

# DDE-GAN: Integrating a Data-driven Design Evaluator into Generative Adversarial Networks for Desirable and Diverse Concept Generation

**Chenxi Yuan**

Department of Mechanical and  
Industrial Engineering,  
Northeastern University,  
Boston, MA 02115, United States  
email: yuan.chenx@northeastern.edu

**Tucker Marion**

D'Amore-McKim School of Business,  
Department of Industrial Engineering,  
Northeastern University,  
Boston, MA 02115, United States  
email: t.marion@northeastern.edu

**Mohsen Moghaddam\***

Department of Mechanical and Industrial Engineering,  
Khoury College of Computer Sciences,  
Northeastern University,  
Boston, MA 02115, United States  
email: m.moghaddam@northeastern.edu

## ABSTRACT

*Generative Adversarial Networks (GANs) have shown remarkable success in various generative design tasks, from topology optimization to material design, and shape parametrization. However, most generative design approaches based on GANs lack evaluation mechanisms to ensure the generation of diverse samples. In addition, no GAN-based generative design model incorporates user sentiments in the loss function to generate samples with high desirability from the aggregate perspectives of users. Motivated by these knowledge gaps, this paper builds and validates a novel GAN-based generative design model with an offline design evaluation function to generate samples that are not only realistic, but also diverse and desirable. A multimodal Data-driven Design Evaluation (DDE) model is developed to guide the generative process by automatically predicting user sentiment for the generated samples based on large-scale user reviews of previous designs. This paper incorporates DDE into the StyleGAN structure, a state-of-the-art GAN model, to enable data-driven generative processes that are innovative and user-centered. The results of experiments conducted on a large dataset of footwear products demonstrate the effectiveness of the proposed DDE-GAN in generating high-quality, diverse, and desirable concepts.*

**Keywords:** generative design, generative adversarial networks, design evaluation, desirability, design diversity, user-centered design

## 1 Introduction

The generation of innovative, diverse, and user-centered design concepts is an essential phase in the early stages of the product development process and is known to have a significant impact on the quality and success of the design [1–4]. Creating a wide range of solutions that differ significantly from each other can benefit the ideation process of designers and therefore increase the possibility of creating high-quality concepts [5–8]. Various approaches in the literature focus on automatically developing diverse and innovative concepts. The argument is that a large set of concepts promote creativity and logically allows the selection of better ideas from the set [5,9]. However, it is difficult for designers to manually generate a large set of samples with great diversity and novelty because designers naturally tend to fixate on specific design specifications [10–12]. Moreover, most

\*Corresponding author.

existing design problem-solving practices rely heavily on the designers' experiences and preferences. They lack advanced computing methods to help navigate larger solution spaces by generating more diverse, unexpected, and viable solutions [5, 11, 13, 14].

Developing methods to assess and improve creativity has historically been challenging due to its intangible and subjective nature. Significant research in engineering design is currently focused on studying methods and tools to improve the effectiveness and efficiency of creative tasks, such as concept development [4, 15–17]. Creativity is an essential and central part of the ideation process [18]. In human-led design practices, ideation is often an iterative and exploratory process [19], where designers share, modify, and use various stimuli to generate new ideas and concepts [20]. Humans approach this process through various cognitive processes, which research has classified into types and has been shown to affect the effectiveness of ideation [21]. Over the past 25 years, research on computers and Artificial Intelligence (AI) has increasingly focused on how these systems can be used to enhance the creative ideation process [22, 23]. With its ability to synthesize data and make predictions at great speed, the potential for AI to be a generator of new and creative design ideas and concepts has garnered substantial attention from both academia and industry [16, 19].

The methods and frameworks used to apply AI and machine learning in design and engineering are numerous. Deep learning and generative modeling have recently attracted researchers' attention for their potential impact. Recent advances in AI research have made remarkable progress in the machine's ability to generate design ideas [24]. AI can be an inspiration tool in the creative process and a generative tool to assist designers in developing design concepts. AI-powered generative design tools can potentially augment designers' ability to create concepts faster and more efficiently due to their increased speed and efficiency. The power of AI lies in the speed with which it can analyze large amounts of data and suggest design adjustments. The designer can then choose and approve adjustments based on these data.

An emerging research area on using AI to generate novel and realistic design concepts is the use of Generative Adversarial Networks, or GANs [25]. A typical GAN architecture comprises two neural network architectures: a generator and a discriminator. The generator neural network is trained to generate samples (e.g., images) almost identical to real samples. On the other hand, the discriminator neural network learns to differentiate between them. GANs have made significant progress in synthesizing and generating "realistic" images as their central objective. Several successful GAN architectures have recently been proposed, mainly for synthesizing and generating facial images. Examples include CycleGAN [26], StyleGAN [27], Pix2Pix [28], Text2Image [29], and DiscoGAN [30]. These powerful image synthesis models can generate a large number of high-resolution images that are often difficult to distinguish from authentic images without close inspection. Nevertheless, the question remains on leveraging these models in early-stage product design to generate realistic but also novel and diverse concepts. Several technical limitations restrict the ability of GANs to generate diverse and novel designs. These include network architectures, training issues, and a lack of reward mechanisms to generate outputs that satisfy metrics other than realism, such as diversity, novelty, or desirability. Taken together, these represent an impediment to design, where novelty and diversity are critical factors in producing beneficial outcomes [31].

This paper presents a data-driven generative design model that integrates a Data-driven Design Evaluator (DDE) [23] into GANs, called the DDE-GAN model, to improve the performance of GANs through large-scale user feedback on previous designs for diverse and desirable generative design. The main contributions of this paper are as follows:

1. This paper empirically evaluates the potentials and limitations of GANs for generative design. The observations point to the fact that state-of-the-art GAN models and architectures such as StyleGAN [27] are not capable of undertaking generative design tasks due to the lack of mechanisms to ensure diversity and desirability. Empirical evaluation of StyleGAN on a large-scale dataset of footwear products reveals that although the model can generate realistic samples, the generated samples are remarkably similar to authentic products in the training dataset. The results may not benefit designers or promote their creativity, as the samples are neither novel nor aligned with user needs.
2. This paper proposes a novel neural network architecture that integrates a GAN-based model with a multimodal data-driven design evaluation model, or the DDE model for brevity, which was previously developed by the authors [23]. The proposed DDE-GAN model tackles the challenging problem that existing GAN-based generative design solutions lack efficient mechanisms to guide the generator toward generating samples that are not only realistic but also diverse and desirable (i.e., have high expected sentiment scores, both overall and attribute-level) by devising a novel DDE-GAN model enhanced with DDE as a new loss function for automated design evaluation. The DDE-GAN model can predict user sentiments for each attribute of generated samples and generates design concepts with high quality, desirability, and diversity.
3. This paper conducts extensive experiments on a large dataset, scraped from a major online store for apparel and footwear, to demonstrate the effectiveness of the proposed DDE-GAN model in improving the diversity

of generated design samples, as well as their desirability based on predicted user sentiments by comparing it with the StyleGAN model [27] as a baseline. The model is applicable to any other domain, as long as both user data (e.g., reviews, comments) and product data (e.g., images, technical descriptions) are available.

As the majority of cutting-edge generative models are built to create visual designs with great efficacy and success, creating a design concept with descriptive phrases that can automatically convey a novel design concept remains a challenge. This work merely deploys the pre-trained ResNet network of the DDE model [23] to examine and evaluate the visual samples generated. The DDE model, which excludes inputs from the product description, was incorporated into the architecture of the DDE-GAN model presented in Section 3. Future research should focus on building a multimodal DDE-GAN model that couples images and descriptions for automated generation and evaluation of design concepts.

The remainder of this paper is organized as follows. Section 2 provides a detailed overview of related work and topics in traditional and GAN-based generative design. Section 3 provides the details of the proposed DDE-GAN model. Section 4 presents the experimental results, analyses, and performance evaluation. Section 5 provides concluding remarks and directions for future research.

## 2 Related Work

This section provides an introductory overview of GANs and their advantages and limitations for generative design, followed by a review of five main traditional generative design approaches and their comparison with GANs.

### 2.1 GANs for Generative Design: Advantages and Limitations

Deep generative modeling is one of the most promising areas of modern AI studied within the engineering design community to enhance diversity and performance. One way of design exploration is through generative design, which involves programming that alters design geometry parametrically and evaluates the performance of design output versus configurable constraints. The generative model is an architecture that, given a training dataset, can learn its probability distribution and generate new samples with the same statistics as the training data. Among the generative models, GANs [25], offer excellent capabilities and success in generating realistic design images and continue to attract growing interest in the deep learning community. GANs are generative models that involve a minimax game of two players between two models: a discriminative network  $D$  and a generative network  $G$ . The generator aims to learn a generative density function from the training data to produce realistic samples. In contrast, the discriminator attempts to discern whether an input sample is part of the original training set or a synthetic one generated by the generator in such a way as to distinguish fake samples from real ones. GANs have been applied to various domains such as computer vision [32, 33], natural language processing [34], and semantic segmentation [35]. Specifically, GANs have shown significant recent success in the field of computer vision in a variety of tasks such as image generation [36], image-to-image translation [37], and image super-resolution [38].

GANs have been applied to the generation of engineering design, such as the generation of 3D aircraft models in native format for complex simulation [39], numerous wheel design options optimized for engineering performance [40], realistic samples from the distribution of paired fashion clothing and the provision of real samples to pair with arbitrary fashion units for style recommendation [41], and new outfits with precise regions that conform to a description of a language while maintaining the structure of the wearer's body [42]. Most of these models are usually built with quality to ensure high quality and usefulness; however, their intrinsic diversity is limited. The rationale behind the lack of diversity is that, during the training process, the GAN generator is encouraged to generate samples close to the training data distribution to fool the discriminator in a minimax game. GANs illustrate this proposition, as it prompts the generator  $G$  to map an arbitrary noise distribution to realistic samples. On the contrary, the discriminator  $D$  tries to distinguish the generated samples from the real ones, inevitably resulting in limited diversity and creativity. However, due to the property that the generator attempts to learn to mimic the data, GANs are "emulative" [43, 44] and have inspired researchers to investigate areas where diversity and creativity can be promoted in GANs [43].

Extensive research has been conducted to enhance the diversity of GAN-generated image styles [45, 46]. The model can produce diverse outputs by injecting noise vectors, such as the style variation sampled from a normal distribution, into the generator and sampling different style codes [46]. Some studies introduce modes as an additional input to transform conditional input into the target distribution [47]. The predetermined label is fed to the generator. It helps the model produce deterministic outputs that can map different visual domains and styles, which has successfully generated diverse outputs from a given source domain image. It is also observed that the generators are most likely to generate samples from certain major modes/styles in the data but ignore the other modes, for example, the modes that take a small count of distributions. This problem is known as the "mode collapse" and is a primary factor in the lack of diversity in GAN-generated samples. To address this problem,

some researchers propose a regularization term to maximize the distance between the generated outputs and the generated samples with latent codes injected [48]. DivAugGAN [49] further prevents mode collapse and improves the diversity of generated images by using three randomly sampled latent codes and two relative offsets. The model exerts a constraint on the generator to ensure that the changing scale of the generated samples is consistent with the various vectors injected into the latent space. Some researchers [50,51] believe that introducing a regularizer to GAN can address the model collapse problem and thus improve the diversity and quality of generated samples. If GANs are pushed too far from the data distribution for design generation, the quality and realism of the generated samples will be negatively affected. Elgammal [44] proposes modifications to the GAN objective to allow it to generate creative art by maximizing the deviation from established styles while minimizing the deviation from the art distribution. Some researchers [32,52] suggest that only improving diversity will cause GANs to deviate slightly from the original distribution. With this motivation, this paper develops a new GAN architecture that can guarantee high quality and also improve the diversity and desirability of the generated samples.

## 2.2 Traditional Generative Design Methods versus GANs

Various generative design methods have been developed to assist designers in the creative ideation process. Generative design is one of the design exploration methods that can enable simultaneous exploration, validation, and comparison of thousands of design alternatives to support designers and/or automate parts of the design process. There are five commonly used generative design methods, including cellular automata [53], L-systems [54], shape grammars [55], genetic algorithms [56], and swarm intelligence [57]. As a popular generative strategy, Cellular automata, are characterized by the simplicity of its mechanisms on the one hand and the potential complexity of its outcomes on the other. Cellular automata can modify the design specification according to predefined rules and produce unexpected design concepts [58]. Cellular automata, as a popular generative strategy, are characterized by the simplicity of its mechanisms on the one hand and the potential complexity of its outcomes on the other. Cellular automata can modify design specifications according to predefined rules and produce unexpected design concepts [59]. Shape grammars are geometry-based generative systems that describe how complex shapes are built from simple entities and how a complex shape can be decomposed into simpler sub-shapes. Unlike conventional generative design methods that designers often communicate initially, shape grammars involve designers more in making decisions throughout the generative process stage [58]. Genetic algorithms, the most widely used method in generative design exploration, are applied as a generative and search procedure to look for optimized design solutions and has the ability to modify the sequence of the rule of design generation process to assist the designer in generating specific parts of a solution [60,61]. Swarm intelligence is inspired by natural phenomena in which flying or swimming animals move together in packs and allows the system to interact locally with autonomous computational agents to achieve heterogeneous phenomena in generative processes [62]. Despite these generative design methods' significant progress and success, several critical knowledge gaps remain. Most importantly, product forms in these quantitative design methods are typically expressed with a mathematical representation such as vectors, trees, graphs, and grammars, therefore are limited by the trade-off between flexibility and realism [63].

Deep generative models have recently been proposed in the literature to enable more effective and diverse concept generation as an alternative solution for generative design. Specifically, generative adversarial networks (GAN) [25] have shown tremendous success in a variety of generative design tasks, such as topology optimization [40], material design [64], and shape parametrization [39,63]. GANs are composed of a generator trained to generate new samples and a discriminator trained to detect whether the generated samples are real. To better understand how GANs work for generative design purposes, a brief comparison between GANs and the five conventional generative design methods is conducted as follows.

*GANs versus cellular automata.* In conventional cellular automata, generative rules are predefined, usually following much more basic transformations. GANs are composed of many convolutional layers, and cellular automata can be represented using a convolutional neural network with a network-in-network architecture. Therefore, it is noticed that a sufficiently complex neural network architecture, such as GAN, can be used to approximate each rule that fully comprises the cellular automata function. Moreover, the states of neurons in a neural network are continuous, whereas cells in cellular automata have discrete states. In addition, neural networks are primarily concerned with the output and not with the states of individual neurons, whereas the output of cellular automata is a collection of its states.

*GANs versus L-systems.* L-systems is a programmable rewriting paradigm for producing samples. It is challenging to predict the final rendering from the expression of the L-system alone since it is particularly sensitive to changes in expression. The deterministic L-system does not solve the lack of variability for more realistic outputs [65]. However, GANs automatically discover and learn production rules by reading a large dataset. Beyond deterministic restrictions, GANs investigate alternative rules and relationships between characteristics. Because of the powerful processing power of GANs, they are smart enough to comprehensively



learn the distribution of the training samples and reconstruct them. Consequently, GANs can guarantee the quality and realism of the results generated.

*GANs versus shape grammars.* Shape grammars allow for the addition and subtraction of shapes that are eventually perceived as shape modifications. If the shape on the left side matches a shape on a drawing, then the rule can be applied, and the matching shape changes to match the right side of the rule. The generator and discriminator of a GAN model are similar to the left- and right-hand sides of a shape grammar, respectively. The generator sample (equivalent to the left side of a shape grammar) is validated as real by the discriminator (equivalent to the right side of a shape grammar). The generating rule (latent representation learned by GAN) can then be reinforced in the next iteration of the training process, similar to shape grammars.

*GANs versus genetic algorithms.* Genetic algorithms are evolutionary algorithms widely used to explore and optimize the generative design. The adversarial training procedure of GAN can be regarded as an evolutionary process. That is, a discriminator acts as the environment (i.e., provides adaptive loss functions), and a population of generators evolves in response to the feedback from that environment. Genetic algorithms use a form of sampling to measure the relationship between a change in a parameter and a change in the fitness (loss). In contrast, neural networks give a means to directly calculate that relationship without sampling. Therefore, the speedup you experience when training a neural network is the result of not needing to gather as many samples as the number of parameters you wish to tune.

*GANs versus swarm intelligence.* Swarm intelligence involves a collective study of how individuals act in their surrounding environment and interact with each other. It has shown benefits in simplicity, ease of implementation, lack of need for gradient information, and low parameter requirements [46]. Swarm intelligence is the approach that most closely resembles GANs out of the five methods. Although GANs are highly dependent on various parameters and the backpropagation process to alter each layer to affect the loss function, mode collapse is a frequent issue. To prevent mode collapse, swarm intelligence can be employed to improve the generator's performance in GAN and minimize iterations differently from conventional methods [67].

Despite the significant impact and progress made in the literature on GANs, existing work [52,68] is observed to lack sufficient evaluation mechanisms for desirability and diversity that would make GANs suitable for generative conceptual design. The ability of a model to generate concepts with iterative updating from evaluation and feedback has the potential to lead to more creative and valuable design outcomes. The rationale is that the generative process must continually evaluate the generated samples concerning not only realism but also desirability and diversity; otherwise, the number of generated samples with lower desirability or diversity will continue to grow without improvement, making it impossible for designers to consider them meaningfully and accordingly. Some studies in the literature have built GAN-based generative models with such evaluation processes [39, 40, 69]; however, their proposed evaluation tools are exclusively based on physics-based virtual simulation environments that do not necessarily reflect user feedback. To bridge this gap, a user-guided evaluation DDE-GAN model is proposed to enhance the generated design's quality, diversity, and desirability by incorporating synthetic user feedback from an evaluation process for its generated intermediate samples.

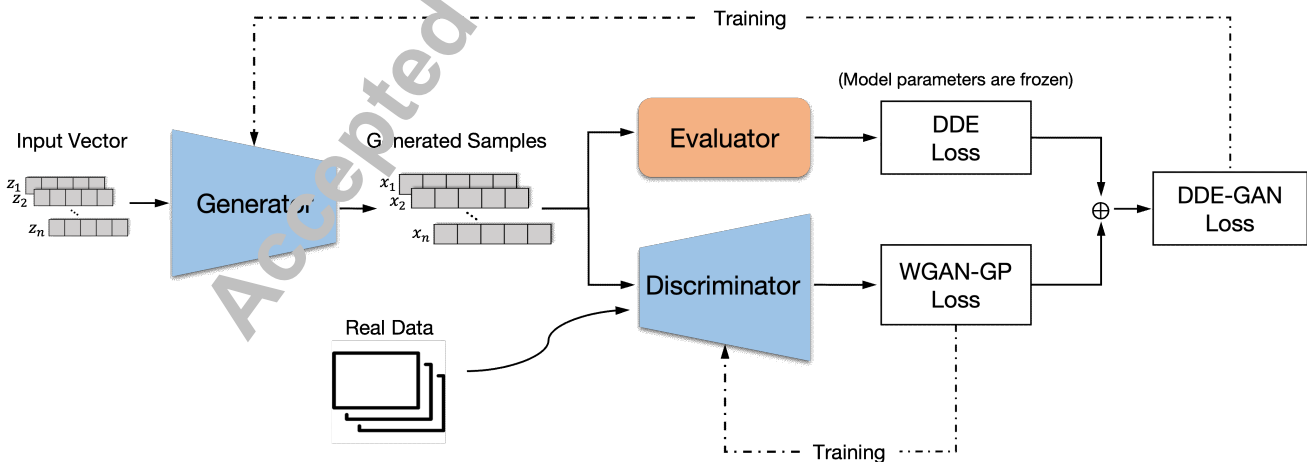


Fig. 1. Architecture of the integrated Automated Design Evaluator-Generative Adversarial Network (DDE-GAN) model. WGAN-GP: Wasserstein GAN + Gradient Penalty [70]; DDE: Data-driven Design Evaluation [23].

### 3 Methodology

This section presents the architecture and formulation of the proposed DDE-GAN model. A schematic of the model architecture is shown in Figure 1. A brief overview of the GAN formulation and StyleGAN is presented first (Section 3.1), followed by the proposed loss function based on the DDE model (Section 3.2). StyleGAN [71] is applied as a baseline in this work, and the novel loss function of the DDE-GAN model is improved over the loss function of StyleGAN. The proposed DDE-GAN is described, followed by details of the DDE model, previously developed by the authors [23], which is used as a newly added loss function in the developed model. The DDE model [23] accurately predicts the overall and attribute-level desirability of a new concept based on large-scale user sentiments and feedback on past designs. This work applies a well-trained DDE model as an augmented discriminator to promote user-centered image generation using the StyleGAN model, to generate realistic, diverse, and desirable samples.

#### 3.1 GAN Formulation

GANs [25] can generate images from random noise and do not require detailed information or labels from existing samples to start the generative process. The standard GAN structure consists of two neural networks: a generator  $G$  and a discriminator  $D$ . The generator  $G$  takes random noise  $z \sim P(z)$  sampled from a uniform or normal distribution as input and maps the noise variable  $z \sim P(z)$  to the data space  $x = G(z)$ . The discriminator  $D$  distinguishes whether an image is real or fake (i.e., made by the generator). The output  $D(x)$  is the probability that the input  $x$  is real. If the input is a fake image,  $D(x)$  would be zero. Through this process, the discriminator  $D$  is trained to maximize the probability of assigning the correct label to both real samples and fake samples. Generator  $G$  is encouraged simultaneously to fit the true data distribution. The adversarial training processes that update the parameters of both networks through backpropagation are formulated as the following learning objective:

$$\min_G \max_D \mathbb{E}_{x \sim \mathbb{P}_r} [\log(D(x))] + \mathbb{E}_{\tilde{x} \sim \mathbb{P}_g} [\log(1 - D(\tilde{x}))], \quad (1)$$

where  $\mathbb{P}_r$  is the distribution of the real image  $x$  and  $\mathbb{P}_g$  is the model distribution implicitly defined by  $\tilde{x} \sim G(z)$ ,  $z \sim P(z)$ . The generator's input,  $z$ , is sampled from a simple noise distribution  $P$ , such as the uniform distribution or a spherical Gaussian distribution.

GAN models come in various forms, StyleGAN [21, 71] being one of them. The StyleGAN extension to the GAN architecture proposes a major modification to the generator model, which uses (1) a mapping network to transfer latent space points to an intermediate latent space, (2) an intermediate latent space to regulate style at each point in the generator model, and (3) an addition of noise as a source of variation at each point in the generator model. It features a brand-new style-based generator architecture that creates high-resolution images with cutting-edge visual quality. In addition to generating stunningly realistic high-quality images, the model also provides control over the style of the generated image at various degrees of detail by adjusting the style vectors and noise. Moreover, most GAN-variants are generally sensitive to the problem domain. They perform exceptionally well in processing generative tasks using large popular datasets such as human faces and animals, where the novelty, diversity, or desirability of the generated samples are not important. However, their limitation becomes evident in generative design tasks where the quality, diversity, and desirability of samples must be optimized simultaneously [41]. Inspired by the success and popularity of StyleGAN, this paper develops a novel GAN model based on StyleGAN's architecture, enhanced with the DDE model [23]. The loss function utilized in StyleGAN is WGAN-GP [70], which is the most widely used loss function. WGAN-GP is constructed with the Wasserstein GAN [72] formulation along with a gradient norm penalty to achieve Lipschitz continuity. The Wasserstein loss formulation applies the Wasserstein-1 distance using a value function based on Kantorovich-Rubinstein duality [73]. The loss function (1) is then modified as follows:

$$\min_G \max_{D \in \mathcal{D}} \mathbb{E}_{x \sim \mathbb{P}_r} [D(x)] - \mathbb{E}_{\tilde{x} \sim \mathbb{P}_g} [D(\tilde{x})], \quad (2)$$

where  $\mathcal{D}$  is the set of 1-Lipschitz functions and  $\mathbb{P}_g$  is the distribution of the model implicitly defined by  $\tilde{x} = G(z)$ ,  $z \sim p(z)$ . The Wasserstein loss is approximated given a set of  $k$ -Lipschitz, and the weights of the discriminator are clipped to some range. By adding the Wasserstein loss with gradient penalty, WGAN-GP enforces a soft restriction on the gradient norm of the discriminator's output with respect to its input rather than clipping network weights

as in [72], to guarantee the Lipschitz requirement. The objective function of StyleGAN is then formulated as follows:

$$L = \mathbb{E}_{\tilde{\mathbf{x}} \sim \mathbb{P}_g} [D(\tilde{\mathbf{x}})] - \mathbb{E}_{\mathbf{x} \sim \mathbb{P}_r} [D(\mathbf{x})] + \lambda_{GP} \mathbb{E}_{\hat{\mathbf{x}} \sim \mathbb{P}_{\hat{\mathbf{x}}}} \left[ (\|\nabla_{\hat{\mathbf{x}}} D(\hat{\mathbf{x}})\|_2 - 1)^2 \right], \quad (3)$$

where  $\hat{\mathbf{x}}$  is a random sample,  $\mathbb{P}_{\hat{\mathbf{x}}}$  is defined as samples along straight lines between pairs of points that come from the true data distribution and the generator distribution, and  $\lambda_{GP}$  is a weighing factor.

### 3.2 Data-driven Design Evaluator Loss

The preliminary experiments conducted by the authors to generate images of footwear products using StyleGAN revealed that although the model is capable of generating realistic samples, the generated samples are remarkably similar to the real products in the training dataset. These similarities can even be detected by simple visual inspections (see Figure 2). With a sufficiently trained generator, even the discriminator would be unable to distinguish between the generated samples and the real ones. The second finding is that, although the generated images are realistic, they may not benefit designers or promote their creativity as the samples are not necessarily novel or aligned with user needs. Although the model training procedure considers algorithmic quality, it does not consider how users will receive and react to these computer-generated designs. This paper argues that this problem comes from the sole objective of existing generator-discriminator architectures to maximize “realism”. That is, there is an absence of a loss function that can incorporate other critical metrics in addition to realism, such as the alignment of the generated samples with the perspectives and needs of users, which could cause the discriminator to fail when updating the generator in terms of learning and producing features that maximize the usefulness of a design. To convey the measurement of the design performance score back to the generator for subsequent iteration improvements, the authors believe that new loss functions are needed to force the discriminator to identify and



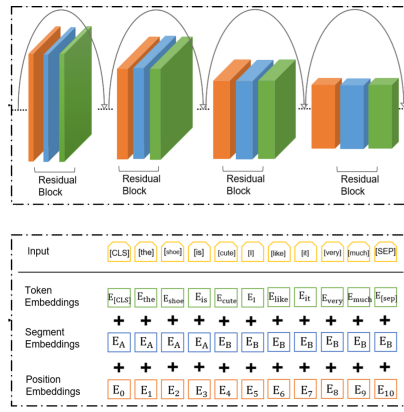
Fig. 2. Examples of sneaker images generated using StyleGAN.

# Design Concept Renderings

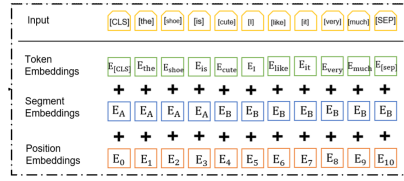
## Design Concept Description

- Mesh upper is lightweight and breathable
- Rigid heel counter for stability
- Padded tongue and collar for cushioned ankle support
- Bounce midsole for stability and energy return
- Rubber traction outsole for grip on the court

## Image Model (ResNet-50)



## Text Model (BERT)



## Predicted Performance

Attributes	Rating
Traction	-0.07
Shape	0.58
Heel	0.37
Cushion	0.49
Color	0.12
Fit	0.78
Absorption	-0.21
Durability	-0.47
Permeability	0.13
Stability	0.48
Overall	0.65

Fig. 3. The multimodal Data-driven Design Evaluation (DDE) model (adapted from [23]).

locate other metrics, such as novelty or desirability. This observation inspired the authors to investigate when to incorporate the user-guided assessment mechanism into the discriminator, as described below, if the similarity between the produced and real images is effectively reduced.

This paper applies DDE [23] as a user-centered design evaluation model to evaluate the generated samples with respect to the expected quality and performance of the generated designs. DDE is a multimodal deep regression model that uses an attribute-level sentiment analyzer [23] to predict user-generated product ratings based on online reviews. It was created to automate design evaluation and improve decision-making by domain experts. Based on extensive user evaluations of existing designs, the DDE model offers designers a precise and scalable means to forecast new concepts' overall and attribute level desirability. DDE is an end-to-end design assessment system that can interpret visuals, plain language, and structured data. As shown in Figure 3, the DDE system uses a ResNet-50 model [75] to evaluate and interpret images of a product. ResNet-50 can represent complex functionality and learn features at many different levels of abstraction to understand the connections between orthographic representations of design concepts (inputs) and user sentiment intensity values (outputs). The Bidirectional Encoder Representation from Transformers (BERT) model, a different model in the DDE system, extracts and analyzes product descriptions written in natural language [76]. The BERT model can determine the connection between a product's technical description and the user's emotional sentiment level. The DDE system then integrates the various meaningful data collected from the Internet platform and models the relationships between images, text, and statistics. The DDE model synthesizes different modes of data using a novel fusion mechanism to develop a more accurate context about the product and the associated user feedback [23]. The DDE model was trained on a large-scale dataset that was scraped from a major online footwear store. In the dataset, each product has four types of information: six orthographic images, one numerical rating score, a list of textual product descriptions, and real textual customer reviews from an e-commerce platform, where images and feature descriptions are the inputs to the DDE model and the numerical rating score and sentiment intensity values from customer reviews are the outputs. The dataset is constituted of a total number of 8,706 images and 113,391 reviews for 1,452 identified shoes. Numerical experiments on this large dataset indicated promising performance by the DDE model with 0.007 MSE loss and over 99.1% accuracy.

The DDE model can accurately predict user sentiments for a new design concept based only on its orthographic images and description and provide numerical design performance values associated with each attribute of the generated concept. This paper builds a novel loss function based on the DDE model, called the DDE loss, into the GAN's discriminator to enable an accurate and scalable prediction of the new concepts' overall desirability. By integrating the DDE loss into the StyleGAN's discriminator, the DDE-GAN model is created (Figure 1). The DDE loss integrated into the discriminator can measure the intermediate samples generated by the generator in each iteration and convey the loss back to the generator for a new set of parameters. The DDE loss evaluates the results of each round in the iterative training process, which is then used to backpropagate and optimize the generator and the discriminator. The DDE-GAN architecture is expected to result in better designs from the user's point of view and simultaneously maintain excellent image quality.

The objective function of the DDE-GAN model is therefore formulated as follows:

$$L = \mathbb{E}_{\tilde{\mathbf{x}} \sim \mathbb{P}_g} [D(\tilde{\mathbf{x}})] - \mathbb{E}_{\mathbf{x} \sim \mathbb{P}_r} [D(\mathbf{x})] + \lambda_{GP} \mathbb{E}_{\tilde{\mathbf{x}} \sim \mathbb{P}_{\tilde{\mathbf{x}}}} \left[ (\|\nabla_{\tilde{\mathbf{x}}} D(\tilde{\mathbf{x}})\|_2 - 1)^2 \right] + \lambda_{DE} \mathbb{E}_{\tilde{\mathbf{x}} \sim \mathbb{P}_g} [\mathcal{L}_{DE}(\tilde{\mathbf{x}})] \quad (4)$$

$$\mathcal{L}_{DE}(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N (f_i(\mathbf{x}) - \hat{\mathbf{y}}_i)^2, \quad (5)$$

where  $\lambda_{DE}$  is a constant that defines the loss weight,  $\mathcal{L}_{DE}$  is the DDE loss function to evaluate the feature of the generated samples in the characteristic of performance,  $f_i(\mathbf{x})$  is the prediction for all design  $\mathbf{x}$  generated by the DDE model, and  $\hat{\mathbf{y}}_i$  is the desired design evaluation score and is set as 1 for each attribute, indicating that the models are trained to generate samples with the highest possible expectation. The StyleGAN loss terms regulate the high quality of the produced pictures and the DDE loss guarantees that the produced samples have high user sentiment scores. Combining the two elements allows the proposed DDE-GAN model to simultaneously create high-quality images and high user sentiment ratings. This new set of loss functions provides a more accurate and evaluation-guided generator and discriminator in DDE-GAN compared to previous work and can be easily tuned. Information on the constants and other implementation details is provided in Section 4.

## 4 Experiments and Results

In this section, the dataset and implementation details of the proposed DDE-GAN model are first described, followed by the introduction of metrics established to investigate the effectiveness of the developed DDE-GAN model in generating realistic samples with high desirability and diversity. The results of the experimental analyses are presented next, comparing the outcomes generated by the developed DDE-GAN model and the state-of-the-art StyleGAN model as a baseline.

### 4.1 Dataset and Implementation Details

To test and validate the performance of StyleGAN in generating realistic and diverse images, a large-scale dataset was scraped from a major online footwear store to perform numerical experiments. The collected large-scale dataset contains a total of 7,642 images with a size of  $256 \times 256 \times 3$ . Several brands of footwear are included in the dataset to avoid mode collapse and increase the diversity of the dataset, including Adidas, ASICS, Converse, Crocs, Champion, FILA, PUMA, Lacoste, New Balance, Nike, and Reebok.

The DDE model is pre-trained and serves as an offline network added to the new StyleGAN loss. The implementation of the pre-trained DDE model is discussed in this section. The experiments were carried out with  $k$  fold, with  $k = 10$ , to randomly split the dataset into train, validate, and test sets with a 7 : 1 : 2 ratio. All experimental results were conducted five times and reported as mean  $\pm$  std to alleviate the randomness effect. All neural networks were trained on PyTorch [77]. Adam [78] optimizer with  $\beta = (0.9, 0.999)$  and the learning rate of  $= 0.01$  were used to train the model parameters for 50 epochs and save the model with the best loss in the validation dataset. To avoid overfitting, a dropout layer was added to the self-attention fusion model with a dropout rate of  $P_{drop} = 0.1$ . The DDE model was trained over 40 epochs. The training time cost per epoch was 5-7 minutes, which added up to 3-4 hours. All training and testing experiments were conducted on a single NVIDIA RTX 3090 GPU (24GB VRAM), an AMD Ryzen 9 5950X CPU, and 64GB memory.

The weight of StyleGAN,  $\lambda_{GP}$ , was precisely calibrated by the original paper with the best performance achieved [27]. Therefore, this paper follows the exact same value of 0.8192 as suggested in the original StyleGAN paper. The weight of DDE loss,  $\lambda_{DE}$ , is defined by binary search from 0.1 to 2 in general, and finally is set as 0.5 to meet the trade-off between high image quality (FID) and predicted sentiment scores. Adam [78] was used as an optimizer with a learning rate of 0.0025 to optimize the model and set  $\beta$  as  $(0.9, 0.999)$ , representing the coefficients used for computing running averages of gradient and its square. Beyond that, data augmentation methods such as random flip, rotation, scale, brightness, and contrast were applied to improve data diversity. The model was trained 20,000 times for each experimental setting, and the average performance statistics were reported.

### 4.2 Evaluation Metrics

**Frechet Inception Distance** FID [79] is used to assess the quality of the images created by a generative model. FID evaluates the statistics for both the target and output images simultaneously. It compares the distribution



of the generated images with the distribution of real photos used to train the generator. FID can also identify intraclass mode dropping and quantify the variety and quality of produced samples, making it a valuable tool for assessing the quality and diversity of synthetic images. A lower FID score intuitively indicates a closer distribution between the objectives and the results, which corresponds to a better performance of the generative model. The proposed DDE-GAN model is compared with the state-of-the-art GAN architecture, StyleGAN, in terms of FID, using the following equation [79]:

$$\text{FID} = \|\mu_r - \mu_g\|^2 + \text{Tr}(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{1/2}), \quad (6)$$

where  $\text{Tr}$  refers to the trace linear algebra operation,  $(\mu_r, \Sigma_r)$  and  $(\mu_g, \Sigma_g)$  refer to the mean and covariance matrices in the feature order of the embeddings obtained in real and generated images, respectively.

**Diversity Assessment** To quantitatively measure whether the DDE-GAN model has guided the generator to synthesize new designs with greater variety, a kernel-based statistical test method called Maximum Mean Discrepancy (MMD) is used to determine the similarity between two distributions [80, 81]. MMD is defined by the idea of representing the distances between the distributions as the distances between the mean embeddings of the features. Given two sets of data  $X$  and  $Y$ , the MMD is calculated as the distance between the feature means of  $X$  and  $Y$ . The expression is formulated as follows:

$$\text{MMD}^2(P, Q) = E_P[k(X, X)] - 2E_{P, Q}[k(X, Y)] + E_Q[k(Y, Y)], \quad (7)$$

where  $k$  is the kernel function,  $P$  is the distribution over a set of input data  $X$ , and  $Q$  is the distribution over a set of generated data  $Y$ . This paper uses two different kernel functions: a linear kernel and a polynomial kernel. The linear kernel is defined as:

$$k(x, y) = x^\top y, \quad (8)$$

and the polynomial kernel is defined as:

$$k(x, y) = (\gamma x^\top y + c_0)^d, \quad (9)$$

where  $x$  and  $y$  are the input vectors,  $d$  is the degree of the kernel,  $\gamma$  is the weight, and  $c_0$  is a constant. In the experiments, the polynomial kernel  $\gamma$  is set to 1, the kernel degree  $d$  is set to 0.5 and the coefficient  $c_0$  is set to 0. Details of the results are discussed in the following sections.

### 4.3 Results and Analyses

To test and validate the performance of the proposed DDE-GAN model for design generation with improved desirability and diversity, a set of experiments was performed on a real dataset of footwear products with StyleGAN as the baseline model. This section first presents the visual design samples generated by the DDE-GAN models. The performance of the proposed model is then compared with the baseline using the FID score, followed by an MMD analysis to examine the similarity between the generated images and the real images. Lastly, DDE is applied to test the images generated by the DDE-GAN and StyleGAN models to evaluate their desirability prediction scores.

**Visual Results** As shown in Figure 4, the DDE-GAN generated samples deliver the expected high quality and realism, which are also observed in the StyleGAN generated samples (Figure 2). The overall images are realistic, vibrant, clear, and have an aesthetic understandable to the human mind. Although Figure 4 reveals some differences, the uniqueness and diversity of the images are discovered in some samples. Some images contain features that might sound novel or even strange. However, this paper defines these characteristics as novelty and diversity. The authors noticed that most of the generative model samples generated in the current GAN-based design literature emphasize quality, while the images are somehow similar to existing products. Yet, that may

hinder innovation in the generative process, because a conventional GAN discriminator may easily label a “novel” sample that could potentially be an interesting sample from the design perspective as “fake”, simply because it does not look like any real item within that category and contains unknown features. This, in turn, would discourage the conventional GAN generator from generating more of these potentially novel samples. The DDE-GAN model proposed in this paper introduces an additional loss to encourage the generator to produce more novel and distinctive images. Therefore, the authors define attributes such as “strange” or “never seen before” as one of the diversity criteria. Among the large size of the generated samples, 16 distinct images are manually selected and presented in Figure 5, as they are identified as designs with novelty and diversity. It is clearly seen that these sneakers are far from “similar” to existing sneakers, compared to the other samples shown in Figure 4. They look distinguishable with more novelty and diversity. To further validate the effects of the DDE-GAN model on novelty

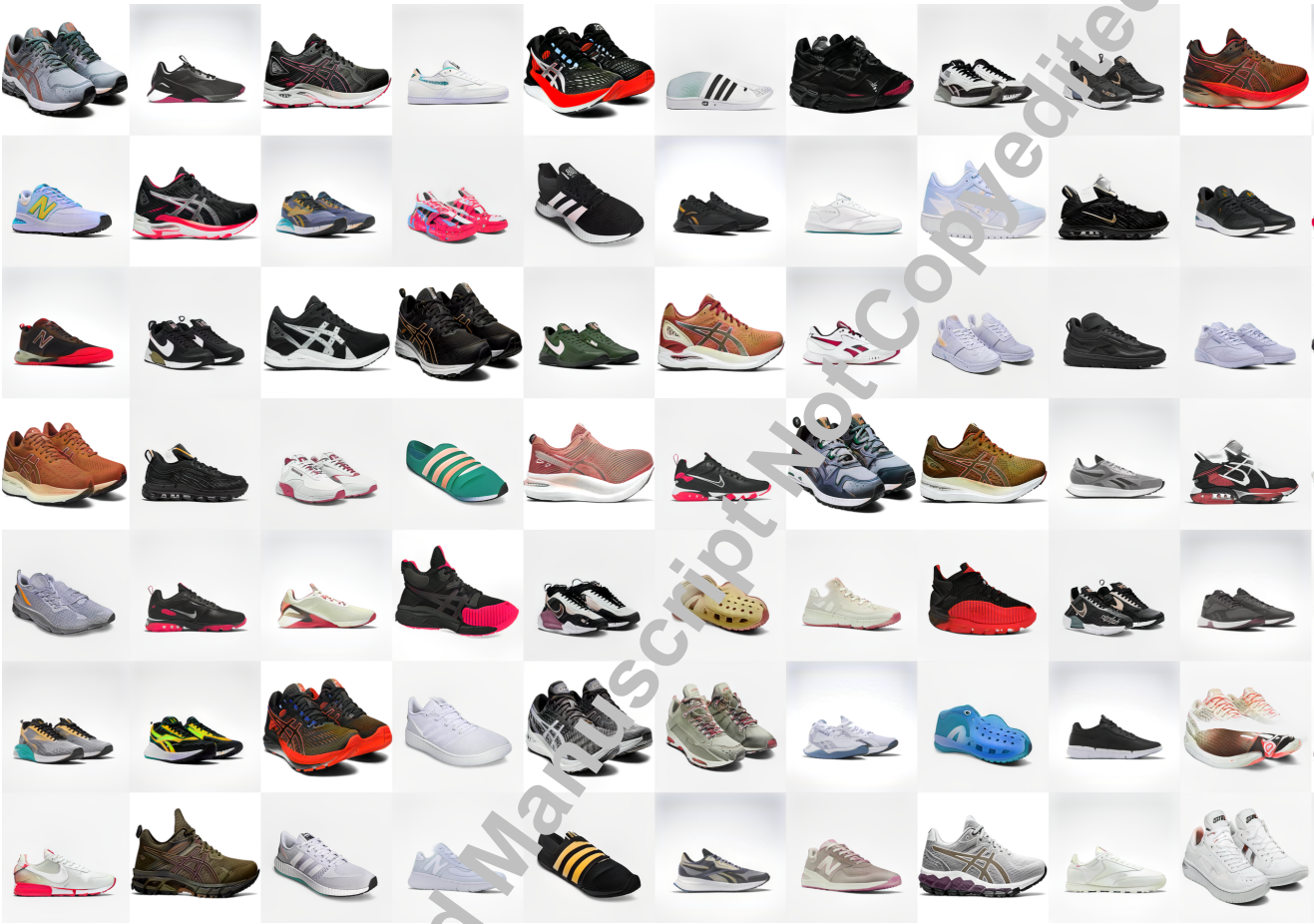


Fig. 4. Examples of novel designs generated by the proposed DDE-GAN model.

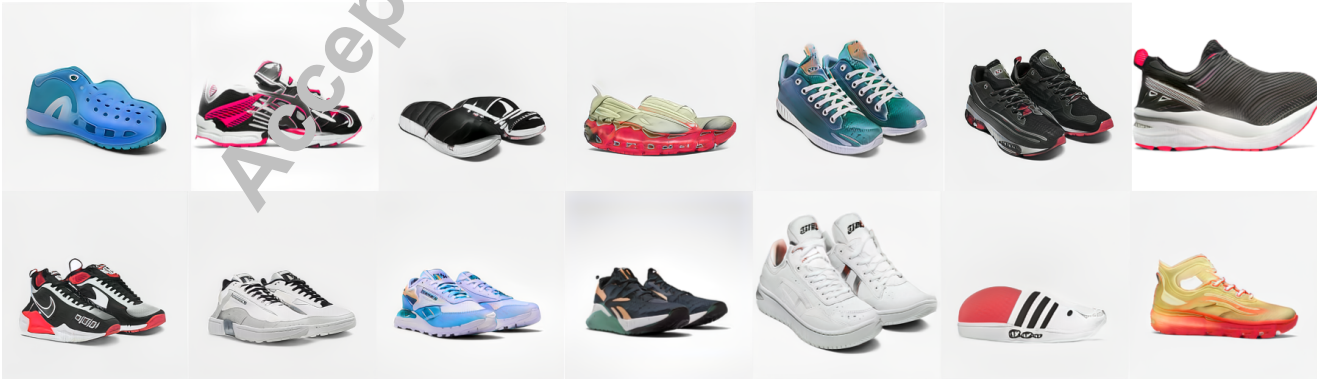


Fig. 5. Examples of novel designs with high diversity observed.

Algorithm	FID Score
StyleGAN	$6.22 \pm 0.17$
DDE-GAN	$6.45 \pm 0.21$

and diversity, a quantitative analysis of FID and MMD is conducted next.

**Quality Test** Table 1 shows the FID scores of the best samples for each generation when training models with the collected dataset. The FID scores are the mean values for ten different training results. As shown in Table 1, the StyleGAN model produces a lower FID score than the DDE-GAN model. A lower FID score means that the model is more stable and correlates better with higher-quality images. However, the FID score of the DDE-GAN model and its standard deviation are close to StyleGAN with only a small change (a 0.23 decrease), and it is empirically concluded that an FID score below 10 is sufficient to demonstrate the effectiveness of a generative model [82, 83]. In addition, the difference between DDE-GAN (mean = 6.45) and StyleGAN (mean = 6.22) is verified with t-test,  $P = 0.0026$ . Therefore, the DDE-GAN model performs well in achieving high-quality results. FID can also be explained as a similarity metric, because it calculates the distance between the feature vectors calculated for the real and generated images. Lower scores indicate the two groups of images are more similar, or have more similar statistics, with a perfect score being 0.0 indicating that the two groups of images are identical. Therefore, from the perspective of similarity, StyleGAN with lower FID represents that the generated samples are more similar to real images compared with the DDE-GAN with a higher FID score. DDE-GAN with higher FID reveals that the generated samples are distinct from existing images, which is further validated next.

**Diversity Test** The primary rationale behind the proposed DDE-GAN model is to promote the diversity of images generated by GAN. The similarity between the produced samples and the original input is calculated using the MMD metric to estimate the diversity of novel samples. A higher similarity value indicates that the generated samples contain less diversity, and vice versa. The MMD (Maximum Mean Discrepancy) values are calculated based on the results of the proposed model and the baseline model, using linear and polynomial kernels, as shown in Figure 6. The proposed DDE-GAN model is observed to produce higher MMD scores than the baseline StyleGAN model, indicating a significantly lower similarity between the real training dataset and the samples generated by the DDE-GAN model. For the linear kernel, StyleGAN receives a mean of 124.77 and a standard deviation of 3.74, and DDE-GAN obtains a lower mean of 110.15 with a lower standard deviation of 5.02. The mean and variance of the polynomial kernel are (0.145, 0.003) and (0.164, 0.002) for DDE-GAN and StyleGAN, respectively. A statistical test was employed to examine the difference between the performances of the two models. Results of assessments using linear kernel MMD show that the DDE-GAN model (mean =

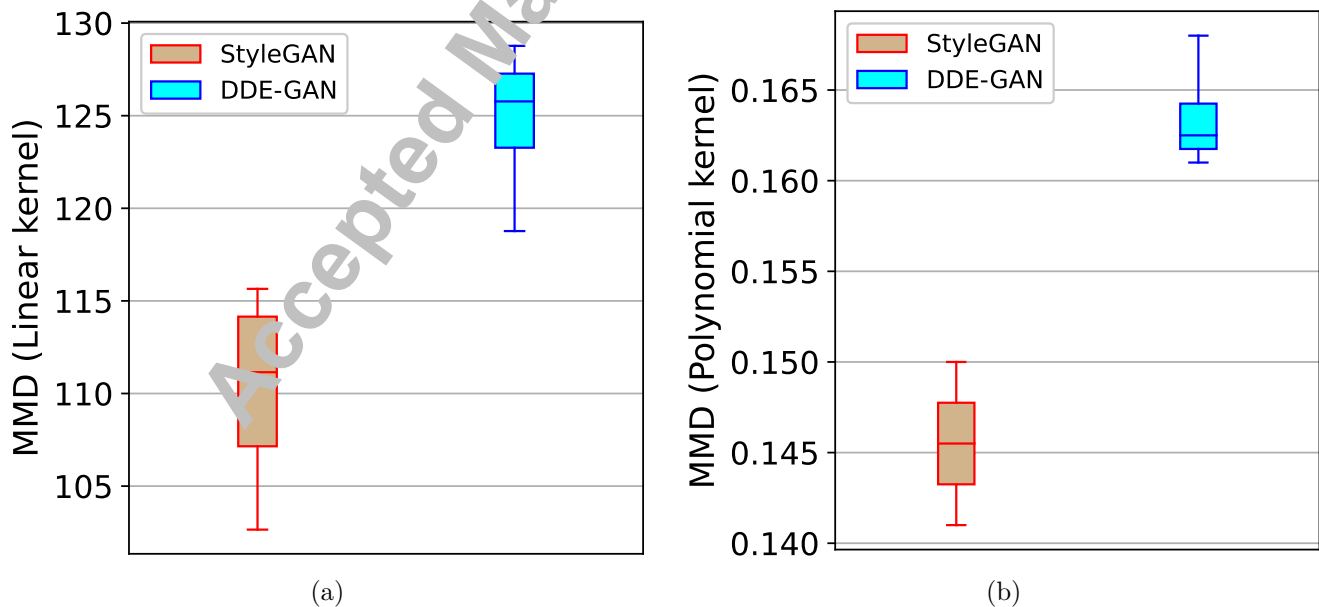


Fig. 6. The MMD linear kernel (a) and MMD polynomial kernel (b) results for StyleGAN and DDE-GAN.



110.15; t test  $P = 3e - 08$ ) significantly outperforms StyleGAN (mean = 124.77) in generating samples with high diversity. Likewise, StyleGAN, assessed using polynomial kernel MMD, is shown to perform worse in generating diverse samples (mean = 0.144) compared to DDE-GAN (mean = 0.163; t test  $P = 4e - 14$ ). Overall, DDE-GAN performs well in generating images with less similarity to the original dataset and more diversity.

**Desirability Test** In addition to enhancing the diversity and novelty of the generated images, another objective of this paper is to build a user-guided automated generative design model that can produce designs that meet the desirability requirements. The DDE model was trained on a large dataset of 1,452 design images labeled with user sentiments to learn and capture the relationship between images and attribute performance. The DDE model creates a collection of rating scores representing the performance of relative attributes when images of testing products are imported into the model. The number of products with an absolute value of the desired values and the predicted values below a threshold was counted using the Prediction Accuracy Rate (PAR) metric. The percentage of the counted number to the overall testing number serves as the accuracy metric. The well-trained DDE model was verified to predict user sentiments for a new design concept based only on its orthographic images and provides the numerical values of the design performance associated with each product attribute with a prediction accuracy of 76.54% [23]. To test whether the new designs produced by DDE-GAN perform better than the designs created by StyleGAN, all 480 images were selected from the output of two models and tested using a well-trained DDE model to predict their overall and attribute-level desirability based on large-scale user reviews on existing products. The average numerical values of user sentiments in 10 attributes and the overall performance of the designs are shown in Table 2 in which the sentiment intensity of users ranges from  $[-1, 1]$  with  $-1$  and  $1$  representing extremely negative and extremely positive sentiment, respectively. DDE-GAN is observed to generate designs with higher expected user sentiment values for most individual attributes and overall performance. In general, the predicted sentiment values of individual attributes of the samples created by DDE-GAN obtained increases of 9%-56% compared to StyleGAN, except for the attribute "Fit". To further explore the differences between the two models, the predicted sentiment values of the two models are analyzed by two-tailed independent samples t-tests, tested for significance at  $P < .05$ . As shown in Table 2, there is a significant improvement associated with the attributes "Traction", "Shape", "Heel", "Cushion", "Color", "Impact absorption", "Permeability", "Stability" and the "Overall" rating of which p-values are much less than 0.05. However, The prediction performance of the two models was not significantly different for the attributes "Fit" and "Durability" ( p-value 0.0803 and 0.0334, respectively). The potential reason is that these two features are not easy to be captured by generative models which is more capable of learning the latent representation among visual features. This is an interesting and open issue for generative design tasks and authors are inspired to tackle the problem in future work. The statistical

Table 2. Results of the DDE test [23] regarding "predicted sentiment values" on 480 randomly selected samples generated by StyleGAN and DDE-GAN.

Attributes	Predicted Sentiment Value		Change(%)	P-value
	StyleGAN	DDE-GAN		
Traction	$0.1652 \pm 0.018$	$0.2064 \pm 0.018$	25%	0.0035
Shape	$0.2831 \pm 0.016$	$0.3097 \pm 0.024$	9%	0.0074
Heel	$0.3736 \pm 0.020$	$0.5142 \pm 0.015$	38%	<0.0001
Cushion	$0.1924 \pm 0.019$	$0.3005 \pm 0.031$	56%	<0.0001
Color	$0.2783 \pm 0.021$	$0.4179 \pm 0.019$	50%	<0.0001
Fit	$0.2350 \pm 0.015$	$0.2168 \pm 0.012$	-8%	0.0803
Impact absorption	$0.2303 \pm 0.027$	$0.3211 \pm 0.016$	39%	<0.0001
Durability	$0.2409 \pm 0.039$	$0.2714 \pm 0.034$	13%	0.0334
Permeability	$0.1471 \pm 0.020$	$0.1916 \pm 0.017$	30%	<0.0001
Stability	$0.1892 \pm 0.031$	$0.2073 \pm 0.025$	10%	<0.0001
<b>Overall</b>	$4.536 \pm 0.0754$	$4.735 \pm 0.0718$	4%	0.0002

test strongly verifies that the additional loss function in conjunction with the discriminator successfully improves the generator to learn features corresponding to user sentiments and design desirability. The evaluation results indicate that DDE-GAN-generated design samples will lead to greater user satisfaction compared to StyleGAN-generated samples that focus only on the “realism” of the generated samples. The proposed DDE-GAN model is optimized to serve effectively as a user-centered generative design framework.

## 5 Conclusions and Future Research Directions

This paper takes a different approach to promote diversity and desirability in GAN-based generative design models. The lack of these critical design metrics in the samples generated by existing GANs is caused by the limitation of adversarial training between the generator and the discriminator to generate only “realistic” samples. To address this problem, a multimodal data-driven design evaluation model, DDE, is introduced in the discriminator to encourage the generator to get creative and generate more “unfamiliar” and potentially novel samples. Another problem this paper addresses is devising a user-centered generative model that can generate real products with high usefulness and attractiveness from the user’s perspective. To bridge this gap, the DDE model is applied to predict the performance of the generator samples in each iteration. The predicted values are integrated with other loss functions and transmitted to the models for backpropagation. The generator is updated and optimized for integrated DDE loss and finally is enforced with the capability to generate well-performed designs. To investigate the effectiveness of the developed DDE-GAN model in generating images with high quality, high diversity, and desirability, the FID metric, the MMD, and the DDE testing tool are deployed to conduct the DDE-GAN experiment analysis with the baseline StyleGAN model. Visual output and quantitative analysis validate the improvement of DDE-GAN. Specifically, the generated images contain novel features and characteristics from human observation and further quantitative analysis. Average FID scores confirmed the stability of the newly devised DDE-GAN and stated the sufficient ability of DDE-GAN to generate high-quality images. Lower MMD values again indicate that the DDE-GAN enhances the generator’s ability to create more diverse samples. The DDE offline model was applied to test the two sets of novel images of DDE-GAN and StyleGAN, and DDE-GAN has demonstrated the ability to design samples with improved desirability and popularity. The developed DDE-GAN model was successfully tested in a sneaker design case study, but is flexible enough to be readily expanded to other product categories and can serve as an intelligent tool to produce photorealistic renderings of new concepts in other design applications.

This work starts the journey of generative model integrated with user data to build the foundation for data-driven, user-centered design. Potential AI-augmented design tools can range from user-centered design evaluation, design generation, design selection, to design recommendation. In future iterations, because DDE-GAN integrates user sentiment, the influence of extreme users on design novelty will be explored. Integrating extreme user behaviors, needs, and sentiment has been shown to increase design creativity and novelty [84]. What’s more, the DDE-GAN model is developed to aggregate user feedback in the loss function to generate samples with high desirability from the perspective of users, which is a limitation that the produced sample conveys most of the user feedback. Therefore, future work can build a model that can generate design concepts based on the individual user. Furthermore, another potential for enhanced user-guided GANs would be the ability to simultaneously generate images and textual languages. The DDE model was devised to extract visual and textual features and identify the dependency among various data types, such as image, text, and structure data. This work only partially used the image evaluation tools in DDE to inspire the generator to create enhanced and guided samples. Therefore, in the next step, the authors will continue to develop a multimodal GAN that can generate a visual image and natural language as a detailed description of design samples. Additionally, the DDE model will be efficiently used to assess samples with images and text information for a more accurate generative model. There is the potential to broaden model’s usefulness to other facets of the innovation process, including design, marketing and product management. Moreover, future research will explore the increasing inclusion and reduction of bias in the model, as this represents a significant issue in many AI applications [85]. Lastly, future research should conduct semi-structured interviews and post-evaluation activities involving design experts to evaluate the results of generative design, both objectively and subjectively. Further qualitative validation can evaluate how many ideas designers can generate, the novelty of ideas in terms of how innovative and unexpected they are, how well the design space is explored (variety), and how feasible they are when comparing them with design specifications (quality). The relevant attributes of the product in each dimension must be identified and evaluated according to the recommendations and considerations of Shah [31] and Nelson [86].

## Acknowledgments

This material is based upon work supported by the National Science Foundation under the Engineering Design and System Engineering (EDSE) grant #2050052. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

## References

- [1] Liu, Y.-C., Chakrabarti, A., and Bligh, T., 2003. "Towards an 'ideal' approach for concept generation". *Design studies*, **24**(4), pp. 341–355.
- [2] Yilmaz, S., Daly, S. R., Seifert, C. M., and Gonzalez, R., 2015. "How do designers generate new ideas? design heuristics across two disciplines". *Design Science*, **1**.
- [3] Simon, H. A., 2019. *The sciences of the artificial*. MIT press.
- [4] Cross, N., 2021. *Engineering design methods: strategies for product design*. John Wiley & Sons.
- [5] Osborn, A. F., 1953. "Applied imagination".
- [6] Akin, Ö., and Lin, C., 1995. "Design protocol data and novel design decisions". *Design Studies*, **16**(2), pp. 211–236.
- [7] Brophy, D. R., 2001. "Comparing the attributes, activities, and performance of 'divergent, convergent, and combination thinkers". *Creativity research journal*, **13**(3-4), pp. 439–455.
- [8] Atman, C. J., Chimka, J. R., Bursic, K. M., and Nachtmann, H. L., 1991. "A comparison of freshman and senior engineering design processes". *Design studies*, **20**(2), pp. 131–152.
- [9] Christiaans, H., and Dorst, K. H., 1992. "Cognitive models in industrial design engineering: a protocol study". *Design theory and methodology*, **42**(1), pp. 131–140.
- [10] Jansson, D. G., and Smith, S. M., 1991. "Design fixation". *Design Studies*, **12**(1), pp. 3–11.
- [11] Purcell, A. T., and Gero, J. S., 1996. "Design and other types of fixation". *Design studies*, **17**(4), pp. 363–383.
- [12] Viswanathan, V. K., and Linsey, J. S., 2013. "Design fixation and its mitigation: a study on the role of expertise". *Journal of Mechanical Design*, **135**(5), p. 051008.
- [13] Beitz, W., Pahl, G., and Grote, K., 1996. "Engineering design: a systematic approach". *Mrs Bulletin*, **71**.
- [14] Kirton, M. J., 2004. *Adaption-innovation: In the context of diversity and change*. Routledge.
- [15] Sarkar, P., and Chakrabarti, A., 2011. "Assessing design creativity". *Design studies*, **32**(4), pp. 348–383.
- [16] Crilly, N., 2019. "Creativity and fixation in the real world: A literature review of case study research". *Design Studies*, **64**, pp. 154–168.
- [17] Buhl, A., Schmidt-Keilich, M., Muster, V., Mazzejewski, S., Schrader, U., Harrach, C., Schäfer, M., and Süßbauer, E., 2019. "Design thinking for sustainability: Why and how design thinking can foster sustainability-oriented innovation development". *Journal of cleaner production*, **231**, pp. 1248–1257.
- [18] Cash, P., and Štorga, M., 2015. "Multifaceted assessment of ideation: using networks to link ideation and design activity". *Journal of Engineering Design*, **26**(10-12), pp. 391–415.
- [19] Dorst, K., and Cross, N., 2001. "Creativity in the design process: co-evolution of problem–solution". *Design studies*, **22**(5), pp. 425–437.
- [20] Sosa, R., 2019. "Accretion theory of ideation: Evaluation regimes for ideation stages". *Design Science*, **5**.
- [21] Gonçalves, M., and Cash, P., 2021. "The life cycle of creative ideas: Towards a dual-process theory of ideation". *Design Studies*, **72**, p. 100988.
- [22] Renner, G., and Ekárt, A., 2003. "Genetic algorithms in computer aided design". *Computer-aided design*, **35**(8), pp. 709–726.
- [23] Yuan, C., Marion, T., and Moghaddam, M., 2022. "Leveraging end-user data for enhanced design concept evaluation: A multimodal deep regression model". *Journal of Mechanical Design*, **144**(2).
- [24] Vasconcelos, L. A., Cardoso, C. C., Sääksjärvi, M., Chen, C.-C., and Crilly, N., 2017. "Inspiration and fixation: the influences of example designs and system properties in idea generation". *Journal of Mechanical Design*, **139**(3).
- [25] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y., 2014. "Generative adversarial networks". *arXiv preprint arXiv:1406.2661*.
- [26] Teng, L., Fu, Z., and Yao, Y., 2020. "Interactive translation in echocardiography training system with enhanced cycle-gan". *IEEE Access*, **8**, pp. 106147–106156.
- [27] Karras, T., Laine, S., and Aila, T., 2019. "A style-based generator architecture for generative adversarial networks". In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 4401–4410.
- [28] Chen, L., Maddox, R. K., Duan, Z., and Xu, C., 2019. "Hierarchical cross-modal talking face generation

- with dynamic pixel-wise loss”. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7832–7841.
- [29] Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., and Lee, H., 2016. “Generative adversarial text to image synthesis”. *arXiv preprint arXiv:1605.05396*.
  - [30] Kim, T., Cha, M., Kim, H., Lee, J. K., and Kim, J., 2017. “Learning to discover cross-domain relations with generative adversarial networks”. In International conference on machine learning, PMLR, pp. 1857–1865.
  - [31] Shah, J. J., Smith, S. M., and Vargas-Hernandez, N., 2003. “Metrics for measuring ideation effectiveness”. *Design studies*, **24**(2), pp. 111–134.
  - [32] Chen, W., and Ahmed, F., 2021. “Padgan: Learning to generate high-quality novel designs”. *Journal of Mechanical Design*, **143**(3).
  - [33] Xu, K., Zhang, G., Liu, S., Fan, Q., Sun, M., Chen, H., Chen, P.-Y., Wang, Y., and Lin, X., 2020. “Adversarial t-shirt! evading person detectors in a physical world”. In European conference on computer vision, Springer, pp. 665–681.
  - [34] Cheong, H., and Shu, L., 2014. “Retrieving causally related functions from natural-language text for biomimetic design”. *Journal of Mechanical Design*, **136**(8), p. 081008.
  - [35] Behzadi, M. M., and Ilies, H. T., 2022. “Gantl: Toward practical and real-time topology optimization with conditional generative adversarial networks and transfer learning”. *Journal of Mechanical Design*, **144**(2).
  - [36] Lee, M., Park, Y., Jo, H., Kim, K., Lee, S., and Lee, I., 2022. “Deep generative trend pattern design framework for efficient conceptual design”. *Journal of Mechanical Design*, **144**(7), p. 071703.
  - [37] Kazemi, H., Seepersad, C. C., and Alicia Kim, H., 2022. “Multiphysics design optimization via generative adversarial networks”. *Journal of Mechanical Design*, **144**(12), p. 121702.
  - [38] Regenwetter, L., Nobari, A. H., and Ahmed, F., 2022. “Deep generative models in engineering design: A review”. *Journal of Mechanical Design*, **144**(7), p. 071704.
  - [39] Shu, D., Cunningham, J., Stump, G., Miller, S. W., Yukish, M. A., Simpson, T. W., and Tucker, C. S., 2020. “3d design using generative adversarial networks and physics based validation”. *Journal of Mechanical Design*, **142**(7).
  - [40] Oh, S., Jung, Y., Kim, S., Lee, I., and Kang, N., 2019. “Deep generative design: Integration of topology optimization and generative models”. *Journal of Mechanical Design*, **141**(11).
  - [41] Yuan, C., and Moghaddam, M., 2020. “Attribute-aware generative design with generative adversarial networks”. *IEEE Access*, **8**, pp. 190710–190721.
  - [42] Liang, X., Lin, L., Yang, W., Luo, P., Huang, J., and Yan, S., 2016. “Clothes co-parsing via joint image segmentation and labeling with application to clothing retrieval”. *IEEE Transactions on Multimedia*, **18**(6), pp. 1175–1186.
  - [43] Heyrani Nobari, A., Rashad, M. F., and Ahmed, F., 2021. “Creativegan: Editing generative adversarial networks for creative design synthesis”. In International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Vol. 85383, American Society of Mechanical Engineers, p. V03AT03A002.
  - [44] Elgammal, A., Liu, B., Elhoseiny, M., and Mazzone, M., 2017. “Can: Creative adversarial networks generating “art” by learning about styles and deviating from style norms”. In 8th International Conference on Computational Creativity, ICCO 2017, Georgia Institute of Technology.
  - [45] Nobari, A. H., Chen, W., and Ahmed, F., 2022. “Range-constrained generative adversarial network: Design synthesis under constraints using conditional generative adversarial networks”. *Journal of Mechanical Design*, **144**(2).
  - [46] Huang, X., Liu, M.-Y., Belongie, S., and Kautz, J., 2018. “Multimodal unsupervised image-to-image translation”. In Proceedings of the European conference on computer vision (ECCV), pp. 172–189.
  - [47] Choi, Y., Choi, M., Kim, M., Ha, J.-W., Kim, S., and Choo, J., 2018. “Stargan: Unified generative adversarial networks for multi-domain image-to-image translation”. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 8789–8797.
  - [48] Chen, Q., Wang, T., Pope, P., Fuge, M., et al., 2022. “Inverse design of two-dimensional airfoils using conditional generative models and surrogate log-likelihoods”. *Journal of Mechanical Design*, **144**(2).
  - [49] Yunlong, M., and Xu, L., 2020. “Diversity augmented conditional generative adversarial network for enhanced multimodal image-to-image translation”.
  - [50] Lin, J., Chen, Z., Xia, Y., Liu, S., Qin, T., and Luo, J., 2019. “Exploring explicit domain supervision for latent space disentanglement in unpaired image-to-image translation”. *IEEE transactions on pattern analysis and machine intelligence*, **43**(4), pp. 1254–1266.
  - [51] Che, T., Li, Y., Jacob, A. P., Bengio, Y., and Li, W., 2016. “Mode regularized generative adversarial networks”. *5th International Conference on Learning Representations*.
  - [52] Chen, W., and Ahmed, F., 2021. “Mo-padgan: Reparameterizing engineering designs for augmented multi-

- objective optimization". *Applied Soft Computing*, **113**, p. 107909.
- [53] Wolfram, S., 2002. *A new kind of science*, Vol. 5. Wolfram media Champaign, IL.
  - [54] Lindenmayer, A., 1968. "Mathematical models for cellular interactions in development i. filaments with one-sided inputs". *Journal of theoretical biology*, **18**(3), pp. 280–299.
  - [55] Stiny, G., 1980. "Introduction to shape and shape grammars". *Environment and planning B: planning and design*, **7**(3), pp. 343–351.
  - [56] Holland, J. H., et al., 1992. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press.
  - [57] Kennedy, J., 2006. "Swarm intelligence". In *Handbook of nature-inspired and innovative computing*. Springer, pp. 187–219.
  - [58] Patel, N. M., Kang, B.-S., Renaud, J. E., and Tovar, A., 2009. "Crashworthiness Design Using Topology Optimization". *Journal of Mechanical Design*, **131**(6), 05. 061013.
  - [59] Hornby, G. S., Lipson, H., and Pollack, J. B., 2001. "Evolution of generative design systems for modular physical robots". In Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No. 01CH37164), Vol. 4, IEEE, pp. 4146–4151.
  - [60] Krish, S., 2011. "A practical generative design method". *Computer-Aided Design*, **43**(1), pp. 88–100.
  - [61] Caldas, L. G., and Norford, L. K., 2002. "A design optimization tool based on a genetic algorithm". *Automation in construction*, **11**(2), pp. 173–184.
  - [62] Lobos, A., 2018. "Finding balance in generative product design".
  - [63] Burnap, A., Liu, Y., Pan, Y., Lee, H., Gonzalez, R., and Papalambros, P. Y., 2016. "Estimating and exploring the product form design space using deep generative models". In International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Vol. 50107, American Society of Mechanical Engineers, p. V02AT03A013.
  - [64] Yang, Z., Li, X., Catherine Brinson, L., Choudhary, A. N., Chen, W., and Agrawal, A., 2018. "Microstructural materials design via deep adversarial learning methodology". *Journal of Mechanical Design*, **140**(11).
  - [65] Valencia-Rosado, L. O., Guzman-Zavaleta, Z. J., and Starostenko, G., 2022. "A modular generative approach for realistic river deltas: When l-systems and cgans meet". *IEEE Access*, **10**, pp. 5753–5767.
  - [66] Huo, Q., Tang, G., and Zhang, F., 2019. "Particle swarm optimization for great enhancement in semi-supervised retinal vessel segmentation with generative adversarial networks". In *Machine Learning and Medical Engineering for Cardiovascular Health and Intravascular Imaging and Computer Assisted Stenting*. Springer, pp. 112–120.
  - [67] Iklima, Z., Adriansyah, A., and Hitimana, S., 2021. "Self-collision avoidance of arm robot using generative adversarial network and particles swarm optimization (gan-pso)". *Sinergi*, **25**(2), pp. 141–152.
  - [68] Chen, W., and Ahmed, F., 2020. "Padgan: A generative adversarial network for performance augmented diverse designs". In International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Vol. 84003, American Society of Mechanical Engineers, p. V11AT11A010.
  - [69] Raina, A., McComb, C., and Cagan, J., 2012. "Learning to design from humans: Imitating human designers through deep learning". *Journal of Mechanical Design*, **141**(11).
  - [70] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. C., 2017. "Improved training of wasserstein gans". *Advances in neural information processing systems*, **30**.
  - [71] Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., and Aila, T., 2020. "Analyzing and improving the image quality of stylegan". In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 8110–8119.
  - [72] Arjovsky, M., Chintala, S., and Bottou, L., 2017. "Wasserstein generative adversarial networks". In International conference on machine learning, PMLR, pp. 214–223.
  - [73] Villani, C., 2016. *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg.
  - [74] Han, Y., and Moghaddam, M., 2021. "Analysis of sentiment expressions for user-centered design". *Expert Systems with Applications*, **171**, p. 114604.
  - [75] He, K., Zhang, X., Ren, S., and Sun, J., 2016. "Deep residual learning for image recognition". In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.
  - [76] Kenton, J. D. M.-W. C., and Toutanova, L. K., 2019. "Bert: Pre-training of deep bidirectional transformers for language understanding". In Proceedings of NAACL-HLT, pp. 4171–4186.
  - [77] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al., 2019. "Pytorch: An imperative style, high-performance deep learning library". *Advances in Neural Information Processing Systems (NeurIPS 2019)*.
  - [78] Kingma, D. P., and Ba, J., 2015. "Adam: A method for stochastic optimization". *International Conference for Learning Representations, San Diego, 2015*.

- [79] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S., 2017. “Gans trained by a two time-scale update rule converge to a local nash equilibrium”. *Advances in neural information processing systems*, **30**.
- [80] Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., and Smola, A., 2012. “A kernel two-sample test”. *The Journal of Machine Learning Research*, **13**(1), pp. 723–773.
- [81] Wilson, A. G., Hu, Z., Salakhutdinov, R., and Xing, E. P., 2016. “Deep kernel learning”. In *Artificial intelligence and statistics*, PMLR, pp. 370–378.
- [82] <https://nealjean.com/ml/frechet-inception-distance/>.
- [83] Ravuri, S., Mohamed, S., Rosca, M., and Vinyals, O., 2018. “Learning implicit generative models with the method of learned moments”. In *International Conference on Machine Learning*, PMLR, pp. 4314–4323.
- [84] Raviselvam, S., Hölttä-Otto, K., and Wood, K. L., 2016. “User extreme conditions to enhance designer empathy and creativity: Applications using visual impairment”. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Vol. 50190, American Society of Mechanical Engineers, p. V007T06A005.
- [85] Chong, L., Raina, A., Goucher-Lambert, K., Kotovsky, K., and Cagan, J., 2022. “The evolution and impact of human confidence in artificial intelligence and in themselves on ai-assisted decision-making in design”. *Journal of Mechanical Design*, **145**(3), p. 031401.
- [86] Nelson, B. A., Wilson, J. O., Rosen, D., and Yen, J., 2009. “Refined metrics for measuring ideation effectiveness”. *Design Studies*, **30**(6), pp. 737–743.

Accepted Manuscript Not Copyedited