:

$$x_a = GAP(h) \tag{19}$$

$$x_m = GMP(h) \tag{20}$$

Among these, $x_a \in \mathbb{R}^{C\times1\times1}$ and $x_m \in \mathbb{R}^{C\times1\times1}$ are the channel features extracted after average pooling and max pooling, respectively, where C denotes the number of channels in the feature map h.

Then, queries, keys, and values are set to capture the semantic relevance between channels and input text. Here, x_a and x_m are used as queries, and s is set as the key and value. The definitions are as follows:

$$q_a = W_a x_a, \quad q_m = W_m x_m \tag{21}$$

$$k_c = W_k s, \quad v_c = W_c s \tag{22}$$

Among them, W_a, W_m, W_k and W_v are projection matrices obtained through 1×1 convolution. The resulting q_α, q_m are meaningful visual features extracted based on average pooling and global pooling. Subsequently, the channel attention weights are calculated as shown in Equations (23) to (25):

$$\tilde{\alpha}_a^c = q_a \cdot k_c^T, \quad \tilde{\alpha}_m^c = q_m \cdot k_c^T \tag{23}$$

$$\alpha_a^c = softmax(\tilde{\alpha}_a^c \cdot v_c) \tag{24}$$

$$\alpha_m^c = softmax(\tilde{\alpha}_m^c \cdot v_c) \tag{25}$$

Among them, $\tilde{\alpha}_a^c$ and $\tilde{\alpha}_m^c$ represent the semantic similarity between channel feature maps and sentence representations. $\alpha_a^c \in \mathbb{R}^{C\times1\times1}$ and $\alpha_m^c \in \mathbb{R}^{C\times1\times1}$ represent the attention weights for each channel after average pooling and max pooling, respectively. After obtaining the attention weights for each channel, the attention weights are multiplied by the original feature map to update the feature map:

$$o_a = \alpha_a^c \otimes h \tag{26}$$

$$o_m = \alpha_m^c \otimes h \tag{27}$$

Here, $\otimes$ denotes element-wise multiplication. By updating the feature map through channel attention, the generator network focuses on convolutional channels that are more semantically relevant to the given text description. At the same time, an adaptive gate network is used to fuse the results of average pooling and max pooling, as shown in Equations (28) and (29):

$$g^c = \sigma(W_1 x_a + W_2 x_m) \tag{28}$$

$$o_c = g^c * o_a + (1 - g^c) * o_m \tag{29}$$

Among them, W_1 and W_2 are learnable matrices, and σ is a sigmoid function. Finally, new feature maps h' are generated by adopting adaptive residual connections, as shown in formula (30):

$$h' = \gamma_c * o_c + h \tag{30}$$

Among them, γ_c is a learnable parameter initialized to 0.

(3) Pixel attention

For a given visual feature map h_1 and global sentence representation s, first perform average pooling and max pooling operations on h_1:

$$e_a = SAP(h_1) \tag{31}$$

$$e_m = SMP(h_1) \tag{32}$$

Among them, SAP and SMP are the average pooling and max pooling of the feature map space dimension, respectively, and $e_a \in \mathbb{R}^{1\times H\times W}$ and $e_m \in \mathbb{R}^{1\times H\times W}$ are the new feature maps after pooling. Then, the sentence representation s is used to obtain keys and values through 1×1 convolution:

$$k_p = W_k s, \quad v_p = W_v s \tag{33}$$

Where $W_k \in \mathbb{R}^{D\times C}$ and $W_v \in \mathbb{R}^{D\times C}$ are matrices learned through 1×1 convolution. Then, the new visual feature maps e_m and e_a are respectively dot-multiplied with k_p to obtain the semantic similarity measures $\tilde{\alpha}_a^p$ and $\tilde{\alpha}_m^p$. The semantic similarity measures are then multiplied by v_p and passed through a softmax function to ultimately obtain the spatial-dimensional attention weights:

$$\tilde{\alpha}_a^p = e_a \cdot k_p^T, \quad \tilde{\alpha}_m^p = e_m \cdot k_p^T \tag{34}$$

$$\alpha_a^p = softmax(\tilde{\alpha}_a^p \cdot v_p) \tag{35}$$

$$\alpha_m^p = softmax(\tilde{\alpha}_m^p \cdot v_p) \tag{36}$$

Where α_a and α_m represent the final attention weights of the spatial pixels after average pooling and max pooling.

Next, matrix multiplication is performed between the attention weights and the initial feature maps to derive the new feature maps o_a and o_m:

$$o_a = \alpha_a^p \otimes h_i \tag{37}$$

$$o_m = \alpha_m^p \otimes h_1 \tag{38}$$

Additionally, o_a and o_m are cascaded, and a nonlinear function σ is used to calculate the final result o_p. Finally, an adaptive residual connection is used to connect h_1 and o_p to obtain the final visual feature map $\hat{h}_1$:

$$o_p = \sigma(W_0[o_a; o_m]) \tag{39}$$

$$\hat{h}_1 = \gamma_p * o_p + h_1 \tag{40}$$

Among them, W_0 is the parameter matrix learned through 1×1 convolution, σ is the ReLU function, and γ_p is a learnable parameter initialized to 0.

III. C. 2) Unidirectional discriminator

Since this paper uses matching-aware gradient penalty (MA-GP) loss in the discriminator, the discriminator can converge without bidirectional discrimination. Therefore, this paper adopts a unidirectional discrimination mechanism from text to image, where the discriminator concatenates image features and text features, and then outputs a adversarial loss through two convolutional layers.

III. C. 3) 3 Loss Function

The generator can apply zero-centered gradient penalties on real data to synthesize more realistic images. The representation of unconditional zero-centered gradient penalties on real data is as follows:

$$\mathrm{L}_{0-GP} = k\mathrm{E}_{x\sim P_r}[\| \nabla_x D(x) \|]^p \tag{41}$$

Where k and p are two hyperparameters used to balance the effectiveness of gradient punishment, and p_r is the true data distribution.

As for the adversarial loss, hingeloss is used to stabilize the training process, so the discriminator adversarial loss is:

$$\begin{aligned}\mathrm{L}_D^{adv} = &-\mathrm{E}_{x\sim P_r}[\min(0,-1+D(x,s))] \\ &-\frac{1}{2}\mathrm{E}_{G(z)\sim P_s}[\min(0,-1-D(G(z),s))] \\ &-\frac{1}{2}\mathrm{E}_{x\sim P_{mi,i}}[\min(0,-1-D(x,s))]\end{aligned} \tag{42}$$

Among them, x is an image sampled from the real data set, $G(z)$ is an image generated by the generator, z is a noise vector sampled from a Gaussian distribution, s is the text representation vector. P_g, P_r, P_{mis} refer to the distribution of generated image data, the distribution of real image data, and the distribution of real images that do not match the text description, respectively. The corresponding generator adversarial loss:

$$\mathrm{L}_G^{adv} = -\mathrm{E}_{G(z)\sim P_t} D(G(z),s) \tag{43}$$

At the same time, the loss function for the matching-aware gradient penalty applied to text-generated images can be obtained from the zero-center gradient penalty (0-GP) formula (41):

$$\mathrm{L}_{MA-GP} = k\mathrm{E}_{x\sim P_r}[(\| \nabla_x D(x,s) \| + \| \nabla_s D(x,s)) \|]^p \tag{44}$$

In which k and p are set as hyperparameters in the network to 2 and 6, respectively. Therefore, the loss function L_G of the generator in the generative adversarial network is given by equation (45):

$$\mathrm{L}_G = \mathrm{L}_G^{adv} \tag{45}$$

The total loss function L_D of the discriminator is:

$$\mathrm{L}_D = \mathrm{L}_D^{adv} + \mathrm{L}_{MA-GP} \tag{46}$$

III. D. Experimental Results and Analysis

In this section, we conduct experiments to validate the effectiveness of the proposed model. We present the dataset and training settings used in the experiments, and compare the generated images with other models in terms of quality and numerical values.

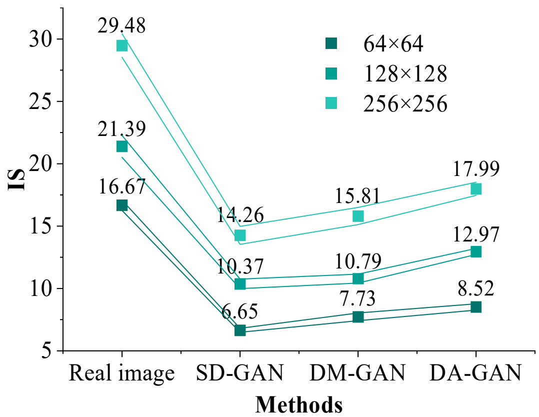
III. D. 1) Dataset and Training Settings

To validate the model's effectiveness, experiments were conducted on the COCO-stuff dataset. This dataset contains approximately 25,000 training images, 1,000 validation images, and 2,000 test images.

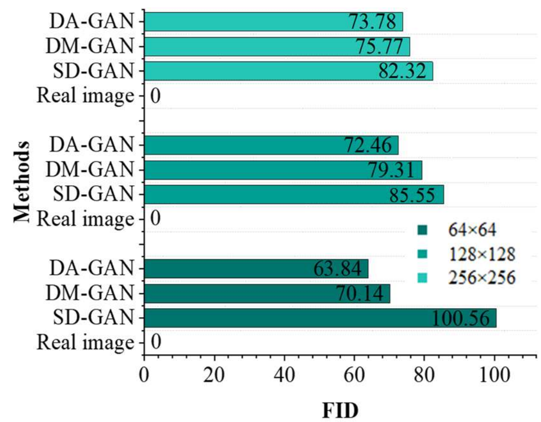
The experimental environment was Python 3.7, with 12 processor cores, 16 GB of memory, and the Ubuntu 18.04 operating system, along with a GeForce GTX 1080 Ti graphics card. The research plan was based on the open-source deep learning framework PyTorch and visualized using TensorBoardX.

III. D. 2) Model comparison experiment

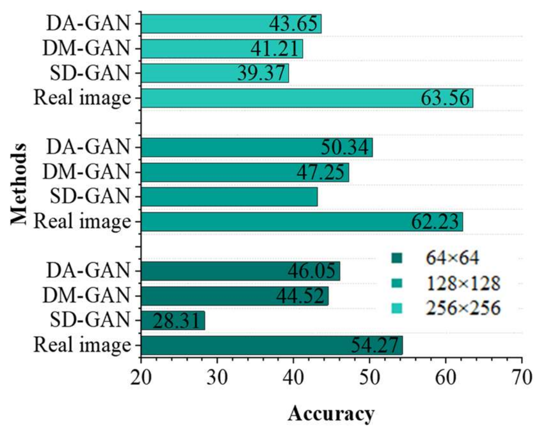
The proposed method was compared with DM-GAN and SD-GAN using evaluation metrics such as IS, FID, and accuracy. The metric results for each model at different resolutions are shown in Figure 6. Regardless of the resolution (64×64, 128×128, or 256×256), the proposed method demonstrates a certain improvement in quality compared to the comparison models. A higher IS value, lower FID, and higher accuracy indicate that the images generated by the proposed method are more similar to real images. At resolutions of 64×64, 128×128, and 256×256, the IS values of the images generated by the proposed method improved by 28.12% and 10.22%, 25.07% and 20.20%, and 26.16% and 13.79%, respectively. The FID values decreased by 2.63% to 36.52%, and the Accuracy values increased by 3.44% to 62.66%. Table 1 shows the values of each model in terms of IoU and R@0.5 and R @ 0.3. Higher IoU and R@0.5 and R values at 0.3 indicate greater accuracy in the size and position of the generated targets. The proposed method achieves improvements in both IoU and R@0.5 and R at 0.3, with increases of 58.54% and 25.00%, 20.00% and 10.53%, and 26.42% and 8.06%, respectively. The generated images produced by the proposed DA-GAN text-to-image generation method exhibit good quality and largely meet the designer's requirements, making them suitable for use in the visual identity design of housing brand imagery, thereby enhancing the efficiency of housing brand image design.



(a) IS



(b) FID



(c)Accuracy

Figure 6: The index results of each model in different resolutions

Table 1: Values of each model on IoU, R@0.5 and R@0.3

Methods	IoU	R@0.5	R@0.3
SD-GAN	0.41	0.35	0.53
DM-GAN	0.52	0.38	0.62
DA-GAN	0.65	0.42	0.67

IV. Conclusion

Digital design of brand imagery involves integrating traditional brand imagery into the information age, which is characterized by openness, virtuality, and interactivity. This paper conducts a questionnaire survey based on housing brand imagery and consumer needs to understand the preferences of potential buyers. Further analyzing the digital transformation of housing brand imagery, this study applies AI-generated technology to design a text-image generation model and conducts experimental validation. Through consumer surveys, over 70% of respondents acknowledged that housing brand imagery influences their home-buying decisions, and over 80% believe that housing construction quality is associated with housing brand imagery. Consumers place greater emphasis on the application systems of brand imagery, preferring fresh and elegant styles as well as opulent and grand styles. These survey results can provide reference for the design of housing brand image and its dissemination on virtual display platforms. Additionally, the text-image generation method designed in this paper outperforms the comparison method in all evaluation metrics, with the IS value and accuracy value of the generated images improving by 13.79% to 28.12% and 3.44% to 62.66%, respectively. While FID values decreased by 2.63% to 36.52%. This indicates that the method can generate high-quality housing brand identity images that meet the visual design requirements for housing brand identities. The digital design expression and promotion of housing brand identity systems not only meet consumers' demands for brand identity digitization but also align with the trend of enhancing brand identities in the digital age.

About the Author

Bing Shen (1989.07-), female, Han, from Heze, Shandong province, lecturer, School of Fine Arts and Art Design, Luoyang Normal University, main research direction is cultural innovation design and communication.

References

[1] Setyaningsih, R., Wati, I., Al Ghifari, M., & Sari, A. I. (2024). Branding Image Management As a Strategy To Enhance Competitiveness. Jurnal At-Tarbiyat: Jurnal Pendidikan Islam, 7(3).

[2] Sihombing, J. K., Octavia, A., & Fitriaty, F. (2024). The INFLUENCE OF BRAND IMAGE AND SERVICE QUALITY ON THE DECISION TO PURCHASE A HOUSE WITH MODERATING EFFECT OF LIFESTYLE: A CASE STUDY ON MORTGAGE LOANS AT BANK TABUNGAN NEGARA BRANCH OFFICE, KABUPATEN BUNGO. JOURNAL OF BUSINESS STUDIES AND MANAGEMENT REVIEW, 8(1), 56-62.

[3] Alptekin, G. A. M. Z. E. (2022). Real Estate 4.0 & brand ('ed housing) positioning. Journal of Construction Engineering, Management & Innovation (Online), 5(4).

[4] Ngoc, N. M., & Tien, N. H. (2021). Branding strategy for gamuda land real estate developer in ho chi minh city vietnam. celadon city project. Psychology and education, 58(5), 3308-3316.

[5] Ding, L., Yu, W., & Shang, C. (2021). Research on the promotion of efficient and energy-saving architecture in brand management. Sustainable Computing: Informatics and Systems, 30, 100503.

[6] Choi, M., Park, M., Lee, H. S., & Hwang, S. (2017). Dynamic modeling for apartment brand management in the housing market. International Journal of Strategic Property Management, 21(4), 357-370.

[7] Yöreş, F. Ç. (2017). Celebrity Endorsement: Developing A Brand Image in Housing Ads. Brand Management: Advancing Insights on, 175.

[8] Suryo, P. A. D., Gde, S. I. P., Made, W., & Ketut, G. I. G. (2023). the Mediating Role of Customer Satisfaction in the Influence of Product Quality and Brand Image on Home Purchase Decisions in Bali: a Study of Bali Surya Residence Real Estate. Russian Journal of Agricultural and Socio-Economic Sciences, 139(7), 71-84.

[9] Wijaya, A. G., & Zulfa, M. (2017). The effect of word of mouth, price perception, brand image and trust on purchase decision using purchase intention as an interveningvariable (study at middle class housing in semarang). IJIBE (International Journal of Islamic Business Ethics), 2(3), 1-12.

[10] Ramadhani, D., & Kodrat, D. S. (2024). Location and Brand Image Outshine Price in Homebuyer Decisions in Indonesia. Academia Open, 9(2), 10-21070.

[11] Cahyadi, H. A., Santoso, B., & Setiawan, R. (2025, March). The Effect of Murabahah Contract and Location on Housing Purchase Intention with Brand Image as an Intervening Variable. In Proceeding of International Conference on Social Science and Humanity (Vol. 2, No. 1, pp. 450-469).

[12] Widjaja, H., & Santoso, H. (2024). Digital Marketing: A Case Study of Social Media Marketing of Indonesia Real Estate Companies. Business Economic, Communication, and Social Sciences Journal (BECOSS), 6(2), 143-157.

[13] Sajuri, W., & Mansor, H. N. (2023). The challenges with digital marketing in the practice of estate agents. In e-Proceeding 6th Undergraduate Seminar on Built Environment and Technology (USBET) 2023 (pp. 1669-1682).

[14] Matidza, I., Ping, T., & Nyasulu, C. (2020). Use of digital marketing in estate agency industry in Malawi. E-Learning and Digital Media, 17(3), 253-270.

[15] Tarigan, A., & Noverin, D. T. (2024). The Influence of Social Media Marketing and Brand Image on Intention to Purchase Residential Apartments. Journal Ilmiah Manajemen dan Bisnis ISSN, 10(1), 81-96.

[16] ISWARA, A. N. K. P., INDIANI, N. L. P., & WAHYUNI, N. M. (2024). The Mediating Role of Brand Image on the Influence of Digital Marketing and Word of Mouth on Property Purchase Intention in Bali. International Journal of Environmental, Sustainability, and Social Science, 5(3), 679-695.

[17] Hsiao, S. H., Wang, Y. Y., & Lin, T. L. (2024). The impact of low-immersion virtual reality on product sales: Insights from the real estate industry. Decision Support Systems, 178, 114131.
[18] Prema, K., Chhetri, A., Singh, S. P., & Rao, D. Y. (2024, December). Virtual Reality for Real Estate: Revolutionizing Property Visualization and Enhancing Client Engagement. In 2024 IEEE 4th International Conference on ICT in Business Industry & Government (ICTBIG) (pp. 1-7). IEEE.
[19] Xiong, C., Cheung, K. S., Levy, D. S., & Allen, M. (2024). The effect of virtual reality on the marketing of residential property. Housing Studies, 39(3), 671-694.
[20] Huang, S., & Zheng, H. (2023). Brand experience and brand image building based on virtual reality technology. Computer-Aided Design and Applications, 20, 49-60.