Journal of Engineering Design

ISSN: 0954-4828 (Print) 1466-1837 (Online) Journal homepage: www.tandfonline.com/journals/cjen20

Human-AI co-ideation via combinational generative model

Pan Wang, Xun Zhang, Liyan Wei, Peter Childs, Stephen Jia Wang, Yike Guo & Maaike Kleinsmann

To cite this article: Pan Wang, Xun Zhang, Liyan Wei, Peter Childs, Stephen Jia Wang, Yike Guo & Maaike Kleinsmann (2026) Human-AI co-ideation via combinational generative model, Journal of Engineering Design, 37:2, 458-494, DOI: 10.1080/09544828.2025.2504309

To link to this article: https://doi.org/10.1080/09544828.2025.2504309

Published online: 01 Jun 2025.

JOURNAL OF ENGINEERING DESIGN
2026, VOL. 37, NO. 2, 458–494
https://doi.org/10.1080/09544828.2025.2504309

Human-AI co-ideation via combinational generative model

Pan Wang [a], Xun Zhang [b], Liyan Wei [c], Peter Childs [d], Stephen Jia Wang [c], Yike Guo [e] and Maaike Kleinsmann [a]

[a]Industrial Design Engineering, Delft University of technology, Delft, Netherlands; [b]Laboratory for Artificial Intelligence in Design, The Hong Kong Polytechnic University, Hong Kong, China; [c]School of Design, The Hong Kong Polytechnic University, Hong Kong, China; [d]Dyson School of Design Engineering, Imperial College London, London, United Kingdom; [e]Department of Computer Science, The Hong Kong University of Science and Technology, Hong Kong, China

ABSTRACT

Ideation is a critical step in the engineering design process, enabling designers to develop creative and innovative concepts and prototypes. Currently, the ideation workflow requires designers to generate new designs based on product requirements, heavily relying on their personal expertise and experience. To advance human-AI collaboration design and assist designers in the idea-generation process, this paper proposes an Object Combination Generative Adversarial Network (OC-GAN) for combinational creativity. The proposed method includes an image encoder module and a cross-domain object combination generator module. The image encoder module captures and encodes image structure information into latent space, while the cross-domain object combination generator module leverages GANs to combine object images based on user preferences, producing new design images. A design case study is used to evaluate the new ideation approach and reveal not only strong cross-domain concept combination capabilities but also improvement in designers' workflow and provision of novelty to the design case.

Highlights

- An AI approach to improve the efficiency of idea generation in the design process.
- A case study evaluates its support for idea generation and design creativity.
- The OC-GAN is used for multi-domain object image combining tasks.
- Exemplifies the feasibility of human-AI collaboration design for enhancing creativity.

ARTICLE HISTORY
Received 18 November 2024
Accepted 7 May 2025

KEYWORDS
Ideation; combinational creativity; visual stimuli; artificial intelligence; generative adversarial networks

CONTACT Pan Wang p.wang-2@tudelft.nl Landbergstraat 15, 2628 CE Delft, Netherlands

1. Introduction

Conceptual design serves as a pivotal phase in the design process, facilitating a seamless transition from user requirement analysis to the generation of viable concepts. This stage unfolds as an iterative progression, transforming coarse, vague, and abstract ideas into refined, clear, and concrete solutions (Beitz, Pahl, and Grote 1996; Ulrich and Eppinger 2016). Central to this evolution, ideation governs conceptual design by crafting solutions precisely aligned with user needs (Chou 2014). Conventionally, this process hinges on designers' expertise and experience to generate original concepts, which makes ideation process both labour-intensive and time-consuming (Yilmaz et al. 2015). Consequently, the design industry has prioritised research into methods that streamline and expedite brainstorming without compromising its innovative and creative essence (L. Chen, Cai et al. 2024).

Extensive research has been devoted to advancing ideation methods within the design process (L. Chen, Wang et al. 2019; L. Chen, Xiao et al. 2024; L. Chen, Zhang et al. 2024; Chou 2024; Jiang et al. 2022). Among the diverse strategies explored, combinational creativity distinguishes itself by synthesising concepts from seemingly unrelated or indirectly related domains–encompassing modalities such as images, videos, or text–to yield innovative design solutions (Fu et al. 2014; Linsey, Wood, and Markman 2008; Smith, Troy, and Summers 2006). This approach inspires designers' creative processes and enhances their productivity by extending the scope of conceptual exploration (Fu et al. 2014; Linsey, Wood, and Markman 2008; Smith, Troy, and Summers 2006).

However, traditional approaches to combinational creativity are often limited by their reliance on textual representation or heavily dependent on professional expertise (Frigotto and Riccaboni 2011; Henriksen and Mishra 2014; Yang and Zhang 2016), rendering the ideation process labour-intensive, particularly for designers with limited artistic skills or experience. While these methods aim to integrate concepts from diverse domains, their reliance on manual effort restricts accessibility and efficiency, especially for novices. Recent program-based methods have offered interactive tools that simultaneously present text and images in the same window to support designers during ideation (Han et al. 2016, 2018). However, these tools typically provide only basic image overlay, which constrains their ability to facilitate meaningful synthesis.

The rapid advancement of artificial intelligence (AI), driven by the exponential growth of internet data, has opened new pathways for exploring creativity, equipping designers with data-driven tools to enhance ideation (Chen and Fuge 2017; Chen and Li 2024; Han, Childs, and Luo 2024; Jin et al. 2024; Lee et al. 2024). Among these, generative models (Mou et al. 2025) stand out for their ability to intuitively visualise design concepts through synthesised content. Recent advancements in AI-based tools have utilised generative models to enhance the creative process, enabling designers to transform random noise or textual prompts into visual outputs or to achieve integration of features from two images (L. Chen, Wang et al. 2019; L. Chen, Zhang et al. 2024; Wang, Tan, and Ma 2024). However, these approaches predominantly focus on whole-image generation or combination, lacking the fine-grained control necessary to produce precise and inspiring design images, which constrains their effectiveness in facilitating targeted ideation.

To address these challenges, we propose an Object Combination Generative Adversarial Network (OC-GAN), an AI-driven generative model designed to enhance design ideation. Our approach enables the extraction and integration of semantic visual structures from

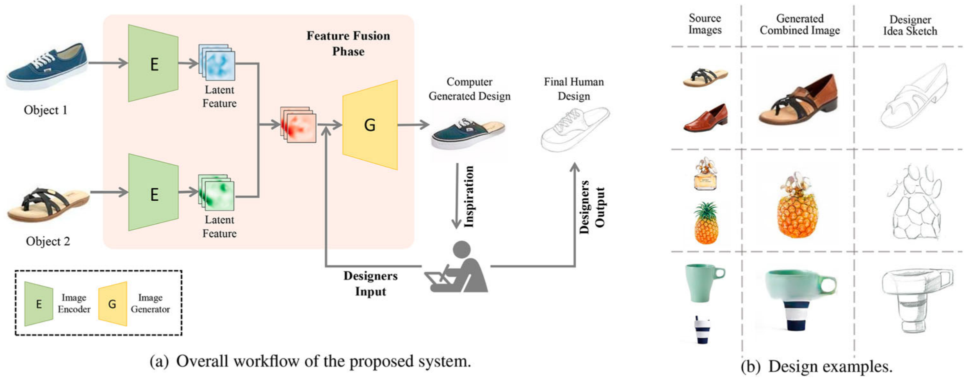


Figure 1. The overall workflow of the proposed system and design examples. The left section illustrates the overall process of the system. Designers select two object images as inputs to the system and provide ideas for feature fusion during the feature fusion phase. The system then performs the fusion of the images based on the designers' requirements and generates the fused image. The generated image serves to inspire the designers and aids in the completion of their sketch designs. The right section displays a subset of design outcomes achieved by designers utilising the proposed system. (a) Overall workflow of the proposed system. (b) Design examples.

selected images based on designers' requirements, facilitating the creation of novel design concepts. The framework incorporates an intuitive human-computer interaction interface, allowing designers to exert high-level control over the generated design images while significantly reducing cognitive load and minimising the expertise required for combinational creativity tasks (Figure 1). To evaluate the effectiveness and practical utility of our method, we conduct a comprehensive design case study. The results validate the efficacy of our approach in supporting ideation, demonstrating a reduction in cognitive workload and an improvement in creative efficiency compared to traditional design ideation methods. The primary contributions of this study are as follows:

(1) A novel AI generative approach is proposed to enhance the efficiency of idea generation within the design process.
(2) We Implemented the OC-GAN model for latent space vector extraction, image generation, and local editing, achieving the task of editing and merging multidomain object images.
(3) The suggested method is evaluated using a design case study that assesses its support for the idea-generation process and the creativity of the generated designs.
(4) Our approach exemplifies human-AI collaboration design in ideation process, showing that such collaboration boosts creativity and is feasible. It serves as an example for future implementations to enhance design processes and reduce designers' workload.

2. Related work

2.1. AI-based data-driven ideation

In the design process, ideation is critical for creating, developing, and communicating ideas (Childs et al. 2022; Sun, Zhang et al. 2025). Advances in computational creativity have led to AI-based strategies that enhance idea generation, broadly categorised into stimulus retrieval and stimulus generation (Tang et al. 2019). Stimulus retrieval often employs data mining and information retrieval techniques to extract insights from large databases. For instance, Chan et al. (2011) highlighted the role of analogical data in innovative design through a patent database case study, while Hao et al. (2019) developed an evolutionary computing system to derive design language from 500,000 patents. Similarly, Shi et al. (2017) proposed an ontology network for unsupervised learning, linking design and engineering information via text mining and semantic analysis. However, these approaches depend heavily on existing data patterns, restricting designers' ability to explore truly novel concepts. Other studies (Lin et al. 2013; Sarica et al. 2021; Toh et al. 2017) have improved creativity assessment and learning but often fail to directly support the generation of innovative design concepts, especially for novel visual styles or combinations. The emergence of image generation models has advanced idea generation, enabling AI to produce imagery that supports ideation. Numerous studies (Archana Balkrishna 2024; L. Chen, Song et al. 2025; X. Chen et al. 2024; Hayakawa, Noji, and Kato 2024; Heyrani Nobari, Rashad, and Ahmed 2021; Honda, Yanagisawa, and Kato 2022; Jin et al. 2024; Karadag, Güzelci, and Alaçam 2023; Khan et al. 2023; W. Li, Zhang et al. 2024; Lu et al. 2024; Ren et al. 2024; Rosati et al. 2024; Zang et al. 2024; Zhang, Li, and Zheng 2024) have explored generative AI's potential across various design domains. For example, X. Li, Su et al. (2021) combined an affective recognition model with a GAN to generate user-aligned conceptual images, though reliance on manual preprocessing and a small dataset limits practical effectiveness. Similarly, Jiang and Fu (2017) and Sbai et al. (2018) focussed on fashion style generation, with the latter introducing a creativity loss to enhance originality, while Heyrani Nobari, Rashad, and Ahmed (2021) and Khan et al. (2023) developed ShipHullGAN and Creative-GAN for ship hull and bicycle designs, respectively. Recent advances in Stable Diffusion (SD) have shown promise in industrial design, with Liu and Hu (2023) achieving high-quality sketching and rendering outcomes, and Cao et al. (2023), Sun, Zhou et al. (2023), and Zhang et al. (2021) advancing conditional and reference-based synthesis for fashion design, yielding state-of-the-art (SOTA) results. L. Chen, Song et al. (2025) investigated the roles of Generative AI in conceptual design and found that it mainly supports problem definition and idea generation, while idea selection and evaluation remain largely human-driven. However, these methods often prioritise broad style control or complete image generation over detailed editing, reducing designers' control over the outputs. In contrast, our proposed GAN-based approach facilitates high-quality, user-editable design concept generation, enhancing designer engagement in the ideation process using generative models.

2.2. Combinational creativity for the ideation process

According to Boden (2004), creativity is the ability to generate novel, surprising, and valuable ideas or products, encompassing three primary ideation processes: exploratory, transformational, and combinational creativity. Exploratory creativity involves serendipitous discoveries within conceptual spaces, while transformational creativity entails reconfiguring these spaces across domains. Combinational creativity focuses on integrating two concepts–spanning images, videos, or texts–to produce innovative outcomes. Frigotto and Riccaboni (2011), Henriksen and Mishra (2014), and Yang and Zhang (2016) highlight the significance of synthesising pre-existing knowledge to foster creativity, framing it as a process of merging concepts into novel forms. Recent studies on combinational creativity have explored various techniques: Toivonen and Gross (2015) investigated data mining's potential, and Bacciotti, Borgianni, and Rotini (2016) examined cross-dimensional concept integration for product ideation. However, these approaches predominantly rely on textual data, lacking the intuitive stimulation of visual inputs. For image-based combinatorial creativity, Han et al. (2016, 2018) developed a computer-based tool that displays text and images within a single window to support designers in generating ideas. However, its functionality is limited to overlaying images without synthesis, constraining its ability to foster creative exploration. With the advent of generative models, researchers have begun leveraging their potential for combinational creativity. L. Chen, Wang et al. (2019) proposed a visual concept combination model that generates images by merging partial structures from distinct domains, while Wang, Tan, and Ma (2024) introduced a generative AI tool enabling innovative image experimentation. Both methods perform uncontrolled fusion of images as a whole, providing users with no precise control over the fusion regions or techniques, thereby limiting the user experience. Other studies (L. Chen, Xiao et al. 2024; L. Chen, Zhang et al. 2024; Wang et al. 2023) further explore the interpretation and evaluation of combinational designs. Our proposed method harnesses the local editing capabilities of generative models, allowing users to precisely control the regions of combination based on their needs. This generates tailored combination results, enhancing creative exploration and improving the user experience.

2.3. Generation models and local editing

The visual stimuli-based ideation process relies on photo-realistic images as visual cues to enhance inspiration. Recent advancements in GANs (Brock, Donahue, and Simonyan 2018; Isola et al. 2017; Radford, Metz, and Chintala 2015; J. Y. Zhu et al. 2017) have yielded promising results in generating such images. Several studies (Abdal, Qin, and Wonka 2019, 2020; Donahue, Krähenbühl, and Darrell 2017; Karras, Laine, and Aila 2019; Karras et al. 2020; Park et al. 2020; Pidhorskyi, Adjeroh, and Doretto 2020; J. Zhu et al. 2020) have explored inter-domain, inter-attribute, inter-concept, and inter-style image generation to produce clearer, more realistic outputs; however, these GAN-based methods focus solely on generation and lack local editing capabilities. Local editing, an extension of inter-domain generation, targets specific image regions (e.g. nose, eyes, mouth, background) rather than altering global appearance. Collins et al. (2020) introduced an effective approach for semantically aware local edits by manipulating style vectors from generated images, while Alharbi and Wonka (2020) replaced the learned constant from StyleGAN with multiple noise codes to achieve local and global editing. Nevertheless, these methods are limited to editing GAN-generated images and cannot modify real images. To address real-image editing, P. Zhu et al. (2020) manipulated per-region style codes, though their model requires training with paired images and segmentation masks. Similarly, Suzuki et al. (2018) proposed a CNN-based technique that adjusts the feature-space representation within a trained GAN to edit semantic content at user-specified locations, but inaccuracies in source region selection can lead to unintended feature blending. Recent large vision-language models (LVLMs) (Liu et al. 2023; C. Wu et al. 2024; Yan et al. 2025) offer a new paradigm for image fusion and generation by leveraging large-scale vision-language modelling to achieve semantically controllable composition. Although these models demonstrate impressive capabilities in blending images based on high-level textual descriptions, they still primarily focus on semantic-level fusion and lack pixel-level precise alignment and blending. To overcome these limitations, our proposed method integrates a spatially-aware style mapping strategy within a GAN framework, enabling high-quality image generation alongside precise, interactive local editing and blending of real images.

3. Proposed ideation approach

Prior research has demonstrated that high-quality visual stimuli serve as a critical factor in the ideation process (Laing 2018; Venkataraman et al. 2017; J. Wu et al. 2024; Zhao et al. 2021), and studies on combinational creativity have emphasised the importance of image combination capabilities of the supporting tools to manipulate visual elements to better align with their creative intent (Chilton, Petridis, and Agrawala 2019; Han et al. 2016). based on these needs, we adopted the StyleMapGAN (Kim et al. 2021) model architecture because of its excellent capabilities in image generation and local editing. The model can be trained according to the specific requirements of different design tasks with different design datasets for design image generation and local editing.

The detailed architecture of the proposed GAN model is presented in Figure 2. To elucidate its functionality, we describe the roles of the three pipelines and their relations depicted in the Figure 2. The framework integrates random image generation, real image encoding, and fake image encoding to enable high-quality, editable design image synthesis. The first pipeline (Figure 2(a)) trains the image generator to achieve high-quality synthesis from randomly sampled gaussian noise, allowing the model to learn the training data distribution. The second pipeline (Figure 2(b)) employs an image encoder to map real images into the latent space, enabling accurate reconstruction of the original image and supporting the local edition ability with existing real images. The third pipeline (Figure 2(c)) further enhances the ability of image generator and image encoder by training them with generated images, ensuring robust latent representations for both real and fake images, which improves consistency and performance in tasks like local image editing. Together, these pipelines balance generation, reconstruction, and editing capabilities, making the proposed GAN model highly effective.

In the following sections, we will introduce the details of the model's generator, training procedure, and local editing process.

3.1. Object generator

We constructed the object combination GAN (OC-GAN) based on StylemapGAN (Kim et al. 2021) and introduced a multiscale condition encoder (MCE) to enable condition control for image generation. The detailed structure of the proposed conditional object generator is shown in Figure 3. StyleGAN2 uses the style mapping network to produce style vectors from random noise z to control the feature maps. As an expansion of the original StyleGAN2, StylemapGAN changes the output style vectors into style features with spatial dimensions to increase the effectiveness of the real-image projection and enable local editing. Our approach follows StylemapGAN's style mapping network architecture, which

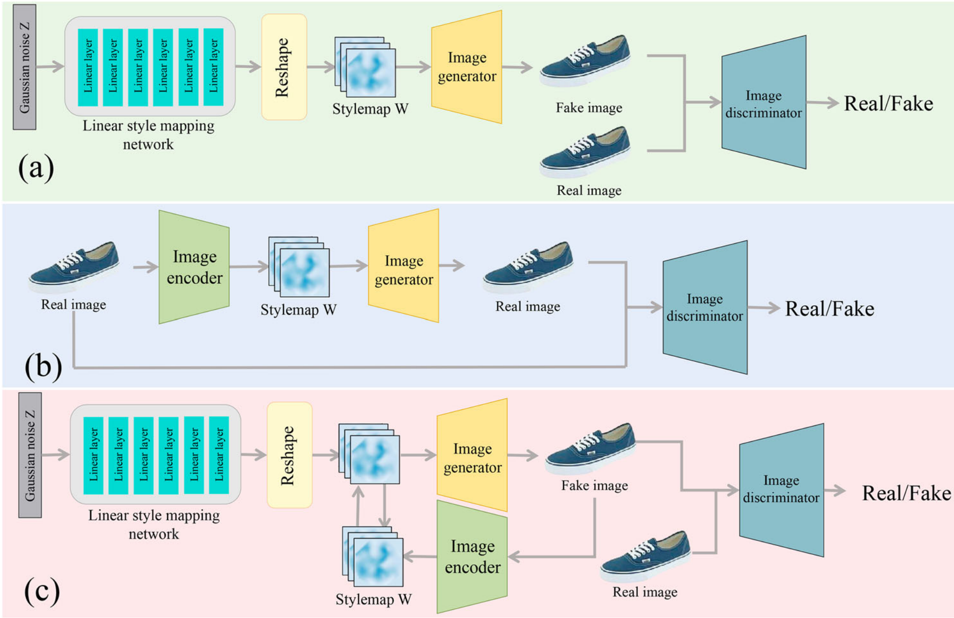


Figure 2. The overall architecture of the proposed GAN model, comprising three distinct pipelines tailored to specific training objectives. (a) Depicts the random image generation pipeline, which enables the model to learn the training data distribution for random generation. (b) Shows the image encoder training pipeline for real image encoding, training the encoder to extract latent features from real images suitable for the generator. (c) Presents the image encoder training pipeline for fake image encoding, enhancing the encoder's feature extraction capability using randomly generated images.

consists of eight fully connected layers and a reshape layer. The channel size of the fully connected layers is 64, except for the final layer, which is 4096. The reshape layer takes the final output vector of fully connected layers and transforms it into the stylemap.

The stylemap resizer and the image synthesis network gradually increase the size of the stylemap and the feature map to the output size to acquire generated images. The stylemap resizer consists of several stages of convolution and upsampling blocks to increase the resolution of the stylemap. Each block consists of two convolutional layers and an upsampling layer, which produce two $W+$ space stylemaps to match the input of the corresponding synthesis block. With learned convolutions, the stylemap resizer resizes and transforms the stylemap to express more organised and detailed styles. The affine transform generates parameters for the modulation according to stylemaps. The formulas of the modulation operation are as follows:

$$p_i = \frac{h_i - \mu_i}{\sigma_i} \tag{1}$$

$$h_{i+1} = (\gamma_i \otimes p_i) \oplus \beta_i, \tag{2}$$

where i represents the i-th layer, μ_i and σ_i represent the mean and standard deviation of the activation parameters, and h_i stands for the activation of the i-th layers. γ_i and β_i are the parameters of the affine transform modulation. $\otimes$ is the element-wise multiplication, and $\oplus$ is the element-wise addition.

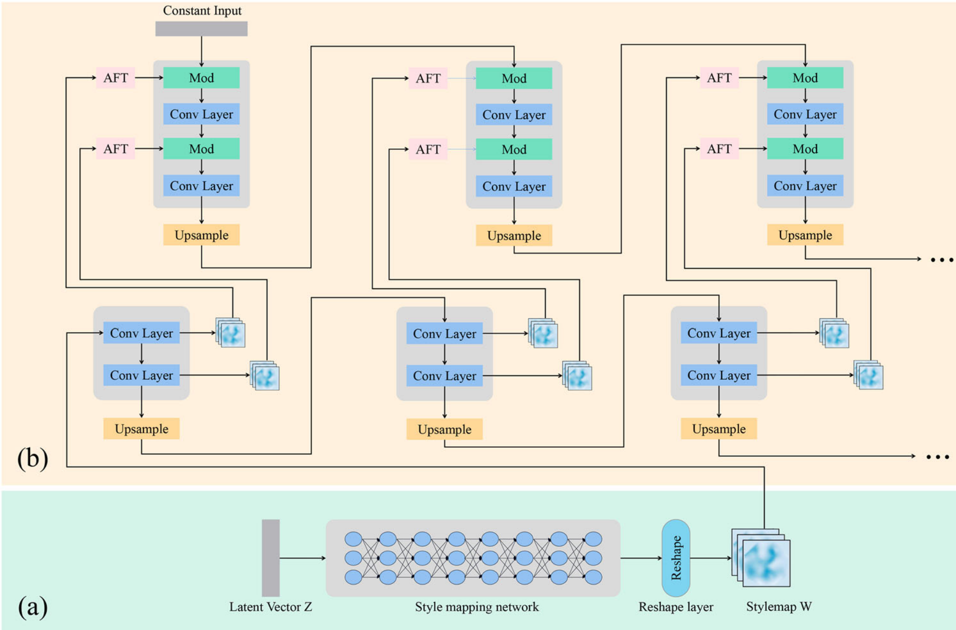


Figure 3. The structure of the image generator. The style mapping network is an eight layer fully connected neural network. 'Conv layer' stands for a conventional layer with a 3 × 3 kernel. 'Constant input' is a random generated tensor in the shape of 512 × 8 × 8. 'AFT' stands for a learned affine transform. 'Mod' is a modulation consisting of element-wise multiplication and addition.

Six losses are included in the training process to ensure the best performance of the proposed model. The introduction of each loss function is as follows.

Adversarial loss. Adversarial loss is used in the training of the discriminator and the generator. Fake images are denoted as generated images from the generator, and the real images are denoted as images from the training dataset. The discriminator tries to classify generated images as fake images and images from the training dataset as real images. The generator tries to fool the discriminator by generating more realistic images. The discriminator and the generator are trained adversarially to improve their performance. The choice for the adversarial loss is a non-saturating loss. The formulas are as follows:

$$l_{Ad}^{g} = -\log p\left(y_d = 1|I^{\text{generate}}\right) \tag{3}$$

$$l_{Ad}^{d} = -\log p\left(y_d = 1|I^{\text{real}}\right) - \log p\left(y_d = 0|I^{\text{generate}}\right) \tag{4}$$

Domain-guided loss. J. Zhu et al. (2020) proposed domain-guided loss for image inversion and real-image editing. The model is trained in an adversarial training manner. However, instead of using random Gaussian noise, the generator's input is changed to a latent space vector produced by an image encoder. The image encoder encodes images into latent space, and the generator tries to reconstruct the original image according to the latent space vector. The encoder and the generator try to reconstruct an image that is as like the original image as possible, while the discriminator tries to distinguish the original image from the generated image. As a result of the loss, the projected latent code is compelled to stay in the GAN's original latent space, providing better image editing and interpolation results. The formulas are as follows:

$$\begin{aligned} l_e = & \| I^{\text{real}} - G\left(E\left(I^{\text{real}}\right)\right) \|_2 \\ & + \lambda_{vgg} \| F_{vgg}\left(I^{\text{real}}\right) - F_{vgg}\left(G\left(E\left(I^{\text{real}}\right)\right)\right) \|_2 \\ & - \lambda_{adv}\left(-\log p\left(y_d = 1 | I^{\text{real}}\right) - \log p\left(y_d = 0 \mid G\left(E\left(I^{\text{real}}\right)\right)\right)\right) \end{aligned} \tag{5}$$

$$\begin{aligned} l_d = & -\log p\left(y_d = 1 \mid I^{\text{real}}\right) \\ & - \log p\left(y_d = 0 \mid G\left(E\left(I^{\text{real}}\right)\right)\right), \end{aligned} \tag{6}$$

where F_{vgg} is the VGG feature extraction model and E and G represent Encode and Generator, respectively.

Latent reconstruction loss. The image encoder was trained to find the best latent code for a given image for reconstruction. To better train the image encoder, we created training pairs using random Gaussian noise and the corresponding generated images using the generator. The generated image was the input and the random Gaussian noise is the target latent code. The encoder was trained under the above supervision, which resulted in better latent code generation ability and reduced the negative bias toward pixel-level reconstruction. The formula is as follows:

$$l_{\text{latent}} = \text{MSE}\left(z, E\left(G\left(z\right)\right)\right) \tag{7}$$

Image reconstruction loss. The goal of image reconstruction loss is to eliminate pixel-level variations between generated and real images such that they are visually similar. The formula is as follows:

$$l_{\text{pixel}} = \| I^{\text{real}} - G\left(E\left(I^{\text{real}}\right)\right) \|_2 \tag{8}$$

Perceptual loss. Pixel-level loss usually causes overfitting of the encoder. Perceptual loss uses feature-level differences instead of pixel-level differences to asses the accuracy of the reconstruction, which results in stable training of the encoder and better image reconstruction accuracy at both pixel and feature levels. Several popular studies use features extracted from a pretrained VGG model as the image representation. We used learned perceptual image patch similarity (LPIPS) for perceptual loss due to better image feature representation. The formula is as follows:

$$l_{perc} = \text{MSE}\left(F_{vgg}\left(I^{\text{real}}\right), F_{vgg}G\left(E\left(I^{\text{real}}\right)\right)\right) \tag{9}$$

R1 regularisation. The training of the GAN model is unstable. By penalising large variations in the output of certain neural network layers, R1 regularisation can stabilise the training process. The R1 regularisation is applied after every 16th step of the discriminator. The formula is as follows:

$$R_1\left(\psi\right) = \frac{\gamma}{2} E_{p_D(x)}\left[\| \nabla D_\psi\left(x\right) \|^2\right], \tag{10}$$

where ψ is the discriminator weights; $E_{p_D(x)}$ represents the regularisation-only focus on real images; γ is the hyperparameter.

3.2. Training procedure

The proposed model contained four modules: the style mapping network, the object generator, the discriminator, and the image encoder. The image encoder and the generator were jointly trained, as opposed to being trained individually, to achieve stable training and higher performance. Adversarial loss was used to train the generator and the discriminator in an adversarial fashion by having them compete with each other. Both pixel-level and perceptual-level reconstruction accuracy were measured for the original image and generated image to train the image encoder and the generator. In addition, the image encoder was also trained with mean square loss (MSE) to reconstruct the stylemap from generated images using random noise z. Finally, the domain-guided loss was applied to the image encoder, the generator and the discriminator for better stylemap generation to improve image editing. The steps mentioned above were necessary to obtain the best model performance.

3.3. Visual object combination and local editing

Visual object combination is intended to combine images from two concepts into one image, which should contain feature concepts from both images. However, the traditional combination process only randomly combines the feature concepts, which results in poor combination results. Local editing addresses the shortcomings of random combination methods by introducing a user-defined mask into the combination process. Local editing uses a mask to pinpoint the area to be altered, which is intended to transplant certain portions of a reference image to an original image. The user can choose any position in the reference image and use masks in any shape, which provides the best human-computer interactive experience.

Two images were chosen to serve as the reference image and the original image for the local editing process. The objective was to replace a specific area of the original image with the reference image using user-defined masks. The reference image and the original image are denoted as I_r and I_o, respectively. The image encoder E encodes I_r and I_o into stylemaps w_r and w_o. Then, w_r and w_o pass through the stylemap resizer so that the stylemaps are generated in the $w+$ space. The stylemaps in the $w+$ space are merged according to the mask m in different resolutions, and mask m uses max pooling to downsample itself to map the resolution of the corresponding stylemap. The local editing workflow of the proposed approach is shown in Figure 4. The edited stylemap $w*$ is calculated using the following formula:

$$w* = m \otimes w_r \oplus (1 - m) \otimes w_o \tag{11}$$

4. Implementation and experiments

4.1. Datasets

To collect suitable datasets for training and testing our proposed model, we proposed a data collection technique to acquire and filter suitable data from internet databases. Initially, we employed web crawlers to download images from commonly used search

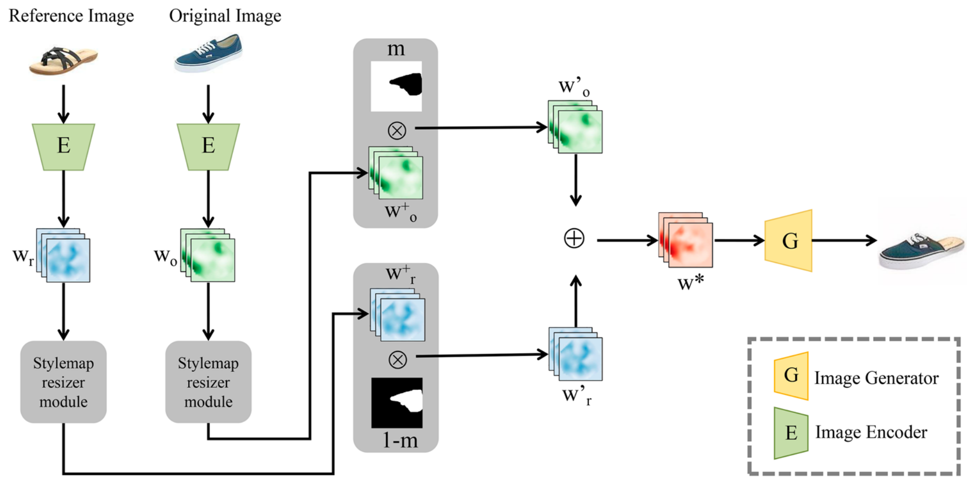


Figure 4. The local editing workflow of the proposed approach. A reference image and an original image are encoded using the same image encoder, generating the corresponding latent space stylemaps W_r and W_o. Next, W_r and W_o are respectively multiplied by the corresponding mask information m and $1-m$, producing the masked stylemaps. Subsequently, W_r' and W_o' are added together to form a new stylemap W^*, which contains information from both the reference image and the original image, and it is ultimately used by the generative model to produce the image combination outcomes.

engines such as Google, Baidu, and Amazon. We used the categories we needed to collect as keywords for the searches. The categorisation of the objects includes shoes (boots, flats, heels, sandals, and sneakers), backpacks, bags, bulbs, cups, lamps, leaves, perfume bottles, pineapples, tables, and wine glasses. These categories were selected based on our design task requirements and the ease of data collection, aiming to maximise the effectiveness of our proposed model. At this initial stage, we collect a total of 55,855 images, and the number of images per category in the dataset is shown in Table 1. Due to the high randomness and uncertainty of the images collected from the internet, we conducted a detailed data filtering to exclude those with irrelevant content, overly complex backgrounds, excessively tiny target objects, multiple target objects in a single image, and extreme image pixel ratios. Given the large volume of data collected from the internet, we first used the YOLO v8 object detection model to detect objects in the images. During this process, we obtained the categories of objects in the images as well as their bounding boxes. We retained only those images that contained our specified categories and where the bounding box of the category occupied more than one-third of the image area. After the initial screening, we manually reviewed the filtered image dataset again. During this process, we removed images with overly complex backgrounds, multiple target objects in a single image, extremely positioned objects, and extreme image pixel ratios. Through these two rounds of data filtering, we obtained images with high category relevance, simple backgrounds, centrally positioned objects, and appropriate pixel ratios as our model's training dataset. After data filtering, we obtained 48,336 images that fully meet our requirements, with the number of images per category presented in Table 1. For the shoe category, the number of images remained the same before and after filtering because multiple specialised shoe datasets were used during collection, resulting in high initial data quality. For

Figure 5. Dataset examples. Each row represents several image examples from one category in the object dataset.

other categories, the images were mostly collected and downloaded using web crawlers from search engines like Google and Amazon, leading to less ideal initial data quality and a significant reduction in the number of images per category after filtering. Other processes included image type adjustment and image resizing. The ideal processed dataset image was a rectangle containing just one category of objects, with the object occupying the majority of the image area and the background being white. Examples from our dataset are shown in Figure 5.

4.2. Implementation detail

The proposed model was implemented using the PyTorch framework. The size of the input images was $256 \times 256 \times 3$, and the size of the output images was also $256 \times 256 \times 3$. The style mapping network used random noise with 64 dimensions as the input and generated a 4096 dimension style vector. The style vector then passed through a reshape layer to generate an $8 \times 64 \times 64$ stylemap. The image synthesis module took the stylemap as input and

Table 1. Number of images for each category.

Categories	Numbers before filtering	Numbers after filtering
shoes	34172	34172
backpacks	2541	1806
bags	1937	1491
bulbs	5036	3862
cups	1062	809
lamps	926	662
leaves	2106	680
perfume bottles	1657	1283
pineapples	4117	1992
tables	862	698
wine glasses	1439	881

Note: The second column shows the number of images after the initial data collection, and the third column shows the number of images after filtering.

generated the final output image. The four proposed modules were trained jointly so that the training would be stable. Six losses were included to guide the model training process for the best result. The model optimiser used was the Adam optimiser, and the learning rate was 0.0002. We trained the generation models based on the experiment groups to evaluate the model's performance on different object generation tasks. Each model could generate at least one category of the object. The model was trained on four NVIDIA A100 GPUs.

4.3. Evaluation matrix

The generation model was evaluated using an inception score (IS) and Fréchet inception distance (FID). FID calculates the feature vector distance between trained dataset images and images generated using random Gaussian noise. The image feature extraction model was an ImageNet-pretrained Inception-V3. We counted the number of images in the training dataset and used the generative model to generate the same number of images for a fair comparison. The realism and diversity of the generation results were quantified using an IS. The realism reflected the performance of the generation model, and the diversity proved the model was not in a state of mode collapse. The image projection and reconstruction ability were evaluated using pixel-level and feature-level accuracy. The pixel-level accuracy was defined as the mean square error between the target images and the reconstructed images. The feature-level accuracy was calculated using the learned perceptual image patch similarity between the target images and the reconstructed images.

4.4. Object generation

By analysing the object dataset collected from the internet, we separated the objects into single-category, single-concept objects and single-category, multiconcept objects. More specifically, if there were few differences in colour, structure and shape between the images in a category, the category was considered as a single-category, single-concept object. Conversely, if there were large differences in colour, structure or shape between the images in a category, the category was considered as a single-category multiconcept object. For example, the pineapple was considered a single-category, single-concept object due to there being few differences in colour, structure and shape between images. On the other

Figure 6. Object generation results. The figure illustrates the image generation capabilities for various categories using the proposed model. All images were generated by the trained model using random noise as input.

Table 2. Quantitative evaluation of image generation on shoe dataset.

Method	IS ↑	FID ↓
Vanilla GAN (Goodfellow et al. 2014)	2.6341	39.3192
BigGAN (Brock, Donahue, and Simonyan 2018)	3.1926	31.8160
StyleGAN2 (Karras et al. 2020)	3.6125	25.7204
Image2StyleGAN (Abdal, Qin, and Wonka 2019)	3.7612	25.1663
OC-GAN (Ours)	**4.2928**	**21.7301**

hand, the shoe was considered a single-category, multiconcept object because it can be further divided into subcategories (e.g. boots, flats, heels, sandals, and sneakers), which have large differences in colour, structure and shape. We trained and evaluated the proposed approach according to different sets of object categories. The object generation results are shown in Figure 6.

The image generation performance of the proposed approach was evaluated by comparing the output with several GAN models trained on the shoe category. The quantitative evaluations are shown in Table 2. To demonstrate that the images generated by the model using random noise are not derived from the training dataset, we adopted a feature comparison method to verify the similarity between the output images and the training images. First, we randomly generated four images using the trained model. Then, we used ResNet50 for feature extraction and calculated the mean squared error(MSE) loss between the features of the generated images and each image in the training set to represent their similarity. The results are shown in Figure 7, where 'top 1' represents the image in the dataset with the smallest MSE loss value compared to the generated image. The generated images exhibit differences in structure, color, and details compared to the training images, proving that the model-generated images are not from the training dataset.

Figure 7. Randomly generated images from the model and the top three most similar images from the training dataset. **Top 1** refers to the image in the training dataset with the smallest computed feature distance to the generated image. It can be observed that the generated images differ from the most similar images in the training set in terms of structure, color, and details, demonstrating that the generated images are not derived from the training data.

4.5. Local editing

Since the focus of our research is on combinational creativity in design, an important task of our proposed model is to accurately combine selected regions of two different objects based on user preferences while preserving the original characteristics of each object to create a novelcombination result. To achieve this, we propose using an image local editing approach to implement this functionality. To evaluate the local editing approach proposed in this paper against the latest diffusion-based local editing methods, we compared our model with the paint-by-example (Yang et al. 2023) for the image local editing task. Paint-by-example (Yang et al. 2023) is the most recent diffusion-based image exemplar-guided editing model, which achieves impressive performance and enables controllable editing on in-the-wild images with high fidelity. We adopted the code and checkpoint from the official GitHub repository. In the comparative experiments, all images were resized to 256 × 256, and the mask areas provided to both models were identical to ensure fairness in the experiments. Figure 8 presents the results of the local editing from GAN-based OC-GAN model and the diffusion-based Paint-by-Example model. According to the results, our proposed model exhibits superior quality and pixel-level accuracy in mask-based image fusion compared to the diffusion-based model.

To analyze the local editing results of the GAN-based (Ours) and diffusion-based (Paint-by-example) models in detail, we refer to the first row of examples in Figure 8. In these

Figure 8. Comparison of the local editing effects between the proposed GAN-based OC-GAN model and the diffusion-based Paint-by-Example model. The OC-GAN model demonstrates superior pixel-level image fusion capabilities, making it highly suitable for combinational design tasks. In the left example from the first row, the goal was to merge the front of a flat shoe with the back of a sandal to create a hybrid design. Our model achieves precise pixel-level fusion with a smooth, natural transition, closely matching the intended outcome. In contrast, Paint-by-Example model fails to achieve pixel-level fusion, merely adjusting the color of the flat shoe's back half rather than truly combining features. We provide a more detailed analysis of the examples in the figures in Section 4.5.

examples, both models attempt to combine specific features from two provided shoe images to generate a new shoe design. For instance, in the left example, the goal was to merge the front half of a flat shoe with the back half of a sandal to create a hybrid design. The results show that the GAN-based model achieves highly precise pixel-level image combination, with a smooth and natural transition at the junction of the two shoe types. The final design closely matches the intended outcome. In contrast, the diffusion-based model, while capable of generating a new shoe design, fails to achieve pixel-level combination of the source images. The result does not align with the desired outcome, as it merely adjusts the color of the flat shoe's back half rather than achieving a true combination of features.The right example exhibits a similar pattern. The GAN-based model successfully fuses the front half of a sandal with the back half of a dress shoe, producing a design with a seamless transition. However, the diffusion-based model generates a design that lacks the characteristics of a sandal, and the front half of the resulting shoe appears unnatural. The second row of Figure 8 illustrates examples of designing a lamp and a cup. Here, the GAN-based model again outperforms the diffusion-based model in pixel-level combination. The transitions at the feature fusion points are more natural in the GAN-based results, with no noticeable structural or size inconsistencies. The third row of Figure 8 involves designing a new perfume bottle by combining the structural features of a pineapple with those of a perfume bottle. The GAN-based model accurately transfers the features of the source perfume bottle onto the corresponding position of the pineapple image, achieving an excellent fusion effect. In contrast, the diffusion-based model struggles to blend the perfume bottle's features with the pineapple image, resulting in a less coherent and lower-quality image in terms of realism and clarity.

Table 3. Mask area MSE loss for the image combination results produced by the Diffusion-based model and the GAN-based model.

Pairs	Diffusion-based model	GAN-based model (Ours)
pairs 1	0.362	**0.083**
pairs 2	0.537	**0.106**
pairs 3	0.511	**0.092**

Through this analysis of the local editing results shown in Figure 8, it is clear that the GAN-based model significantly outperforms the diffusion-based model in terms of precision, coherence, and naturalness of the fusion. These findings strongly demonstrate the advantages of our proposed model for the task at hand. We also calculated the mask area MSE loss to quantitatively analyze the pixel-level image fusion results. In this process, we first provide both models with identical image pairs and mask regions. Subsequently, we generate 50 sets of fusion results based on the given masks. We then compute MSE loss between the mask area of the generated images and those in the mask area of the original reference images. The results are shown in Table 3.

The potential reason for this is that although diffusion models may surpass GAN models in image generation quality, the diffusion-based model requires repeated passes through a UNet model to predict noise during image reconstruction. This process can cause deviations of the feature points due to the extended prediction chain, preventing the model from achieving pixel-level fusion of the reference area into the source image. Furthermore, diffusion models are generally trained on large datasets of natural images, leading to a more abstract understanding of features. Consequently, during image fusion, the model focuses more on the abstract categories of objects in the reference image rather than pixel-level features, resulting in poor pixel-level fusion performance.

In contrast, while GAN models may suffer from mode collapse and other issues, our proposed GAN-based model can more easily align pixel-level features with latent-level features. This allows the model to find the corresponding relationship between each latent-level feature and the original image pixel-level values more effectively during feature fusion in the latent space, resulting in a stronger pixel-level image fusion capability. This capability is more aligned with the requirements of our proposed combinational design task, leading us to ultimately choose a GAN-based model for our approach.

4.6. User interface

We designed an intuitive and structured user-friendly interface to facilitate local image editing tasks. The interface is web-based, enabling quick access through a web browser. Its layout comprises three main functional areas–image selection, image processing, and generation–ensuring clear task segmentation and an efficient workflow.

- Image selection area: users can follow button prompts to select and confirm both the original image and reference image, ensuring that images are correctly loaded into the image processing area without ambiguity.

- Image processing area: users can freely paint masks over target regions to define transfer and blending areas. This flexible selection method allows for precise adjustments without constraints on shape or size.
- Generation area: users can regenerate images multiple times to explore different synthesis effects. A side-by-side comparison view enables direct visual comparison between the generated image, original image, and reference image, ensuring alignment with user expectations.

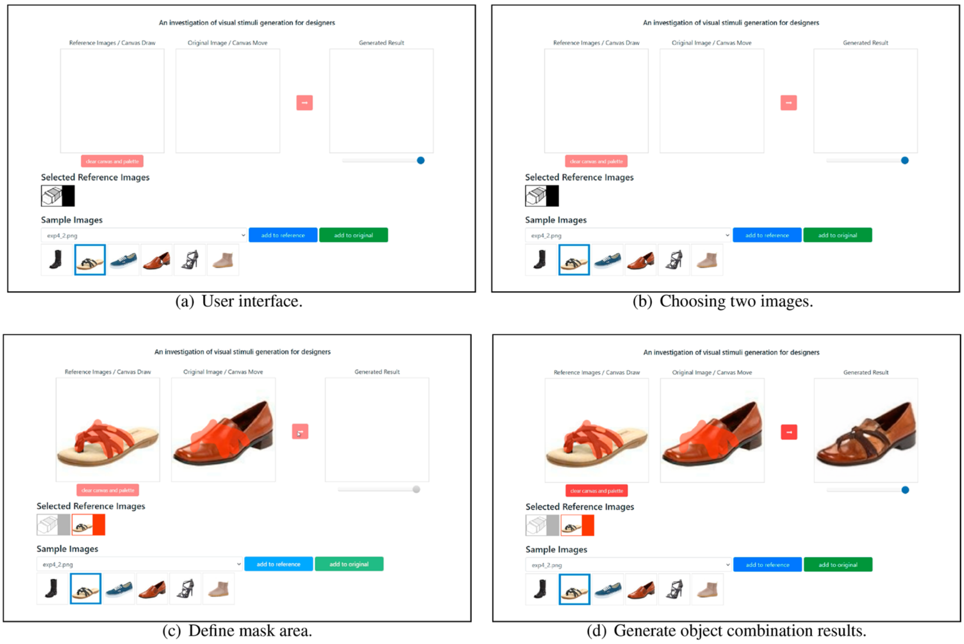
These structured functional divisions and intuitive interaction mechanisms enhance usability, allowing users to navigate and complete tasks efficiently. To further assess the user-friendliness of our interface, we conducted a user study using a Likert scale questionnaire, and the results of these evaluations are reported in Section 5.4. Its layout is shown in Figure 9(a). The interface follows the concept of minimalist design. The reference image and original image are two input images that users can select from the sample image section below. Users can add images to the respective areas by clicking the 'add to reference' and 'add to original' buttons, which are shown in Figure 9(b). Users can draw a mask area within the reference image region using the mouse. Simultaneously, an identical mask area will appear in the original image region. This mask indicates which part of the reference image will be blended into the original image. The mask in the reference image region cannot be repositioned after drawing, but users can modify the mask area in the original image region by manually dragging it to determine which part of the original image will be replaced by the information from the reference image mask area. An example of the mask area in the user interface is shown in Figure 9(c). Once users have finalised the size and position of the mask area, they can click the right arrow button on the page. At this point, the model will run in the background to blend the two input images based on the user-provided mask. After obtaining the result, users can slide the slider below the generated result section to observe the transformation process between the two images, thereby gaining a more comprehensive understanding of the blending process, as shown in Figure 9(d).

5. Case study

5.1. Design tasks

The proposed model was evaluated using five design tasks intended to verify the model's influence on designers' ideation processes, especially combinational creativity for idea generation. Professional designers and ordinary people were considered potential participants in this design case study. The degree of background knowledge required to comprehend the design problem was to be low, allowing participants to concentrate only on the ideation process. The image generation model required separate training for the different design tasks to achieve the best performance. To ensure the fairness of the experiment and simplify the experimental process, the object categories were predefined for each design task instead of chosen randomly by the subjects.

According to the rules described in the previous paragraph, we proposed five design tasks in our case study. The overall design task was to combine the structure and features of the provided object images and design a new object. Due to the limitations of the training



(a) User interface.

(b) Choosing two images.

(c) Define mask area.

(d) Generate object combination results.

Figure 9. The user interface of the proposed approach and an example of performing an object combination task using the user interface. (a) represents the initial interface, (b) shows the selection of the reference image and the target image, (c) displays the mask area drawn by the user, and (d) presents the final generated result.

dataset, each design task used pre-selected images. To better leverage the model's capabilities and validate its generalisation on the test set, all images were sourced from the model's test dataset. Furthermore, when selecting images for the case study, we adhered to the following criteria: consistency in the number of images for the two categories, a certain similarity in the overall structure of the objects, central positioning of the objects in the images, absence of complex background, consistency in the angles of the objects, and an appropriate proportion of the objects within the images. We ultimately decided to select six images for each case study, with three images per category. This number was chosen based on prior research on cognitive load in ideation tasks (Briggs and Reinig 2007; Reinig and Briggs 2013), ensuring that participants were provided with a diverse yet manageable set of stimuli. Additionally, limiting the number of images helped maintain experimental consistency across sessions and prevented excessive decision-making time, which could otherwise hinder the creative process. The titles of the design tasks and the images provided are shown in Figure 10.

The experiment was separated into the introductory session, the system trial run session and five experiment sessions. The duration of the experiment was about 110 minutes. The introductory session lasted about five minutes. During this time, the goal of the experiment and the specifics of each session were explained to the participants to enhance their understanding. For the experimental group, a system trial run session that lasted about five minutes was included to help the participants understand how to use the design assistance system. The images in the system trial run session were not included in the experiment

Figure 10. Design tasks. This paper includes five distinct design cases, each represented by a row. The **design task** column outlines the requirements given to participants for each design case, while the **provided images** column indicates the images supplied to participants. Each design case includes six images as visual stimuli for the participants.

sessions. Each of the five small experiment sessions lasted fifteen minutes, followed by a five-minute break. The design task was to combine the features and structures in the images and design a new object, and the participants were provided with six images for each session. Fifteen minutes were provided for recording ideas, including sketches and descriptions, on white paper. The generated ideas were to be inspired by the provided images, including the shape, colour and structure. The generated ideas were also to meet the usage requirement of the desired category.

In the case study, all design tasks were completed by two distinct groups of participants. Thirty-eight students with different educational backgrounds and work experiences were invited to join this experiment. These students were further randomly separated into two groups, each with 19 participants. The participants in the experimental group used the object combination model for their ideation process. For ease of use, a user-friendly interface was provided to the experimental group, allowing participants to choose images and define the combining areas. Another control group did not have access to the above-mentioned computational tools, but the participants used PowerPoint slides to access the provided images. All experiments for both the experimental and control groups were conducted simultaneously in the same quiet laboratory to ensure participants were not disturbed by external factors. The laboratory was equipped with necessary equipment, including computers, display devices, and drawing tools such as paper and pens. In each experiment session, we provided both the experimental and control groups with the same six images, which were predefined by us before the experiment began, rather than being randomly selected. The selection criteria for these images were based on the requirements of the design case in the experiment sessions and were chosen from the corresponding pre-trained model test datasets. Each image was sized at 256 × 256 pixels. The number, type, and selection criteria of the images provided to participants were consistent to ensure control over experimental conditions, thereby avoiding the influence of variable inconsistencies. The only difference in the experimental setting between the experiment group and the control group is that the experimental group was allowed to use the object combination model we provided to support their ideation process, while the control group could not use the object combination model and could only rely on their imagination to support their ideation process. The other experimental settings for the experiment group and the control group were identical, ensuring the fairness of the entire design case study.

Model preparation work for the experimental group was conducted before the experiments started. Since each experimental task involves the combination of two different categories of images, we specifically collected a corresponding 2-category training dataset for each experimental session. The dataset contains a large number of images from the two selected categories, all of which were carefully selected. The number of training data used for each experimental session is shown in Table 4. These images were then preprocessed to adjust their size and proportion, resulting in a final size of 256 × 256 for each training image. Five separate visual object combination models were trained for different experimental sessions so that the best performances could be obtained. To guarantee fairness, each model was trained using the same configuration and GPU device, and the model with the lowest overall loss was selected for future usage. Multiple image generation datasets were also used to train and evaluate the proposed model, with similarly compelling results. We utilised our own dataset to train the model to adapt it to the experiment's specifications. Finally, the well-trained models served as a backend to aid the experimental groups' idea-generation process. A user-friendly interface was also provided as a supporting tool to help the participants use the machine learning models more efficiently. Depending on their choices, the participants could produce new pictures by combining the features and structures of the provided images. The workflow was as follows: (1) Participants needed to choose one image as the original image and a second image as the reference image. (2)

Table 4. Number of training data used for each experimental session.

Experimental session	Number of training data
Session 1	34172
Session 2	1471
Session 3	3275
Session 4	4671
Session 5	3297

The participants were required to create a mask on the reference image to indicate the precise region that needed to be transferred to the original image. (3) The participants would then adjust the position of the mask on the original image to indicate what part of the reference image needed to be transferred. (4) Finally, the supporting system generated a new image automatically by transferring the selected part of the reference image into the original image via a natural transformation. The generated images possibly gave inspiration for combining the features and structures of the provided images while matching the requirements of the object's intended use.

A short introduction and description of the design task were provided prior to the start of each experimental session. All participants were required to document their ideas with sketches and descriptions on a piece of white paper that was provided beforehand. Participants were required to record as many of their ideas as possible until no more could be generated. At the end of each experimental session, all idea sheets were gathered, and their identification information was erased in preparation for data analysis. All answer sheets were then thoroughly shuffled randomly within their respective experimental groups, and any information that could potentially indicate the participants' group or personal identities was meticulously removed to ensure fairness in the subsequent scoring process.

5.2. Data analysis

For the scientific study of the collected data, this paper used the evaluation metric provided by Chan et al. (2011) and Shah, Smith, and Vargas-Hernandez (2003) to evaluate ideation outputs. This standard assessment tool is often used in the design science literature, which demonstrates its validity and fairness. The evaluation considered the novelty, quality, variety and quantity of the ideation output. Novelty was used to determine whether the proposed design solution had sufficient originality and uniqueness, and it also represented the degree of uncommonness of a particular solution within all possible solutions to a design task. Quality was used to evaluate the functional integrity of the object or system. The design solution needed to meet the minimum functionality requirement of the design task. Taking the perfume bottle design task in the experimental sessions as an example, the purpose of the design task was to combine the features and structures in the provided images and design a new perfume bottle given images consisting of three pineapple images and three perfume bottle images. The quality evaluation consisted of determining whether the designed perfume bottle contained features and structures from the given image – which could have been the combination of a pineapple and a perfume bottle or the combination of two perfume bottles – and whether the designed perfume bottle would

Table 5. Number of ideas generated for each experimental session.

Experimental session	Number of ideas
Session 1	209
Session 2	266
Session 3	214
Session 4	235
Session 5	209
Over all	1133

Table 6. The number of ideas generated in each session and the average number of ideas generated per participant for experimental and control groups.

	Total number of ideas		Average number of ideas	
Experimental session	Experimental group	Control group	Experimental group	Control group
Session 1	**105**	104	**5.83**	5.78
Session 2	**142**	124	**7.89**	6.89
Session 3	**119**	95	**6.61**	5.28
Session 4	**128**	107	**7.11**	5.94
Session 5	**115**	94	**6.39**	5.22
Over all	**609**	524	**33.83**	29.11

be suitable for daily use as a perfume bottle. Variety quantifies the variety of design solutions for a particular design task. A high variety score denotes little similarity between the design solutions, which leads to a greater possibility of finding a suitable design idea. Quantity records the number of generated design outcomes. For the design task in this paper, the number of design solutions produced by each individual in each experimental session was calculated for further analysis. These four evaluation metrics mentioned above were usually measured separately.

Consensual assessment technique (CAT) is the most popular metric for evaluating the performance of creative work samples based on the intuitions of qualified raters on their understanding of the design tasks (Amabile 1983). CAT has the benefit of capturing qualities of creative work that are difficult to assess or describe objectively. Based on the above advantages, the collected data's novelty, quality and variety were measured using CAT. Two experts in the product design field were invited to rate the collected data. Each rater had extensive product design experience and had completed at least three years of relevant coursework. The degree of rater expertise in our study was comparable to that of previous studies using CAT (L. Chen, Wang et al. 2019; Daly et al. 2016). Two raters subjectively rated each design solution collected from the experiment. The degree of the judges' consensus was determined by comparing the scores for the same design solution given by different judges.

Before starting the rating phase, we collected and randomly shuffled all the ideation sheets according to the experimental sessions. On the premise that the two raters did not know the ideas were generated using different ideation approaches, they were required to give novelty, quality and variety scores for all ideas independently. The number of ideas generated by the participants for each experimental session is listed in Table 5. The number of ideas generated in each session and the average number of ideas generated per participant for experimental and control groups is listed in Table 6

Table 7. Percentage of adjacent agreement.

Metrics	Experimental session 1	Experimental session 2	Experimental session 3	Experimental session 4	Experimental session 5	Entire Experiment
Quality	82.7751%	90.6015%	83.1775%	88.5106%	89.9521%	87.2021%
Novelty	81.3397%	85.7142%	90.6542%	89.3617%	92.8229%	87.9082%
Variety	88.8888%	94.4444%	97.2222%	100%	94.4444%	95%

Table 8. Calculated Cohen's kappa value.

Metrics	Experimental session 1	Experimental session 2	Experimental session 3	Experimental session 4	Experimental session 5	Entire Experiment
Quality	0.6534	0.7611	0.6081	0.6728	0.6989	0.6849
Novelty	0.6064	0.6031	0.7197	0.5879	0.7159	0.6439
Variety	0.5056	0.5804	0.7872	0.7755	0.5319	0.6309

The scoring range used in this case study is consistent with CAT on a scale of 1 to 7, with 1 representing the lowest score for novelty or quality of a design idea and 7 representing the highest. The two raters provided a score on each ideation sheet based on their knowledge of the design area and the design task. The raters performed several rounds of scoring based on the novelty and quality of each generated idea. In the first round of scoring, the raters assigned each idea an intuitive score. During the following scoring rounds, the raters reevaluated the prior scores and determined if the initial score should be altered. The scoring process concluded when there were no further changes to the idea scores. Overall, each rater rated 1133 generated ideas for their novelty and quality. The variety was evaluated based on individual participants using a similar approach as CAT. The raters provided a variety score between 1 and 7 to the whole collection of ideas generated by each participant during an experimental session. Score 1 represents no variation, and score 7 represents the most variation. The raters evaluated 190 sets of ideas in total. The quantity was calculated according to the experimental sessions on the number of ideas generated by the experimental groups and the control groups.

To better understand the consensus between raters on each metric, Stemler (2004) proposed a calculation of a computed percentage of the adjacent agreement. For the score of two raters to reach a consensus, the score from one rater needed to differ from another rater within one point above or below. Based on the calculation, the overall percentage of the adjacent agreement was 87.2021% for quality, 87.9082% for novelty, and 95% for variety. The details for adjacent agreement for each experimental session are listed in Table 7.

The consistency of the two raters for each metric was calculated using Cohen's kappa. The scores were divided into three categories: scores lower than 2 were 'low'; scores greater than 5 were 'high'; scores between 2 and 5 were 'normal'. All the calculated Cohen's kappa values are greater than 0.50, indicating that the scoring results are acceptable and reliable according to the CAT approach. The detailed results are shown in Table 8.

5.3. Results

All the scores given by the scorers were used for the data analysis. The score of each design solution was the average of the scores given by the two raters. The Shapiro-Wilk test was used to determine the normality of the rated scores. The resulting *p*-value from the score

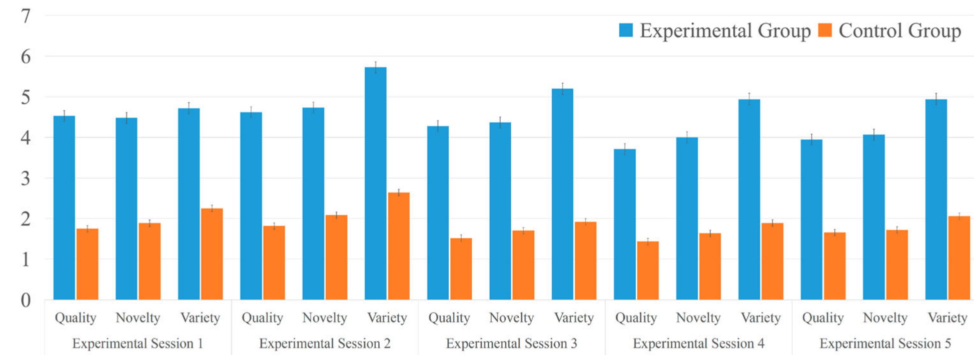


Figure 11. Average CAT score for each experimental session.

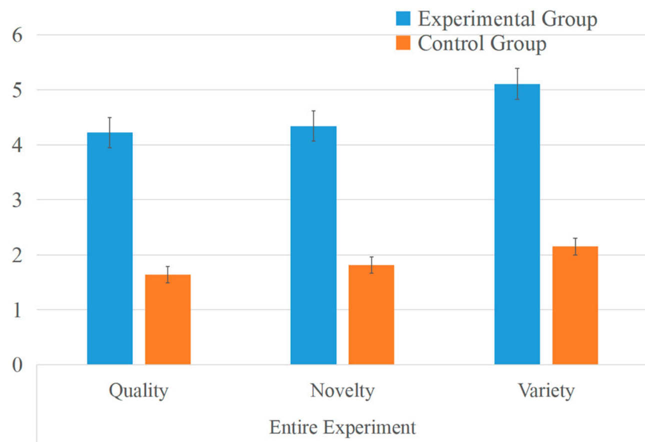


Figure 12. Average CAT score for the entire experiment.

data was less than 0.05, indicating that the scores did not fall into normal distributions. The independent Mann-Whitney U test was used next to determine if there was a significant difference in the mean of the scores between the experimental group and the control group. In terms of novelty, quality and variety, the *p*-value was close to 0 for each experimental session and the overall experiment, indicating that the experimental group outperformed the control group across all metrics. The detailed average score results between different groups for each experimental session is shown in Figure 11. The detailed average score results between different groups for the entire experiment is shown in Figure 12. The detailed novelty, quality, and variety scores for both the experimental and control groups are shown in Table 9. As illustrated in Table 9, Figures 11 and 12, the experimental group surpassed the control group by approximately 2 points in each experimental session and the entire experiment.

In terms of quantity, the number of ideas generated for each group in each experimental session was counted. According to the counted results, the experimental group generated more ideas than the control group. The detailed numbers are shown in Table 6. Several examples of low-score ideas and high-score ideas are shown in Figure 13.

Table 9. Detailed novelty, quality, and variety scores for both the experimental and control groups.

	Experimental group	Control group	
Session 1	Quality	4.528	1.745
	Novelty	4.480	1.88
	Variety	4.722	2.25
Session 2	Quality	4.623	1.814
	Novelty	4.735	2.080
	Variety	5.722	2.638
Session 3	Quality	4.281	1.515
	Novelty	4.369	1.7
	Variety	5.194	1.916
Session 4	Quality	3.710	1.429
	Novelty	4.003	1.630
	Variety	4.944	1.888
Session 5	Quality	3.947	1.654
	Novelty	4.069	1.718
	Variety	4.944	2.055
Average	Quality	4.220	1.639
	Novelty	4.340	1.815
	Variety	5.105	2.15

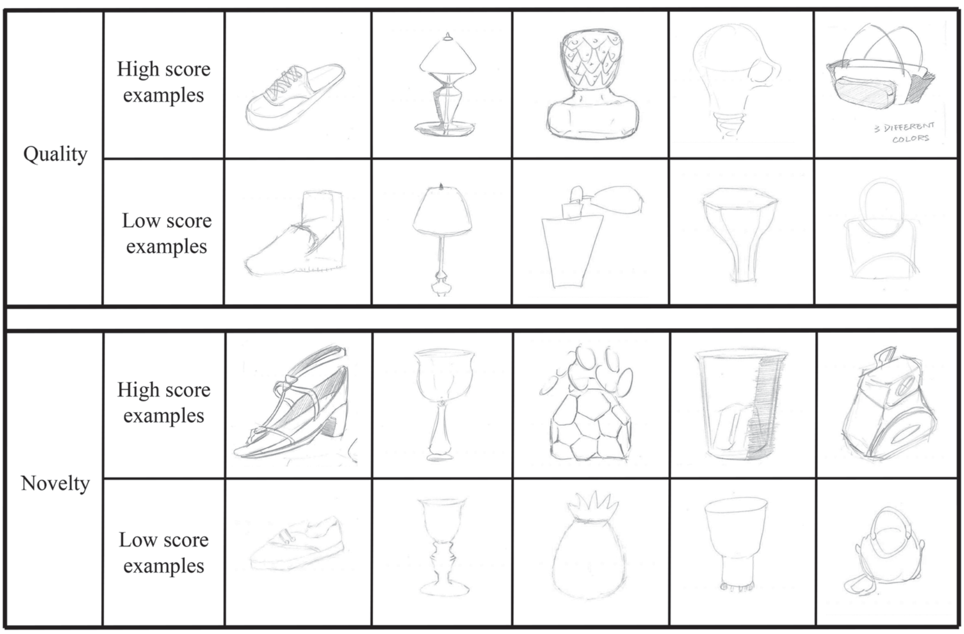


Figure 13. Examples of low-score and high-score ideas. It is evident that under the quality and novelty metrics, high-score examples demonstrate superior design quality and novelty, whereas low-score examples lack innovation and exhibit poor sketch quality.

5.4. Discussion

The scores given by the expert raters in terms of novelty, quality, variety and quantity were undoubtedly better for the experimental group than for the control group. In order to get a deeper understanding of how the object combination model might aid in the ideation process, this study explores the experiment results in further depth.

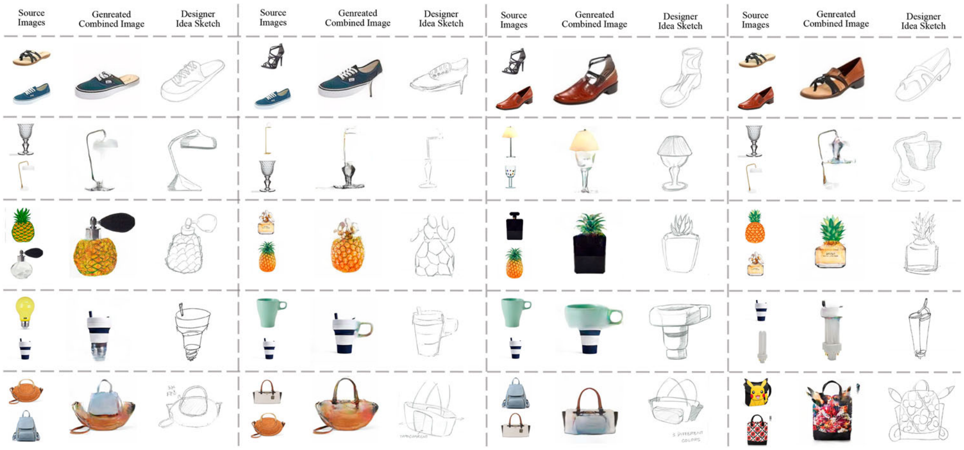


Figure 14. Selected model output and corresponding idea sketches for each experimental session.

First, to demonstrate that the model's object combination results could have helped the experimental group in their ideation process, we select two model outputs from each experimental session and the participants' design sketches based on this output, as shown in Figure 14. It is obvious that the images generated by the object combination model are an excellent combination of the features and structures of the two selected images, and the related idea sketch is relevant to the generated image. This result is an excellent illustration of the effect the model has on the ideation process.

Next, we analysed the average novelty, quality, variety and quantity scores in detail. Regarding quality, the experimental group performed approximately 2.5 points better than the control group on average CAT scores. As seen in Figure 13, the high-scoring examples of idea sketches from the experimental group exhibit excellent object combination visuals and high-level object integrity and detail while meeting the item's fundamental purposes. However, the low-scoring examples of sketches from the control group did not adequately match the above-mentioned criteria. This may be because the model gave the experimental group excellent combination results depending on their preferences, allowing them to generate ideas and sketch designs with great ease. The experimental group was separated from the control group by approximately 2 CAT points in terms of novelty, which was made possible by the model's excellent interactive interface for the participants. The interface enabled users to choose any part of any necessary shape from the reference image and transfer it to any location within the original image. The interface provided limitless options for object combinations, and the vast number of potential combinations made it simpler to generate innovative design ideas. The average variety scores were affected for the same reason. The more possible combinations there were, the more diversified the ideas generated were. The control group did not have access to the model, so they had to combine the objects based on their intuition, which was less efficient. Consequently, the average variety score of the experimental group is roughly 3 points higher than that of the control group. In the aspect of quantity, throughout each experimental session, the experimental group generated one or two more ideas per participant than the control group.

Table 10 shows the statistical findings of the t-test on the ratings of all metrics, provid-

Table 10. Detailed results of the t-test on all metrics (Quality, novelty, and variety) accors all experimental sessions.

Metrics	Sessions	P value	95% Confidence interval	T-statistic value	Standard error of difference
Quality	Session 1	< 0.0001	2.53 to 3.04	21.2783	0.131
	Session 2	< 0.0001	2.60 to 3.01	26.9852	0.104
	Session 3	< 0.0001	2.57 to 2.96	28.1784	0.098
	Session 4	< 0.0001	2.10 to 2.47	24.341	0.094
	Session 5	< 0.0001	2.07 to 2.52	20.1797	0.114
	Overall	< 0.0001	2.48 to 2.68	51.9833	0.050
Novelty	Session 1	< 0.0001	2.36 to 2.83	21.5048	0.121
	Session 2	< 0.0001	2.43 to 2.88	23.3214	0.114
	Session 3	< 0.0001	2.48 to 2.86	27.3342	0.098
	Session 4	< 0.0001	2.19 to 2.56	24.866	0.095
	Session 5	< 0.0001	2.13 to 2.58	20.5594	0.114
	Overall	< 0.0001	2.43 to 2.62	50.7473	0.053
Variety	Session 1	< 0.0001	1.89 to 3.05	8.5247	0.29
	Session 2	< 0.0001	2.44 to 3.73	9.5433	0.323
	Session 3	< 0.0001	2.76 to 3.79	12.6655	0.259
	Session 4	< 0.0001	2.57 to 3.54	12.656	0.241
	Session 5	< 0.0001	2.31 to 3.47	9.9705	0.29
	Overall	< 0.0001	2.70 to 3.21	22.9162	0.129

Note: **P value** means that there is a less than 0.01% probability that the observed difference between the two groups occurred by chance(usually a P value less than 0.05). **95% Confidence Interval** stands for 95% confident that the true difference in means between the two groups. **T-statistic Value** is the calculated t-statistic for the t-test, A higher t-value indicates a greater difference between the groups. **Standard Error of Difference** is the standard error of the difference between the two group means, It provides a measure of the variability of the difference.

ing a detailed comparison between the experimental and control groups across multiple sessions. The metrics analysed include Quality, Novelty, and Variety, with results indicating significant differences favouring the experimental group. The Quality metric consistently showed extremely statistically significant differences across all sessions. The P values for each session were less than 0.0001, indicating that the probability of these differences occurring by chance is exceedingly low. The 95% confidence intervals for the difference in Quality scores ranged from approximately 2.07 to 3.04, with the experimental group consistently scoring higher than the control group. The t-statistic values ranged from 20.1797 to 28.1784, demonstrating a strong effect of the experimental intervention. The overall analysis across all sessions reinforced these findings, with an overall mean difference of 2.48 to 2.68 points, highlighting the substantial improvement in Quality scores due to the experimental conditions. Similar to Quality, the Novelty metric also exhibited highly significant differences in favour of the experimental group. Each session's P value was less than 0.0001, underscoring the robustness of these findings. The 95% confidence intervals for Novelty scores ranged from approximately 2.13 to 2.88, with the experimental group consistently achieving higher scores. The t-statistic values varied from 20.5594 to 27.3342, indicating a significant impact of the experimental intervention on Novelty. The overall analysis across all sessions confirmed these results, with an average difference of 2.43 to 2.62 points. This consistently higher performance in Novelty scores by the experimental group underscores the effectiveness of the experimental conditions in fostering creativity and originality. The Variety metric also showed significant differences favouring the experimental group, with P values for each session being less than 0.0001. The 95% confidence intervals for Variety scores ranged from approximately 1.89 to 3.79, indicating a substantial difference between the experimental and control groups. The t-statistic values ranged from 8.5247 to 12.6655, supporting the presence of a strong effect due to the experimental intervention. The overall analysis revealed a mean difference of 2.70 to 3.21 points, confirming that the experimental group consistently outperformed the control group in terms of Variety. This suggests that the experimental conditions effectively promoted a broader range of ideas and solutions.

The results of the t-test on all metrics (Quality, novelty, and variety) demonstrate that the experimental group significantly outperformed the control group. The consistently low P values (< 0.0001) indicate that these differences are highly unlikely to be due to chance. The confidence intervals and t-statistic values further validate the robustness of these findings. Overall, the experimental intervention proved to be highly effective in enhancing the Quality, Novelty, and Variety of the outcomes, providing strong evidence for the superiority of the experimental conditions over the control conditions.

Finally, we conducted two questionnaire surveys with participants from the experimental group to evaluate both the effectiveness of the design support system and their perceptions of its usability. The first questionnaire assessed the quality of the generated images and the system's ability to inspire design tasks. The second questionnaire measured visual design intuitiveness, interaction design intuitiveness, and system usability to further validate the user-friendliness of the interface. Regarding the quality of the generated images, 67% of participants rated them as excellent, 27% as good, and 6% as poor. When asked whether the system provided inspiration, 83% of participants reported that it offered excellent support for their design tasks, 11% found it somewhat inspiring, and 6% indicated that it provided no inspiration. These findings highlight the system's effectiveness in facilitating the ideation process.

To assess the user-friendliness of the interface, we adapted questions from the standard System Usability Scale (SUS) questionnaire (Brooke 1996) to systematically evaluate system usability. Responses indicated that all users held neutral or positive views regarding the functionalities and ease of interaction within the image selection area, image processing area, and generation area. Furthermore, the majority of participants agreed or strongly agreed that the functionalities of each section were intuitive and that the interaction methods were straightforward and easy to use. In terms of System Usability, most users expressed a willingness to use the system regularly. Over 60% of users strongly agreed that the system was easy to use, while more than 70% strongly agreed that it was easy to learn. Additionally, all users acknowledged that the system's various functions were well integrated and expressed confidence in using the system effectively. The details are shown in Figure 15.

Overall, the suggested integrated approach focuses on combining the features and structures of two objects based on the user's preference to generate an image of the combined object. The created image may serve as a visual stimulus that provides designers with visual inspiration for design tasks. Compared to previous design-supporting tools, our approach provides synthesis results at a visual level rather than a semantic word level, allowing for more immediate inspiration during the ideation process. However, due to the proposed approach requiring high-quality visual-level stimuli to help the designer's ideation process, a high-quality training dataset is required for model training. The dataset used for model training needs to maintain a similar feature space distribution, where objects in the images should have a certain degree of structural similarity, positioned in the centre of the images, and have minimal background complexity. These factors will

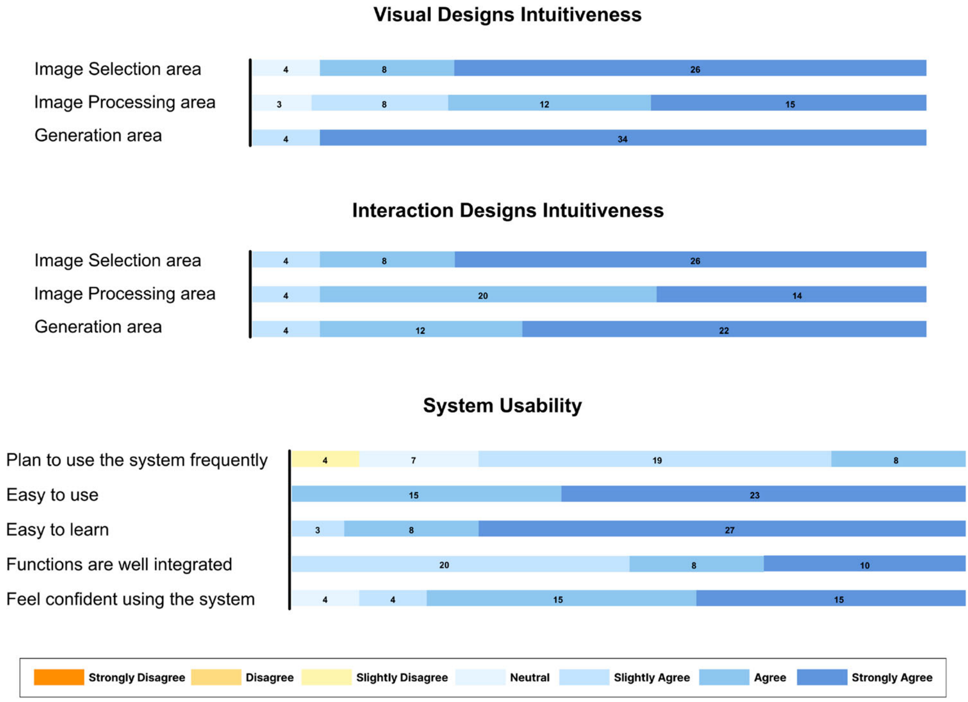


Figure 15. The results of user feedback and usability assessment questionnaire.

help model training process and achieve the best results. Additionally, the number of categories in the training dataset should not be excessive, as too many categories can affect the model's ability to learn the data distribution effectively, thus impacting the model's generative performance. For instance, objects like pineapples and perfume bottles have a certain structural similarity and are suitable for training the dataset to achieve image fusion between these categories. However, categories like pineapples and tables lack structural similarity and are not suitable for model training. Another concern arising from the case study is the hand-picked selection process for the case study images. Instead of providing the entire test set, we randomly selected images from the test dataset for each case study session. This approach was chosen to prevent overwhelming users with an excessive number of image options, which could cause fatigue and distraction, ultimately prolonging each experimental session. The images were randomly selected only once during the case study design stage to ensure fairness across experimental groups. However, this method inevitably introduces selection bias, as it relies on subjective judgment and may unintentionally favour specific data characteristics. Future study could use the entire test dataset for the case study to help mitigate these biases.

6. Limitations and future works

Through the aforementioned design case study, we can see that our proposed object combination model not only facilitates high-quality image generation and local editing tasks but also aids designers in their ideation process, thereby enhancing the quality, novelty, variety, and quantity of design outcomes. However, our proposed system also has certain limitations that need to be discussed in detail.

Firstly, one limitation of this study is the restricted number of stimuli images provided to participants. While six images were chosen to balance cognitive load and maintain experimental consistency, a larger selection might have offered additional inspiration. Future work could explore the effects of expanding the image set on ideation diversity and participant engagement.

Secondly, the proposed model has inherent limitations. Since it is based on a deep learning generative model, it is highly dependent on the dataset. High-quality and extensive training data can effectively improve the model's performance. If the training data is noisy or biased, it may result in lower output quality. Additionally, if the amount of training data is insufficient, the model may not effectively understand the relationships between image features, thereby affecting the local editing performance. This is clearly reflected in the results of our model, as demonstrated in Tables 1 and 4. It is evident that our collected dataset exhibits a certain level of bias, with the number of shoes data significantly exceeding that of other categories. This imbalance leads to the models trained on shoes data performing notably better than those trained on other categories, as observed in our experimental results and case study. Increasing the amount of data for other categories should help improve the model's performance for those specific categories. Moreover, because the generative model learns the data distribution within the dataset to perform image generation tasks, significant structural and color differences among the objects in the training dataset can severely impact the training process, potentially causing mode collapse and thus affecting the model's performance. Consequently, our model is more adept at handling design tasks with certain structural similarities, which limits its application in the design field. It can perform well in specific design domains but not as well in other design tasks. Furthermore, the current model also faces generalisation issues for design tasks. It requires collecting a specific dataset and training a specific model for each design task, preventing it from being trained on a large dataset and generalised to multiple different design tasks, thereby limiting the system's versatility.

Integrating the proposed approach into existing design workflows also presents certain challenges. Firstly, designers may have diverse ways of obtaining inspiration, which vary significantly. The approach proposed in this paper involves generating new design ideas by combining features of existing objects, which may not be a universal method for all designers and design tasks. Secondly, when designers obtain inspiration by combining existing objects, it should not limited to combining two parts but should combine different parts of various objects. These tasks are not suitable for the proposed approach.

Lastly, the proposed approach is not very suitable for complex or abstract design tasks. Our model inspires designers on a visual level by generating images, making it more adept at handling explicit image combination and generation tasks. The proposed system is not suitable for complex design tasks or abstract designs that require more implicit information. However, the proposed model can be adjusted based on specific design tasks, it can inspire designers' creativity through explicit visual stimuli, thereby stimulating more implicit design ideas. Although highly complex and abstract design requirements are not the strengths of the proposed approach, the output requirements of the proposed model can be adjusted from generating explicit, clear design images to generating relatively abstract design patterns, it can also provide inspiration for highly complex and abstract design tasks, ultimately achieving collaborative design with human designers.

For future works: (1) We will collect a large-scale, high-quality image dataset specifically tailored for the design field. This dataset will enhance the model's performance and generalisation capabilities, thereby expanding its application scope in design area. (2) To overcome the current limitations of the model in generating multi-category items and detailed textures, we will first attempt to replace the existing generative model with more effective GAN or Diffusion models to achieve better image pixel-level integration. Additionally, we will experiment with incorporating a texture feature encoder and class labels as additional conditional information into the proposed model. These enhancements are expected to improve texture generation, cross-category image generation, and cross-category image integration capabilities. (3) Finally, to make our proposed tool more user-friendly for designers, we will update the user interface to address its current shortcomings. Enhancements will include the addition of image zooming and rotation functionalities to improve the overall user experience for designers.

7. Conclusion

This paper presents the OC-GAN, a novel AI generative model specifically developed to enhance combinational creativity in design ideation. Unlike conventional methods heavily dependent on textual stimuli or expert knowledge, OC-GAN effectively supports combining the features and structures of two objects depending on the user's preference and generating an image of the combined object. The intuitive human-computer interaction interface reduces cognitive demands and enhances accessibility, especially benefiting designers with limited artistic expertise. Our comprehensive design case study demonstrates that the proposed OC-GAN framework significantly improves ideation efficiency, Quantitative analysis of the case study data support the conclusion that the suggested method can produce high-quality visual stimuli for the ideation process and increase the quality, novelty, variety and quantity of ideas produced. This research underscores the potential of human-AI co-ideation systems to meaningfully enhance creativity in design ideation, paving the way for practical applications across diverse creative domains such as product design, architecture, fashion, and visual arts, particularly benefiting scenarios that require rapid conceptualisation, intuitive visualisation, and reduced cognitive demands on designers.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work described in this paper was supported by a grant from the Laboratory for Artificial Intelligence in Design and Hong Kong Polytechnic University (Project No. P0042736).

ORCID

Pan Wang http://orcid.org/0000-0001-6575-539X
Xun Zhang http://orcid.org/0009-0003-7693-1805
Liyan Wei http://orcid.org/0009-0008-6311-9726

Peter Childs http://orcid.org/0000-0002-2465-8822
Stephen Jia Wang http://orcid.org/0000-0001-9835-9932
Yike Guo http://orcid.org/0000-0002-3075-2161
Maaike Kleinsmann http://orcid.org/0000-0003-4120-5048

References

Abdal, R., Y. Qin, and P. Wonka. 2019. "Image2stylegan: How to Embed Images into the Stylegan Latent Space?" In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Seoul, Korea (South), 4432–4441.

Abdal, R., Y. Qin, and P. Wonka. 2020. "Image2stylegan++: How to Edit the Embedded Images?" In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 8296–8305.

Alharbi, Y., and P. Wonka. 2020. "Disentangled Image Generation through Structured Noise Injection." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 5134–5142.

Amabile, T. M. 1983. "The Social Psychology of Creativity: A Componential Conceptualization." *Journal of Personality and Social Psychology* 45:357. https://doi.org/10.1037/0022-3514.45.2.357.

Archana Balkrishna, Y. 2024. "An Analysis on the Use of Image Design with Generative AI Technologies." *International Journal of Trend in Scientific Research and Development* 8:596–599.

Bacciotti, D., Y. Borgianni, and F. Rotini. 2016. "An Original Design Approach for Stimulating the Ideation of New Product Features." *Computers in Industry* 75:80–100. https://doi.org/10.1016/j.compind.2015.06.004.

Beitz, W., G. Pahl, and K. Grote. 1996. "Engineering Design: A Systematic Approach." *Mrs Bulletin* 71:30.

Boden, M. A. 2004. *The Creative Mind: Myths and Mechanisms*. Routledge.

Briggs, R. O., and B. A. Reinig. 2007. "Bounded Ideation Theory: A New Model of the Relationship Between Ideaquantity and Idea-quality During Ideation." In *2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07)*, Waikoloa, HI, USA, 16–16. IEEE.

Brock, A., J. Donahue, and K. Simonyan. 2018. "Large Scale Gan Training for High Fidelity Natural Image Synthesis." Preprint arXiv:1809.11096.

Brooke, J. 1996. "Sus-a Quick and Dirty Usability Scale." *Usability Evaluation in Industry* 189:4–7.

Cao, S., W. Chai, S. Hao, Y. Zhang, H. Chen, and G. Wang. 2023. "Difffashion: Reference-Based Fashion Design with Structure-Aware Transfer by Diffusion Models." *IEEE Transactions on Multimedia*26:3962–3975. https://doi.org/10.1109/TMM.2023.3318297.

Chan, J., K. Fu, C. Schunn, J. Cagan, K. Wood, and K. Kotovsky. 2011. "On the Benefits and Pitfalls of Analogies for Innovative Design: Ideation Performance Based on Analogical Distance, Commonness, and Modality of Examples." *Journal of Mechanical Design* 133 (8): 081004. https://doi.org/10.1115/1.4004396.

Chen, L., Z. Cai, Z. Jiang, J. Luo, L. Sun, P. Childs, and H. Zuo. 2024. "Asknaturenet: A Divergent Thinking Tool Based on Bio-Inspired Design Knowledge." *Advanced Engineering Informatics* 62:102593. https://doi.org/10.1016/j.aei.2024.102593.

Chen, W., and M. Fuge. 2017. "Beyond the Known: Detecting Novel Feasible Domains over An Unbounded Design Space." *Journal of Mechanical Design* 139 (11): 111405. https://doi.org/10.1115/1.4037306.

Chen, Y., and N. Li. 2024. "A Study of Style Migration Generation of Traditional Chinese Portraits Based on Dualstylegan." *Journal of Engineering Design* 1–14.

Chen, X., Z. Ma, X. Jiang, Y. Jian, X. Yao, and P. Wu. 2024. "Lumos: Ai-Driven Prompt Optimisation Tool for Assisting Conceptual Design." *Journal of Engineering Design* 1–27.

Chen, L., Y. Song, J. Guo, L. Sun, P. Childs, and Y. Yin. 2025. "How Generative Ai Supports Human in Conceptual Design." *Design Science* 11:e9. https://doi.org/10.1017/dsj.2025.2.

Chen, L., P. Wang, H. Dong, F. Shi, J. Han, Y. Guo, P. R. Childs, J. Xiao, and C. Wu. 2019. "An Artificial Intelligence Based Data-Driven Approach for Design Ideation." *Journal of Visual Communication and Image Representation* 61:10–22. https://doi.org/10.1016/j.jvcir.2019.02.009.

Chen, L., S. Xiao, Y. Chen, L. Sun, P. R. Childs, and J. Han. 2024. “An Artificial Intelligence Approach for Interpreting Creative Combinational Designs.” *Journal of Engineering Design* 1–28.

Chen, L., Y. Zhang, J. Han, L. Sun, P. Childs, and B. Wang. 2024. “A Foundation Model Enhanced Approach for Generative Design in Combinational Creativity.” *Journal of Engineering Design* 35 (1): 1–27. https://doi.org/10.1080/09544828.2023.2290914.

Childs, P., J. Han, L. Chen, P. Jiang, P. Wang, D. Park, Y. Yin, E. Dieckmann, and I. Vilanova. 2022. “The Creativity Diamond–a Framework to Aid Creativity.” *Journal of Intelligence* 10:73. https://doi.org/10.3390/jintelligence10040073.

Chilton, L. B., S. Petridis, and M. Agrawala. 2019. “Visiblends: A Flexible Workflow for Visual Blends.” In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 1–14.

Chou, J. R. 2014. “An Ideation Method for Generating New Product Ideas Using Triz, Concept Mapping, and Fuzzy Linguistic Evaluation Techniques.” *Advanced Engineering Informatics* 28 (4): 441–454. https://doi.org/10.1016/j.aei.2014.06.006.

Chou, J. R. 2024. “An Integrative Review Exploring the Development of Sustainable Product Design in the Technological Context of Industry 4.0.” *Advanced Engineering Informatics* 62:102689. https://doi.org/10.1016/j.aei.2024.102689.

Collins, E., R. Bala, B. Price, and S. Susstrunk. 2020. “Editing in Style: Uncovering the Local Semantics of Gans.” In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 5771–5780.

Daly, S. R., C. M. Seifert, S. Yilmaz, and R. Gonzalez. 2016. “Comparing Ideation Techniques for Beginning Designers.” *Journal of Mechanical Design* 138 (10): 101108. https://doi.org/10.1115/1.4034087.

Donahue, J., P. Krähenbühl, and T. Darrell. 2017. “Adversarial Feature Learning.” In *International Conference on Learning Representations,* Toulon, France.

Frigotto, M. L., and M. Riccaboni. 2011. “A Few Special Cases: Scientific Creativity and Network Dynamics in the Field of Rare Diseases.” *Scientometrics* 89:397–420. https://doi.org/10.1007/s11192-011-0431-9.

Fu, K., D. Moreno, M. Yang, and K. L. Wood. 2014. “Bio-Inspired Design: An Overview Investigating Open Questions from the Broader Field of Design-by-Analogy.” *Journal of Mechanical Design* 136 (11): 111102. https://doi.org/10.1115/1.4028289.

Goodfellow, I. J., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. 2014. “Generative Adversarial Networks.” https://arxiv.org/abs/1406.2661.

Han, J., P. R. Childs, and J. Luo. 2024. “Applications of Artificial Intelligence and Cognitive Science in Design.” *Ai Edam* 38:e6.

Han, J., F. Shi, L. Chen, and P. Childs. 2016. “The Combinator: A Computer-Based Tool for Idea Generation.” In *DS 84: Proceedings of the DESIGN 2016 14th International Design Conference*, 639–648.

Han, J., F. Shi, L. Chen, and P. R. Childs. 2018. “The Combinator–a Computer-Based Tool for Creative Idea Generation Based on a Simulation Approach.” *Design Science* 4:e11. https://doi.org/10.1017/dsj.2018.7.

Hao, J., Y. Zhou, Q. Zhao, and Q. Xue. 2019. “An Evolutionary Computation Based Method for Creative Design Inspiration Generation.” *Journal of Intelligent Manufacturing* 30 (4): 1673–1691. https://doi.org/10.1007/s10845-017-1347-x.

Hayakawa, R., R. Noji, and T. Kato. 2024. “Generating Tyre Tread Designs Using a Sensory Evaluation Regression Model and a Generative Model.” *Journal of Engineering Design* 36:19–51. https://doi.org/10.1080/09544828.2024.2411487.

Henriksen, D., and P. Mishra. 2014. “Twisting Knobs and Connecting Things: Rethinking Technology & Creativity in the 21st Century.” *TechTrends* 58:15. https://doi.org/10.1007/s11528-013-0713-6.

Heyrani Nobari, A., M. F. Rashad, and F. Ahmed. 2021. “Creativegan: Editing Generative Adversarial Networks for Creative Design Synthesis. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Virtual, Online, V03AT03A002. American Society of Mechanical Engineers.

Honda, S., H. Yanagisawa, and T. Kato. 2022. “Aesthetic Shape Generation System Based on Novelty and Complexity.” *Journal of Engineering Design* 33 (12): 1016–1035. https://doi.org/10.1080/09544828.2022.2155343.

Isola, P., J. Y. Zhu, T. Zhou, and A. A. Efros. 2017. "Image-to-Image Translation with Conditional Adversarial Networks." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, Hawaii, 1125–1134.

Jiang, S., and Y. Fu. 2017. "Fashion Style Generator." In *IJCAI*, Melbourne, Australia, 3721–3727.

Jiang, Z., H. Wen, F. Han, Y. Tang, and Y. Xiong. 2022. "Data-driven Generative Design for Mass Customization: A Case Study." *Advanced Engineering Informatics* 54:101786. https://doi.org/10.1016/j.aei.2022.101786.

Jin, J., M. Yang, H. Hu, X. Guo, J. Luo, and Y. Liu. 2024. "Empowering Design Innovation Using Ai-generated Content." *Journal of Engineering Design* 36:1–18. https://doi.org/10.1080/09544828.2024.2401751.

Karadag, I., O. Z. Güzelci, and S. Alaçam. 2023. "Edu-Ai: A Twofold Machine Learning Model to Support Classroom Layout Generation." *Construction Innovation* 23 (4): 898–914. https://doi.org/10.1108/CI-02-2022-0034.

Karras, T., S. Laine, and T. Aila. 2019. "A Style-Based Generator Architecture for Generative Adversarial Networks." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, 4401–4410.

Karras, T., S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila. 2020. "Analyzing and Improving the Image Quality of Stylegan." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 8110–8119.

Khan, S., K. Goucher-Lambert, K. Kostas, and P. Kaklis. 2023. "Shiphullgan: A Generic Parametric Modeller for Ship Hull Design Using Deep Convolutional Generative Model." *Computer Methods in Applied Mechanics and Engineering* 411:116051. https://doi.org/10.1016/j.cma.2023.116051.

Kim, H., Y. Choi, J. Kim, S. Yoo, and Y. Uh. 2021. "Exploiting Spatial Dimensions of Latent in Gan for Real-Time Image Editing." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, TN, USA, 852–861.

Laing, S. J. 2018. "The Role of Images in Support of Graphic Design Ideation." PhD diss., The University of Waikato.

Lee, C., J. Liang, K. Yung, and K. Keung. 2024. "Generating Triz-Inspired Guidelines for Eco-Design Using Generative Artificial Intelligence." *Advanced Engineering Informatics* 62:102846. https://doi.org/10.1016/j.aei.2024.102846.

Li, X., J. Su, Z. Zhang, and R. Bai. 2021. "Product Innovation Concept Generation Based on Deep Learning and Kansei Engineering." *Journal of Engineering Design* 32 (10): 559–589. https://doi.org/10.1080/09544828.2021.1928023.

Li, W., W. Zhang, W. Wu, and J. Xu. 2024. "Exploring Human-Machine Collaboration Paths in the Context of Ai-Generation Content Creation: A Case Study in Product Styling Design." *Journal of Engineering Design* 36:298–324. https://doi.org/10.1080/09544828.2024.2396199.

Lin, C. F., Y. C. Yeh, Y. H. Hung, and R. I. Chang. 2013. "Data Mining for Providing a Personalized Learning Path in Creativity: An Application of Decision Trees." *Computers & Education* 68:199–210. https://doi.org/10.1016/j.compedu.2013.05.009.

Linsey, J. S., K. L. Wood, and A. B. Markman. 2008. "Increasing Innovation: Presentation and Evaluation of the Wordtree Design-by-Analogy Method. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Brooklyn, NY, USA, 21–32.

Liu, M., and Y. Hu. 2023. "Application Potential of Stable Diffusion in Different Stages of Industrial Design." In *International Conference on Human-Computer Interaction*, Copenhagen, Denmark, 590–609. Springer.

Liu, H., C. Li, Q. Wu, and Y. J. Lee. 2023. "Visual Instruction Tuning." https://arxiv.org/abs/2304.08485.

Lu, P., S. W. Hsiao, J. Tang, and F. Wu. 2024. "A Generative-Ai-based Design Methodology for Car Frontal Forms Design." *Advanced Engineering Informatics* 62:102835. https://doi.org/10.1016/j.aei.2024.102835.

Mou, C., Y. Wu, W. Wu, Z. Guo, P. Zhang, Y. Cheng, Y. Luo, et al. 2025. "Dreamo: A Unified Framework for Image Customization." Preprint arXiv:2504.16915.

Park, T., J. Y. Zhu, O. Wang, J. Lu, E. Shechtman, A. Efros, and R. Zhang. 2020. "Swapping Autoencoder for Deep Image Manipulation." *Advances in Neural Information Processing Systems* 33:7198–7211.

Pidhorskyi, S., D. A. Adjeroh, and G. Doretto. 2020. "Adversarial Latent Autoencoders." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 14104–14113.

Radford, A., L. Metz, and S. Chintala. 2015. "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks." Preprint arXiv:1511.06434.

Reinig, B. A., and R. O. Briggs. 2013. "Putting Quality First in Ideation Research." *Group Decision and Negotiation* 22 (5): 943–973. https://doi.org/10.1007/s10726-012-9338-y.

Ren, X., N. Wang, J. Pan, and L. Bu. 2024. "Combining Style Generative Adversarial Networks with Particle Swarm Optimisation-Support Vector Regression to Design Affective Social Robot for Public Health Intervention." *Journal of Engineering Design* 36:160–190. https://doi.org/10.1080/09544828.2024.2415830.

Rosati, R., P. Senesi, B. Lonzi, A. Mancini, and M. Mandolini. 2024. "An Automated Cad-to-xr Framework Based on Generative Ai and Shrinkwrap Modelling for a User-Centred Design Approach." *Advanced Engineering Informatics* 62:102848. https://doi.org/10.1016/j.aei.2024.102848.

Sarica, S., B. Song, J. Luo, and K. L. Wood. 2021. "Idea Generation with Technology Semantic Network." *Ai Edam* 35:265–283.

Sbai, O., M. Elhoseiny, A. Bordes, Y. LeCun, and C. Couprie. 2018. "Design: Design Inspiration from Generative Networks." In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops,* Munich, Germany.

Shah, J. J., S. M. Smith, and N. Vargas-Hernandez. 2003. "Metrics for Measuring Ideation Effectiveness." *Design Studies* 24 (2): 111–134. https://doi.org/10.1016/S0142-694X(02)00034-0.

Shi, F., L. Chen, J. Han, and P. Childs. 2017. "A Data-Driven Text Mining and Semantic Network Analysis for Design Information Retrieval." *Journal of Mechanical Design* 139 (11): 111402. https://doi.org/10.1115/1.4037649.

Smith, G., T. J. Troy, and J. D. Summers. 2006. "Concept Exploration through Morphological Charts: An Experimental Study." In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Philadelphia, PA, USA, 495–504.

Stemler, S. E. 2004. "A Comparison of Consensus, Consistency, and Measurement Approaches to Estimating Interrater Reliability." *Practical Assessment, Research, and Evaluation* 9:4.

Sun, Z., Z. Zhang, Y. Zhang, M. Lu, D. Lischinski, D. Cohen-Or, and H. Huang. 2025. "Creative Blends of Visual Concepts." In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. New York, NY: Association for Computing Machinery.

Sun, Z., Y. Zhou, H. He, and P. Mok. 2023. "Sgdiff: A Style Guided Diffusion Model for Fashion Synthesis." In *Proceedings of the 31st ACM International Conference on Multimedia*, Ottawa, ON, Canada, 8433–8442.

Suzuki, R., M. Koyama, T. Miyato, T. Yonetsuji, and H. Zhu. 2018. "Spatially Controllable Image Synthesis with Internal Representation Collaging." Preprint arXiv:1811.10153.

Tang, Y. C., J. J. Huang, M. T. Yao, J. Wei, W. Li, Y. X. He, and Z. J. Li. 2019. "A Review of Design Intelligence: Progress, Problems, and Challenges." *Frontiers of Information Technology & Electronic Engineering* 20 (12): 1595–1617. https://doi.org/10.1631/FITEE.1900398.

Toh, C. A., E. M. Starkey, C. S. Tucker, and S. R. Miller. 2017. "Mining for Creativity: Determining the Creativity of Ideas through Data Mining Techniques." In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Cleveland, OH, USA, V007T06A010. American Society of Mechanical Engineers.

Toivonen, H., and O. Gross. 2015. "Data Mining and Machine Learning in Computational Creativity." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 5 (6): 265–275.

Ulrich, K. T., and S. D. Eppinger. 2016. *Product Design and Development*. New York, NY: McGraw-Hill.

Venkataraman, S., B. Song, J. Luo, K. Subburaj, M. R. Elara, L. Blessing, and K. Wood. 2017. "Investigating Effects of Stimuli on Ideation Outcomes." In *DS 87-8 Proceedings of the 21st International Conference on Engineering Design (ICED 17), Vol 8: Human Behaviour in Design*, 309–318. Vancouver, Canada.

Wang, Z., Z. Tan, and Y. Ma. 2024. "Combinediff: A Genai Creative Support Tool for Image Combination Exploration." *Journal of Engineering Design* 1–27.

Wang, B., Y. Zhu, L. Chen, J. Liu, L. Sun, and P. Childs. 2023. "A Study of the Evaluation Metrics for Generative Images Containing Combinational Creativity." *Ai Edam* 37:e11.

Wu, J., Y. Cai, T. Sun, K. Ma, and C. Lu. 2024. "Integrating Aigc with Design: Dependence, Application, and Evolution-a Systematic Literature Review." *Journal of Engineering Design* 1–39. https://doi.org/10.1080/09544828.2024.2362587.

Wu, C., X. Chen, Z. Wu, Y. Ma, X. Liu, Z. Pan, W. Liu, et al. 2024. "Janus: Decoupling Visual Encoding for Unified Multimodal Understanding and Generation." https://arxiv.org/abs/2410.13848.

Yan, Z., J. Ye, W. Li, Z. Huang, S. Yuan, X. He, K. Lin, J. He, C. He, and L. Yuan. 2025. "Gpt-Imgeval: A Comprehensive Benchmark for Diagnosing gpt4o in Image Generation." https://arxiv.org/abs/2504.02782.

Yang, B., S. Gu, B. Zhang, T. Zhang, X. Chen, X. Sun, D. Chen, and F. Wen. 2023. "Paint by Example: Exemplar-Based Image Editing with Diffusion Models." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Vancouver, BC, Canada, 18381–18391.

Yang, H., and L. Zhang. 2016. "Promoting Creative Computing: Origin, Scope, Research and Applications." *Digital Communications and Networks* 2 (2): 84–91. https://doi.org/10.1016/j.dcan.2016.02.001.

Yilmaz, S., S. R. Daly, C. M. Seifert, and R. Gonzalez. 2015. "How Do Designers Generate New Ideas? Design Heuristics across Two Disciplines." *Design Science* 1:e4. https://doi.org/10.1017/dsj.2015.4.

Zang, T., M. Yang, Y. Liu, and P. Jiang. 2024. "Text2shape: Intelligent Computational Design of Car Outer Contour Shapes Based on Improved Conditional Wasserstein Generative Adversarial Network." *Advanced Engineering Informatics* 62:102892. https://doi.org/10.1016/j.aei.2024.102892.

Zhang, L., Z. Li, and Y. Zheng. 2024. "An Interactive Generative Design Technology for Appearance Diversity–taking Mouse Design as An Example." *Advanced Engineering Informatics* 59:102263. https://doi.org/10.1016/j.aei.2023.102263.

Zhang, Z., J. Ma, C. Zhou, R. Men, Z. Li, M. Ding, J. Tang, J. Zhou, and H. Yang. 2021. "M6-ufc: Unifying Multi-Modal Controls for Conditional Image Synthesis via Non-Autoregressive Generative Transformers." Preprint arXiv:2105.14211.

Zhao, T., J. Yang, H. Zhang, and K. W. M. Siu. 2021. "Creative Idea Generation Method Based on Deep Learning Technology." *International Journal of Technology and Design Education* 31 (2): 421–440. https://doi.org/10.1007/s10798-019-09556-y.

Zhu, P., R. Abdal, Y. Qin, and P. Wonka. 2020. "Sean: Image Synthesis with Semantic Region-Adaptive Normalization." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 5104–5113.

Zhu, J. Y., T. Park, P. Isola, and A. A. Efros. 2017. "Unpaired Image-to-image Translation Using Cycle-Consistent Adversarial Networks." In *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, 2223–2232.

Zhu, J., Y. Shen, D. Zhao, and B. Zhou. 2020. "In-Domain Gan Inversion for Real Image Editing." In *European Conference on Computer Vision*, Glasgow, UK, 592–608. Springer.