Design Science

DCG-GAN: design concept generation with generative adversarial networks

Parisa Ghasemi[1], Chenxi Yuan[2], Tucker Marion[1,3] and Mohsen Moghaddam[1,4]

[1]*Department of Mechanical & Industrial Engineering, Northeastern University, Boston, MA, USA*
[2]*School of Medicine, University of Pennsylvania, Philadelphia, PA, USA*
[3]*D'Amore-McKim School of Business, Northeastern University, Boston, MA, USA*
[4]*Khoury College of Computer Sciences, Northeastern University, Boston, MA, USA*

Abstract

Generative adversarial networks (GANs) have recently been proposed as a potentially disruptive approach to generative design due to their remarkable ability to generate visually appealing and realistic samples. Yet, we show that the current generator-discriminator architecture inherently limits the ability of GANs as a design concept generation (DCG) tool. Specifically, we conduct a DCG study on a large-scale dataset based on a GAN architecture to advance the understanding of the performance of these generative models in generating novel and diverse samples. Our findings, derived from a series of comprehensive and objective assessments, reveal that while the traditional GAN architecture can generate realistic samples, the generated and style-mixed samples closely resemble the training dataset, exhibiting significantly low creativity. We propose a new generic architecture for DCG with GANs (DCG-GAN) that enables GAN-based generative processes to be guided by geometric conditions and criteria such as novelty, diversity and desirability. We validate the performance of the DCG-GAN model through a rigorous quantitative assessment procedure and an extensive qualitative assessment involving 89 participants. We conclude by providing several future research directions and insights for the engineering design community to realize the untapped potential of GANs for DCG.

Keywords: Design concept generation, Generative adversarial networks, Novelty, Diversity, Creativity

Received 20 November 2023
Revised 23 May 2024
Accepted 24 May 2024

Corresponding author
Parisa Ghasemi
ghasemi.p@northeastern.edu

Des. Sci., vol. 10, e14
journals.cambridge.org/dsj
DOI: 10.1017/dsj.2024.13

1. Introduction

Design is a complex cognitive process that requires designers to make creative connections across different areas of knowledge. This process includes carefully identifying and solving problems that may not have been dealt with before or have been approached in unique ways in the past (Dieter, Schmidt & Azarm 2009). Venturing into new territories within the design realm increases the chances of finding new and inventive solutions. However, this kind of exploration can take a long time and may be influenced by preconceived notions, a fixation on initial ideas and personal biases (Linsey *et al.* 2010; Vasconcelos *et al.* 2017). Designers often aspire to navigate the design space uniformly or adapt it to meet specific requirements. Computational technologies, particularly generative artificial intelligence

(AI) methods, offer a promising avenue to accelerate searching and generating novel design concepts within the solution space. Existing generative design approaches can be categorized into five main classes, including shape grammars, L-systems, cellular automata, genetic algorithms and swarm intelligence (Singh & Gu 2012). These approaches typically enhance design generation through the application of mathematical functions or physics-based simulations (Shu *et al.* 2020). The generative design capabilities of commercial CAD packages focus on a limited set of conditions (e.g., spatial constraints) and criteria (e.g., optimizing mass or structural strength) (Buonamici *et al.* 2020). All the methodologies and tools mentioned above are aimed at creating optimized production-ready designs rather than fostering unique and innovative design concepts for faster and more efficient ideation during the early stages of the design process. However, the premise of design concept generation (DCG) is to enhance the efficiency, quality and consistency of the design process by automatically generating numerous and diverse samples for designers to synthesize, choose from and edit, thus elevating their roles to "curators" and "empathizers." In this study, we describe a design concept as a visual representation that captures the fundamental idea behind a product's design. It takes the form of an image.

With the growing abundance of publicly available data (e.g., product data and user reviews) and recent advances in AI methods such as generative adversarial networks (GANs; Goodfellow *et al.* 2014), there has been a recent surge in the adoption of AI-driven approaches for design automation (e.g., Burnap *et al.* 2016; Oh *et al.* 2019; Shu *et al.* 2020). GANs are a relatively recent method in modern AI and have demonstrated state-of-the-art performance in many generative modeling tasks. GANs have been used to solve a variety of generative design problems, from creating 3D aircraft models in native format for detailed simulation (Shu *et al.* 2020) to topology optimization of 2D wheel design (Oh *et al.* 2019) and generating realistic fashion apparel style recommendations (Yuan & Moghaddam 2020). However, traditional GANs limit their suitability for DCG, which requires divergent thinking and imagination, since esthetics and creativity are crucial in design (Buonamici *et al.* 2020).

The existing AI-driven design automation literature lacks a generic computational framework to conduct DCG studies guided by various design conditions and criteria to augment the creative design process. Despite the possibilities of GANs to produce realistic design outcomes, it is not yet clear how existing GAN architectures can support creativity since they tend inherently to replicate the training dataset with the same characteristics due to their sole focus on generating samples that "look real." The lack of creativity is due to the fact that during the training process, the GAN generator is urged to produce samples close to the training data distribution to deceive the discriminator in a minimax game, which ultimately constrains design output, particularly in terms of variety and originality. In this article, we explore the above gaps in AI-driven DCG knowledge by first conducting a thorough quantitative analysis of the limitations of traditional models, then proposing a generic GAN-based architecture for multi-criteria sample generation, and finally customizing it for DCG.

1.1. Knowledge gaps

Creativity, as an indispensable element of the design process, is generally defined as "the capacity to produce original and valuable items by flair" (Gaut 2010). Yet, it is

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

often difficult to objectively assess due to its intangible and subjective nature. In the context of engineering design, the definition of creativity can be translated into maximizing the degree of novelty and usefulness of the design concepts generated (Shah, Smith & Vargas-Hernandez 2003). Novelty can be gauged by how different an idea is relative to others, while usefulness can be measured in terms of the quality and performance of the design (Toh, Miller & Okudan Kremer 2014). In addition, evidence suggests that the quality, performance and originality of the design often correlate with the diversity of the concepts generated and the design space explored (Osborn 1953; Dow *et al.* 2010; Vasconcelos *et al.* 2017). Therefore, we focus on diversity and novelty as two fundamental criteria for objectively assessing the performance of GAN-based DCG in terms of creativity (Wang, She & Ward 2021).

There are two main methods in the design literature for measuring diversity: subjective rating and the genealogical tree approach (Ahmed 2019). An example of subjective rating of design space diversity is categorizing a set of design ideas into various idea pools based on intuitive categories. This method is efficient in terms of time and effort, but the results may not be as valid or reliable since the inferences are based on the rater's mental models. A genealogical tree adopts deterministic rules derived from design attributes to rate the diversity of a set of design ideas. This set of approaches is repeatable and relatively more objective; however, they lack sensitivity and accuracy since they use the same set of formulae for all types of design problems.

Diversity augmentation in GANs is a crucial research focus, aiming to enhance the variety of generated outputs while maintaining quality. Our research categorizes diverse GAN models based on strategies modifying the traditional GAN architecture (Section 2.3). One set, including mode seeking GAN (MS-GAN) and its extensions like diversity sensitive conditional GAN (DS-GAN), diversity balancing GAN, diversity conditional GAN (DivCo GAN) and diversity augmented GAN (DivAug GAN), introduces additional regularization terms to the loss function. Another strategy, exemplified by personalized diversity promoting GAN (PD-GAN), manipulates the generation process within the generator itself. The third category employs data manipulation techniques in models like GAN+ and easy data augmentation coupled with GAN (EDA + GAN). Bagging-inspired methods, such as EDA + GAN, form the fourth category. Models like classification-reinforced GAN (CLS-R GAN) and diversity promoting GAN (DP-GAN) leverage reinforcement learning principles to enhance diversity. Lastly, models like PD-GAN focus on enriching diversity by manipulating latent vectors progressively.

In this article, we depart from employing intra-batch pair-to-pair distance averaging for diversity assessment, as it induces an overall diversity shift rather than ensuring diversification across all generated samples. Instead, we adopt the minimum distance among all pairs, a worst-case scenario metric, as our diversity measure. This approach compels the generator to promote diversity uniformly across all generated samples, avoiding a selective impact on the diversity average. Furthermore, we opt to operate on semantic features within the generator's output, as statistical and probabilistic models employed for diversity assessment lack inherent comprehension of image semantic features. The resultant high-dimensional feature vectors, representing the images, are then evaluated for diversity. Given the high-dimensional nature of the feature space, computational efficiency is crucial. We employ the covering radius upper bound (CRUB) method for its computational efficiency, as it considers diversity across multiple dimensions

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

without involving computationally expensive operations. CRUB's emphasis on the CRUB facilitates a direct assessment of sample spread or coverage within a space, rendering it scalable and adaptable to high-dimensional spaces. However, it is noteworthy that alternative feature extraction and diversity measures may be substituted within our algorithm. Furthermore, drawing inspiration from Wu *et al.* (2019), which diversifies input noise vectors across various categories, we enhance the input noise vectors' diversity through an approach involving extensive sampling of a pool of vectors and then selecting the most diverse subset using stratified sampling, enabling exploration of uncharted areas.

Assessing the novelty of concepts during the design process, with a focus on identifying instances likely to succeed, stands as a formidable challenge. Within the GANs literature, there has been an exploration into the augmentation of novelty, often interchangeably referred to as creativity. This exploration is evidenced by various approaches in the following examples of novelty-augmented GANs. The CreativeGAN method (Heyrani Nobari, Rashad & Ahmed 2021) introduces a procedure for detecting novel components in generated samples by identifying the most unique designs, concealing their novel features and subsequently modifying the GAN architecture to prioritize the generation of designs featuring these unique components. Conversely, creative adversarial networks (CAN; Elgammal *et al.* 2017) adapt the GAN loss function by introducing a style classification loss and a style ambiguity loss. This modification aims to achieve a triple objective: generating novel works, ensuring the generated work remains within the distribution to mitigate excessive arousal and negative hedonic experiences, and enhancing the stylistic ambiguity of the generated outputs. In contrast, Combinets (Guzdial & Riedl 2019) leverages features acquired by generative models trained on existing classes to create new models without supplementary training. This method entails establishing a high-dimensional search space from pretrained models, combining or varying features to ascertain the contribution and inclusion of each feature in the final conceptual expansion.

In this article, we have devised algorithms to quantitatively measure diversity and novelty, drawing upon existing literature on these concepts. A particular focus of our efforts involves distinguishing between these two terms during the formulation of our metrics. Notably, the algorithms formulated for metric computation are designed to be versatile, with each component being interchangeable with alternative methods for improved accuracy or efficiency. This adaptability ensures that the proposed algorithms remain independent of specific components. Moreover, the developed algorithms hold utility as a loss function within various neural architectures, contributing to its broader applicability in diverse computational frameworks.

1.2. Objectives and contributions

This article presents a systematic and objective assessment of the creativity of GAN-based DCG, measured in terms of diversity and novelty, as well as building and validating a new architecture to compensate for the traditional GAN architecture. Specifically, a baseline GAN architecture, Style-GAN2 (Karras *et al.* 2020), is applied to create 2D visual concepts (i.e., images of sneakers) based on a large training dataset. The initial findings demonstrate that although the trained generator is capable of producing realistic and authentic-looking images of sneakers,

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

the generated samples strongly resemble existing products (i.e., the training dataset). As the generator solely concentrates on outsmarting the discriminator by creating samples that look like the training dataset, it results in a lack of originality and variety, which limits its ability to generate creative designs. It must be noted that the purpose of this article is to underscore the limitations of modern generative AI models for DCG, with particular emphasis on key metrics such as novelty and diversity, and propose rigorous quantitative methods for modeling and optimizing for such metrics. Yet, the authors do not claim the outcomes as state-of-the-art in terms of image quality, as more recent backbones such as diffusion models (Dhariwal & Nichol 2021) already outperform GANs in that regard. Specifically, the contributions of this article are as follows:

1. Mathematical models of diversity and novelty concepts have been formulated, informed by an exhaustive literature review, to scrutinize their definitions within the design literature. Subsequently, novel component-agnostic algorithms are developed to assess novelty and diversity in the context of design concept evaluation, with potential applicability to diverse domains and applications.
2. A generic GAN-based architecture is designed and tailored for multi-objective learning scenarios, where the objectives may not be inherently related or aligned. Our findings demonstrate the feasibility of concurrently achieving multiple objectives, as evidenced by the performance of the DCG-GAN, an exemplification of this generic multi-objective architecture. The proposed approach involves incorporating additional evaluators' feedback into the generator's loss function, alongside the discriminator's loss. This regularization enables the generator to simultaneously learn multiple domain-specific and domain-agnostic criteria, making it a versatile and effective generative tool for meeting various predefined benchmarks and performance standards.
3. DCG-GAN, a customized variant of the proposed generic architecture is introduced for DCG. This adaptation is specifically designed to meet the demands of generating design concepts that balance esthetics and functionality. DCG-GAN incorporates design specifications and constraints into the evaluation network, focusing on diversity, novelty, desirability and adherence to a given silhouette in visual design. The loss function is regularized with four additional terms to assess generated samples against specific criteria, utilizing our formulated algorithms and adapting other methods for implementation. Specifically, DCG-GAN has been enhanced by integrating LOF, CRUB + Stratified Sampling, SSIM and DMDE to augment novelty, diversity, geometrical proportionality and desirability, respectively. Each loss term leverages these methods to specifically evaluate distinct attributes of the generated design concepts, ensuring a comprehensive assessment of their esthetic and functional qualities.
4. The culmination of computational and subjective assessment methodologies consistently shows the enhanced capabilities of DCG-GAN compared to the baseline model. The evaluation process involved visual analysis, quantitative metrics and qualitative assessments in a survey format with 90 participants. This comprehensive approach provided a thorough assessment of DCG-GAN's performance in terms of the diversity and novelty of the generated samples. Evidently, DCG-GAN exhibits superior performance in generating design

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

concepts that transcend mere realism, embracing attributes of novelty, diversity, desirability and geometrical proportionality. We speculate that these findings may be generalizable to more recent backbone generative AI models, a hypothesis we intend to explore in future studies.

The remainder of this article is organized as follows: Section 2 presents the architecture of the GAN models used in this article, discusses GAN-based DCG, discusses the terms diversity, novelty and desirability in the context of design, and finally provides a comprehensive review of diversity-augmented GAN models. Sections 3 and 4 present the proposed diversity and novelty evaluation algorithms, as well as the DCG-GAN architecture. Section 5 discusses and analyzes visual results through a rigorous quantitative and qualitative assessment. In Section 6, we discuss the limitations of GANs for concept generation and generative design and elucidate how our novel algorithms and models can effectively mitigate these challenges. Moreover, we discuss the limitations of this study and possible future research directions in this domain.

2. Background

AI-driven DCG can serve as a powerful and transformative tool for designers to efficiently create more original and useful concepts. Advanced data-driven models can be developed to automatically analyze large amounts of product and user data, comprehend intricate patterns, invent new ideas and evaluate them based on existing performance and user data, as well as other requirements and metrics. As a result, the designer can shift their focus from dragging and dropping to iterating over designs, selecting, integrating and modifying AI-generated concepts. GANs are one of the generative models capable of generating realistic images according to the distribution of the input dataset to an extent that is not recognizable as synthetic data by human eyes. Moreover, GANs are capable of producing a large number of solutions in a relatively short period. These properties make GANs a potentially disruptive approach to generate myriad design concepts with little effort. To illuminate the capabilities and limitations of GANs for DCG, this section provides the background necessary for the reader to understand the general logic of the standard GAN model, followed by a description of StyleGAN architecture (Karras, Laine & Aila 2019; Karras *et al.* 2020). Consecutively, this discussion covers recent developments in GAN-based generative design, explores emerging challenges in the field and introduces a data-driven design evaluation method that has the potential to address some of the key limitations associated with GANs in DCG.

2.1. GAN-based DCG and generative design

Generative design refers to an automated design exploration process that analyzes all possible solutions to a design problem based on the specified requirements and constraints and then selects the suitable ones among them. Generative design and DCG share a common iterative approach to exploring a broad solution space. Nevertheless, they diverge in their respective objectives and applications. DCG primarily focuses on generating a multitude of approximate solutions aimed at inspiring designers during the ideation phase, rather than optimizing a design for production. DCG applies a bottom-up approach, in contrast to a traditional

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

designer-based top-down approach, enabling exploration of a wider range of complex solutions. Since there is no single correct answer to a design problem, given the high and even infinite degrees of freedom in product design, searching for all possible solutions could be resource-exhaustive and not practical to be executed by humans. Most of the well-known generative design methods operate on the basis of a set of defined design rules to iteratively evolve, or possibly optimize, an initial (usually selected randomly) solution to satisfy certain requirements. In contrast, GAN models are not limited to predefined rules, but instead attempt to search the design space based on the distribution of the provided dataset. Thus, GANs are a favorable choice for DCG.

Concept design and development are rooted in the visualization of the concepts. These can take the form of human sketches or digitally produced images. Recently, there has been an acceleration of software that enables the quick visualization of design concepts. AI is being used frequently in these applications, which include platforms such as Dall-E and Midjourney. Visualization allows the designer to express concepts and gauge the efficacy of potential solutions (Roozenburg & Eekels 1995). Nearly, all products use a form of visualization of design concepts. In the main, this takes the form of imagination visualization, which enables the designer to experience the creation of a new, never-before-seen expression of a potential solution. In product design, these are most often visual expressions of the physical appearance of a tangible product (such as the side view of a car, or the isometric view of a smartphone) or the user interface of a digital one (the interface of a mobile app) (Ulrich, Eppinger & Yang 2020). These visualizations are essential to the product design and development process. A key in concept development is the ability to have a large variety of concept designs to evaluate, of sufficient variety that includes a diverse set of samples, each with some aspect of novelty (Macedo & Cardoso 2002). Image generation using AI has the ability to greatly increase the efficiency and effectiveness of visual concept design (Li *et al.* 2021). As such, DCG-GAN, with its ability to increase the novelty and diversity of realistic concept visualization, can be applied in nearly any industry where concept visualization is a part of the design process.

GANs have proven their versatility across various design domains, showcasing their efficacy in addressing diverse design challenges (Jiang *et al.* 2022). Jiang *et al.* (2022) introduced a GAN-based platform facilitating mass customization, leveraging user preferences to tailor product structures autonomously, enabling workflow independence from rigid rules. Qian, Tan & Ye (2022*b*) utilized GANs to design high-toughness, high-stiffness architectured composite materials, achieving a significant reduction in required data samples compared to traditional methods. They extended this approach to layout design, surpassing existing methods in accuracy and efficiency (Qian, Tan & Ye 2022*a*). Siriwardane *et al.* (2022) employed GANs to discover stable semiconductors efficiently, using a cubic-GAN-based pipeline for candidate generation and high-throughput screening to evaluate band gaps. Wang *et al.* (2020) explored a neuroscience-inspired design approach using machine learning and EEG signals to capture preferred design attributes. Gui *et al.* (2021) devised a novel visual comfort generative network, demonstrating superior performance in generating underground spaces according to specified comfort levels. Additionally, Yuan & Moghaddam (2020) developed a design attribute GAN model enhancing the accuracy and visual appeal of generated design concepts, signifying GANs' wide-ranging impact in various design contexts.

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

As the successful cases above suggest, GANs are promising in the automation of all or parts of the design process by efficiently producing multiple solutions for a design problem that is usually unachievable without the help of computational tools. However, considering the goal of GANs to mimic the distribution of the input data, these properties may not be achievable using the raw version of these models. Thus, GAN architectures as a design tool require an evaluation method to provide them with feedback on the novelty and diversity of their outputs in the training process. Subsequently, in the following sections, the definitions of diversity and novelty in the design literature are presented.

2.2. Diversity, novelty and desirability in product design

Novelty and diversity are central themes in design and engineering innovation. According to the Osborn rule for brainstorming (Osborn 1953), the availability of a more diverse set of solutions and the uniqueness of the solutions can increase the chances of proposing a successful design instance. A prominent instantiation of design is the esthetic qualities of the design (Norman 2013). Research has shown that the esthetic attributes of a design can positively or negatively impact customer perception and market performance (Landwehr, Labroo & Herrmann 2011, 2013). Research has also shown that the more users are exposed to design, the higher the likelihood that they are drawn to atypical design (Liu *et al.* 2017). As such, we can tie the esthetic qualities of design to emotion and consumer acceptance (Bloch 1995), while at the same time noting that function also has importance (Norman 2002). In the design of consumer products such as sneakers, the link between novelty, esthetics and trendiness is crucial (Hsiao & Chen 2006). As such, designers create and iterate new designs that seek to capture or anticipate design trends and preference. Trendiness and emotion have been identified as key attributes of design novelty (Hsiao & Chen 2006). A higher quantity and diversity of potential esthetic design solutions can increase the likelihood of achieving a novel and pleasing outcome (Marion & Fixson 2018). Generative AI, with its speed and ability to create massive numbers of iterations, stands to greatly impact the quantity of new design output. There are several architectures of generative AI used to create esthetic images. Image generators based on GANs have been studied for several years. Limitations on the output of esthetic designs from GANs include a lack of novelty and diversity which is due to the fundamental architecture of the models. However, GANs are noted for their ability to produce high-quality and realistic images. During new product design and development, initial concepts are refined and made more realistic as they progress through the development cycle (Ulrich *et al.* 2020). It is here where realistic esthetic concept images are extremely valuable, not only in gauging early customer feedback, but transitioning to more detailed engineering. As such, while there are generative AI architectures such as diffusion models that excel at novelty and diversity (Carlini *et al.* 2023), there is still a need in design for realism, particularly when dealing with defined product requirements. Therefore, while an older technology, given the efficiency and quality images produced from GANs, there is value in investigating how might the novelty and diversity of GANs be improved (Yuan, Marion & Moghaddam 2023). And if it can be improved, by what measurable amount.

The three main metrics discussed in this subsection are only a subset of all metrics that could be integrated into GAN-based DCG processes as design

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

conditions and criteria to enhance its capabilities beyond the mere generation of realistic samples. With this motivation in mind, this article first presents a comprehensive quantitative assessment of a GAN architecture in terms of novelty and diversity (Sections 3.1 and 3.2), followed by a new architecture (Section 4) for GAN-based DCG.

2.3. Diversity-augmented GANs

Diversity augmentation has emerged as a vital area of research in GANs, with the primary objective of improving the variety while preserving the quality of the generated outputs. GANs are often plagued by mode collapse, wherein they produce only a limited subset of samples, failing to encompass the entire diversity of the target distribution. Consequently, the generated outputs lack variety and do not accurately represent the full data manifold. In our research on diversity augmentation in GANs, we selected prominent models to the best of our knowledge and categorized them based on their strategies for restructuring the traditional GAN architecture. These strategies involve modifying or extending the standard GAN framework by incorporating supplementary components, regularization terms or loss functions that facilitate the generation of more diverse and novel samples. The following review categorizes and provides an overview of selected GAN models that utilize diversity-augmented approaches.

1. The loss regularization category comprises several models aimed at mitigating mode collapse in GANs by introducing additional regularization terms into the loss function. Within this category, we identified MS-GAN (Mao *et al.* 2019) and its extensions, namely DS-GAN (Yang *et al.* 2019), diversity balancing GAN (Dubinski *et al.* 2022), DivCo GAN (Liu *et al.* 2021*b*) and DivAug GAN (Meng & Xu 2020). MS-GAN, operating on conditional GAN principles, proposes a novel regularization term to maximize the ratio of distances between generated images and their corresponding latent codes. Extending MS-GAN, DivCo GAN introduces a contrastive loss that encourages similarity between images generated from adjacent latent codes while promoting dissimilarity between images from distinct latent codes. DivAug GAN, another extension of MS-GAN, defines a new regularization term to enhance mode diversity by exploring unseen image space, ensuring relative variation consistency and maximizing distinction when injecting different noise vectors. Furthermore, performance augmented diverse GAN (PAD-GAN; Chen & Ahmed 2020) introduces a unique loss function that employs the determinantal point process (DPP) kernel, effectively augmenting quality and diversity simultaneously by establishing a global measure of similarity between pair of items. This kernel ensures a balanced representation of quality and diversity in generated samples. Lastly, CLS-R GAN (Kim & Lee 2023) introduces an additional discriminator-independent classifier that assesses the quality of the generated images.
2. Inside generator augmentation strategies, exemplified by PD-GAN (Wu *et al.* 2019), diversify outputs by manipulating the generation process within the generator itself. PD-GAN employs personalized ranking mechanisms based on diversity metrics, fostering the production of diverse, high-quality samples.
3. In the data augmentation category, models like GAN+ (Yean *et al.* 2021) and EDA + GAN (Wu & Huang 2022) leverage data manipulation techniques to

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

broaden the diversity of generated outputs. GAN+ employs a two-step approach involving dataset sampling using the Dirichlet method, while EDA + GAN integrates data augmentation as a preprocessing step before training.

4. The bagging-inspired category encompasses models that draw inspiration from the principles of bagging methods in machine learning to enhance diversity in GAN-generated samples. One such model within this category is EDA + GAN. In this approach, the utilization of data augmentation techniques serves as a bagging-inspired strategy.
5. Reinforcement-learning-inspired approaches, represented by models like CLS-R GAN and DP-GAN (Xu *et al.* 2018), utilize reinforcement learning principles to promote diversity. DP-GAN employs an LSTM-based discriminator and a reward-based paradigm to guide the generator's behavior toward generating novel samples while penalizing repetitive outputs.
6. Latent Vector Manipulation models, such as probabilistic diverse GAN (PD-GAN; Liu *et al.* 2021*a*), in image inpainting, focus on enriching diversity by manipulating latent vectors. In image inpainting, PD-GAN adjusts latent vectors to progressively increase diversity in areas that allow for higher variation.

3. Modeling diversity and novelty

In this section, we introduce our evaluation algorithms, as outlined in Sections 3.1 and 3.2. These powerful, automated and objective techniques can serve as robust tools for assessing the diversity and novelty of new design concepts across various industries and contexts given a set of ground truth samples for algorithmic evaluation. Each subsection begins with an extensive review of the respective terms within the context of design literature. We then present the underlying concept behind each algorithm, elucidating its alignment with and adherence to established definitions in the literature. Subsequently, we delve into the implementation details.

3.1. Modeling diversity

Quality properties are typically measured according to two main categories of methods in the design literature, namely qualitative assessment carried out by a human expert and mathematical analysis. The qualitative assessment of diversity involves categorizing a set of design ideas into various ideas based on intuitive categories. A common mathematical approach for diversity analysis is to adopt a genealogical tree for a set of design solutions and to estimate the degree of relatedness among the under-review concept and other instances accordingly (Johnson *et al.* 2016).

In alignment with the established definition of diversity in the design literature, which refers to the extent of dissimilarity among design concepts compared to each other, our GAN diversity assessment approach seeks to capture and quantify the diversity of generated solutions for a given design problem. When employing GANs to produce a batch of design solutions, we conduct a thorough analysis of the (dis)similarities among the generated outputs. By employing equations and mathematical models, we precisely assess the degree of diversity within the batch. Our approach not only adheres to the conceptual understanding of diversity as stated in

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

the literature, but also implements it through a mathematical strategy, allowing for an objective evaluation of the diversity.

Diversity assessment within the context of GANs entails a meticulous evaluation of the dissimilarity inherent in the generated design concepts, adhering to the design literature's notion of diversity. Our methodology for diversity evaluation revolves around the generation of a substantial batch of design samples, facilitating a comprehensive inter-sample comparison. A particularly effective approach to elucidate the diversity of this sample ensemble involves visual representation, providing a tangible depiction of the inherent variations among the data points. The details of the algorithm are outlined in Algorithm 1. It utilizes the visualization option for validation, which offers greater informativeness, and employs the score option for integration with neural networks, including GANs.

Algorithm 1. Diversity evaluation algorithm.

Input: *Set of images I, Desired dimension size DimSize*
Output: *Diversity score or Diversity map*
procedure DIVERSITYEVALUATION(I, DimSize)
Feature extraction:
$\Phi(I) \leftarrow ExtractSemanticFeatures(I)$ ▷ Extract semantic features of each image *Dimensionality reduction:*
$\Psi(I) \leftarrow MapToLowerDimension(\Phi(I), DimSize)$ ▷ *Map features to a lower-dimensional space*

Diversity computation:
if $DimSize = 2$ *or* $DimSize = 3$ ***then***
$V \leftarrow Visualize(\Psi(I))$ ▷ *Use visualization technique on the resultant set of 2/3-dimensional points*
$S \leftarrow ComputeDiversityScore(\Psi(I))$ ▷ *Use diversity detection on the feature vectors*
end if
return *S or V*
end procedure

Visualizing this diversity within a two-dimensional space offers a pragmatic solution that aligns with the cognitive mechanisms of the human perception system. Such a visualization strategy enables an enhanced discernment of the intricate relationships, patterns and distinctions encapsulated within the dataset. Among the various techniques available, principal component analysis (PCA) is our method of choice for projecting high-dimensional data points onto a two-dimensional plane. This preference stems from PCA's capability to retain crucial information and structural attributes that might otherwise be compromised in the transformation process. To facilitate the visualization of diversity, a pivotal step involves the transformation of design concepts from their raw image format into a more structured feature format. This conversion ensures the preservation of the most pertinent information that governs the diversity inherent within the concepts generated. To this end, the adoption of a neural network becomes imperative, given its capacity to discern intricate patterns and features within complex image data. In our methodology, we employ the VGG16 neural network to extract these features.

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

Feature extraction. VGG16 model was initially proposed by Simonyan & Zisserman (2014) for image classification and object detection, which is also a powerful feature extraction and image coding CNN model. Therefore, we use VGG16 to embed our dataset before feeding it to PCA. VGG16 is a 16-layer deep neural network model that contains stacked convolutional layers using the smallest possible receptive field of 3×3 that can have a sense of up/down, left/right and center notions. An optional linear transformation layer of the input channel can be added to the top of the network in the form of a 1×1 convolution filter. Among the 13 convolutional layers, 5 are followed by max-pooling layers to implement spatial pooling with a pooling window of size 2×2 and a stride of size 2. The convolution stride is set to 1, but the padding is specified according to the receptive field to preserve the spatial resolution. The convolutional layers are then followed by three fully connected layers, with the first two layers containing 4,096 each, and the last one depending on the number of classes. The topmost output layer is a softmax layer. Layers do not usually contain normalization to avoid high memory consumption and time complexity, as well as to preserve model performance.

Dimensionality reduction. The paper employs PCA as a statistical method to reduce the dimensionality of complex, high-dimensional data extracted from an image set. PCA is used to disentangle the most influential features within the dataset, allowing for the assessment of diversity among generated samples. By projecting data onto a set of orthogonal variables known as principal components, PCA aims to reveal the underlying structure of the dataset. The principal components are linear combinations of the original variables and are derived to maximize variance. The calculation involves determining weight vectors, with the first vector being computed to maximize the variance of the dataset. Subsequent components are orthogonal to the previous ones and maximize the remaining variance. The number of principal components computed depends on the data structure and the desired level of dimensionality reduction. For evaluating diversity, a two-dimensional representation of the samples is considered most informative for visual analysis, enabling comparisons between the exploration areas of the original and generated datasets within the design space.

3.2. Modeling novelty

For novelty evaluation, a natural and convenient approach is to assess the similarity of a design instance to existing concepts either by human judges through developing mental connections between various knowledge sets to score dissimilarities or using predefined rules based on design attributes. This is also the fundamental approach taken by some of the existing novelty assessment work based on the FBS (Function–Behavior–Structure) and SAPPhIRE (State-Action-Part-Phenomenon-Input-oRgan-Effect) models (Sarkar & Chakrabarti 2011), which assess novelty through comparison with previous design (Jagtap 2019). The qualitative assessment of both diversity and novelty, despite being more accurate, is hard to explain and depends on the rater's mental models. On the other hand, algorithmic assessment suffers from a lack of sensitivity and generalizability, as it is relatively more repeatable and objective (Ahmed 2019).

The definition of novelty in the design literature entails various methods to assess novelty, including the "a priori" and "a posteriori" approaches (Shah *et al.* 2003). The former requires identifying a reference solution or a set of solutions to

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

determine the novelty of the examined ideas, whereas the latter calculates novelty based on a specific framework with respect to existing systems. Leveraging this conceptual understanding, our GAN novelty assessment approach is meticulously aligned with the mathematical approach suggested in the literature. Drawing inspiration from the "a priori" approach, we thoroughly compare each GAN-generated design concept with an extensive repository of previous solutions pertaining to the same design problem. This comparison is facilitated by employing a set of mathematical models that enable us to ascertain the novelty value of each design solution based on its unexpectedness within the generated design space, providing a reliable and comprehensive assessment of its novelty in relation to existing design solutions. The details of our proposed novelty detection algorithm are provided in Algorithm 2.

Algorithm 2. Novelty evaluation algorithm.

Input:
Dataset of existing images: D
New image: N
Output:
Novelty score for the new image: $\mathrm{Nov}(N)$
procedure *NoveltyEvaluation*(D,N)
$\Phi(D) \leftarrow ExtractSemanticFeatures(D)$ ▷ *Extract semantic features of existing images*
$\Phi(N) \leftarrow ExtractSemanticFeatures(N)$ ▷ *Extract semantic features of the new image*
$S \leftarrow Array(Zeros(Size(D)))$ ▷ *Initialize similarity scores array*
for $i = 1$ **to** $\mathrm{Size}(D)$ ***do***
$S[i] \leftarrow Similarity(\Phi(N), \Phi(D[i]))$ ▷ *Calculate similarity between the new image and each instance in the dataset*
end for
$i^* \leftarrow argmax(S)$ ▷ *Identify the most similar instance*
$Sim\_normalized \leftarrow Normalize(S[i^*])$ ▷ *Normalize the similarity score*
$Nov(N) \leftarrow 1 - Sim\_normalized$ ▷ *Calculate the novelty score*
return $\mathrm{Nov}(N)$ ▷ *Output the novelty score*
end procedure

Information extraction and object localization. To gain more nuanced insights into design concepts, detailed information extraction plays a vital role. This aspect allows for a comprehensive analysis of design objects, enabling a deeper understanding of their unique characteristics and attributes. Furthermore, novelty in design concepts is not solely determined by their overall appearance, but also by the arrangement and localization of individual design elements. Evaluating the flexibility of object localization ensures that GANs can identify occurrences of design templates regardless of orientation and local brightness. Hence, we apply the template matching algorithm to search for similar areas of a template image (original images) in a source image (generated images), called a training image. Template matching utilizes a sliding-window approach to compare different areas of the template with the source. The comparison method depends on the content of the images and the goal (Basulto-Lantsova *et al.* 2020).

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

The most frequently used similarity scoring methods for this technique include squared difference, cross-correlation and cosine coefficient, as well as their normalized versions, which usually provide more accurate results. After testing the normalized versions of the three methods, the normalized cross-correlation is selected as the best, as it yields slightly more similar matching results. The matching process creates a two-dimensional result matrix R with similarity scores associated with each area of the image, and searches for the highest/lowest value depending on the comparison method. Template matching can be used to identify the most similar part or determine the location of that part. In this article, however, we apply template matching to find the generated image that is most similar to the source image from a set of real images. This method is simple to implement and computationally efficient.

It is worth highlighting that the measures mentioned earlier do come with certain limitations that may not encompass all crucial facets of novelty. Additionally, they might necessitate extensive preprocessing steps, such as background removal, for optimal performance. However, the robustness of our diversity and novelty measurement algorithms lies in their independence from specific techniques and methods. For instance, the similarity detection technique employed in novelty assessment or the feature extraction method integrated to diversity assessment can be substituted with more advanced alternatives in accordance with contemporary state-of-the-art methodologies. Despite these adaptations, the fundamental capabilities of these algorithms, grounded in existing literature and standard definitions, remain understudied. Such algorithms can be adapted for diversity and novelty measurement across diverse settings and applications. This adaptability is facilitated by the ability to replace algorithmic components with domain-specific counterparts, such as the integration of customized feature extraction tools tailored within the required context.

4. DCG-GAN architecture

In this section, we outline the architecture of DCG-GAN. We first introduce the concept of a versatile, multi-criteria and domain-agnostic GAN architecture. Subsequently, this architecture is tailored to the specific requirements of DCG.

The original GAN architecture (Goodfellow *et al.* 2014) is often metaphorically described as a minimax game between an "art forger" (i.e., the generator) and an "art inspector" (i.e., the discriminator), in which the forger's job is to learn to create fake art that looks "realistic" enough to fool the inspector, while the inspector's job is to accurately differentiate between real and fake samples. Although this architecture has led to groundbreaking advances in generative modeling, it inherently limits the generative process to generating samples that are merely "realistic." As a consequence, the data generated by GANs can be biased and lack the creativity needed to explore new and interesting solutions. Furthermore, GANs may struggle when it comes to generating data with a diverse representation of structurally different designs. To address the lack of diversity, researchers have proposed techniques such as conditional GANs, which allow users to control the generated data to ensure a more diverse and inclusive dataset. However, this is not an automatic procedure and would require a field expert to manually check the entire dataset distribution.

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

As discussed in Section 2.1, these shortcomings stem from the evaluation method in GANs (i.e., discriminator feedback) that focuses solely on generating realistic outputs. As a result, other evaluation metrics that account for different sets of performance must be embedded within the model, so that the generated images are not only realistic but also qualify for other aspects. An intelligent way to achieve this is to replace the discriminating network with an evaluation panel of multiple benchmarks, one of which is the degree of being realistic. The benchmarks are not necessarily global and domain-agnostic; instead, they can be chosen with respect to domain-specific requirements, which may not necessarily exhibit inherent correlation or alignment. Such an architectural configuration is particularly advantageous for applications characterized by multifaceted objectives, accommodating diverse and non-aligned requirements within specific domains. An example of such application is generating design concepts, where metrics such as diversity, novelty, desirability and compatibility with geometric constraints (e.g., silhouettes) are often of the utmost importance. These objectives can be integrated to the GAN structure using a large set of hand-evaluated concepts or using arithmetic methods to analyze the generated samples. To avoid the necessity of annotating a large dataset for each product category and benefit from user feedback on new design concepts, we can mathematically model these terms and integrate them into the GAN architecture.

Figure 1 shows a schematic illustration of a new GAN architecture with a "panel of evaluators" rather than a single discriminator, aimed at addressing the limitations of the traditional GAN architecture for DCG. The generator is trained on the basis of the feedback from all evaluators combined, while the evaluators' weights are adjusted according to their own loss functions. Given this, each evaluator is optimized to score the generator's output with respect to a single criterion for which it was designed; however, the generator is optimized to satisfy all these criteria. Yuan *et al.* (2023) recently implemented a basic version of the proposed architecture for the early stage design of desirable concepts by embedding a user-centered design evaluation tool into the GAN architecture.

The final goal of this study is to adapt our suggested generic framework to cater to the needs of visual design recommendation. This involves incorporating design-related specifications and limitations into the evaluation network. Our investigation highlights diversity, novelty and desirability as essential requirements in visual design, along with the need for compatibility with a given silhouette as a constraint. Accordingly, we regularize the loss function by introducing four supplementary terms, each designed to assess generated samples against a specific criterion. The subsequent sections will provide detailed insights into the modeling and integration process for each of these criteria.

4.1. Novelty: LOF loss

While the integration of our novelty evaluation algorithm with the DCG-GAN remains feasible, we have opted for the application of an anomaly detection technique on the extracted semantic features for enhanced efficiency. This decision aligns with the objective of optimizing computational resources while maintaining the efficacy of the feature analysis process. Repetition of the time-consuming comparison process with the entire dataset for each iteration of the training would considerably slow down the training process. To address the requirement for

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

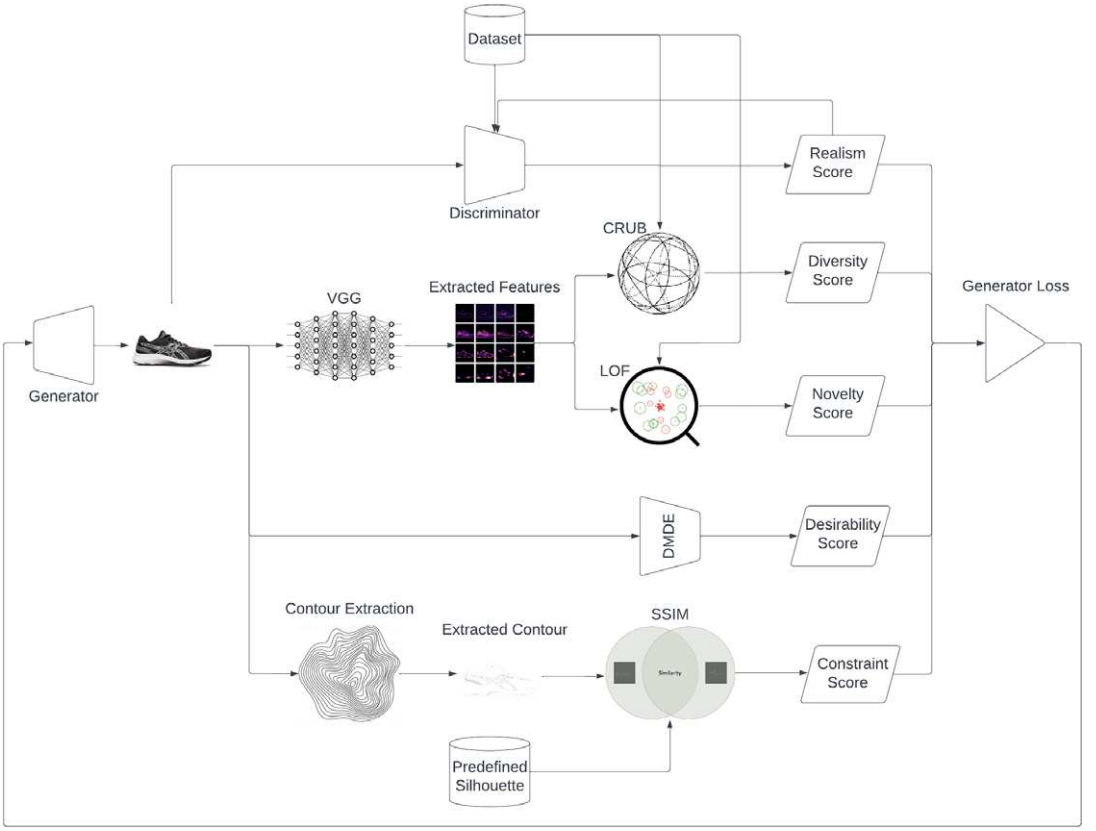

Figure 1. Proposed DCG-GAN architecture for design concept generation: Instead of merely inspecting how realistic the generated samples are using a discriminator network, a "panel of evaluators" must be created to simultaneously assess each generated samples with respect to multiple conditions and criteria.

novelty assurance in the generated design concepts, we incorporated the concept of local outlier factor (LOF; Breunig *et al.* 2000) serving as a density-based anomaly detection technique, designed to assess the distinctiveness of a generated sample in relation to the input dataset. In our methodology, the LOF model is initially trained to utilize the original dataset. During each iteration, features are extracted from the generated samples. These features are then evaluated for novelty relative to the original dataset employing the pre-trained LOF model. This method identifies potential anomalies by evaluating the local deviation of a given data point concerning its neighboring points. LOF operates by comparing the local density of a data point with the densities of its k nearest neighbors. If the density of the point is considerably lower than that of its neighbors, it is deemed an outlier.

In the LOF algorithm, the core idea revolves around measuring the typical distance at which a point can be reached from its neighbors, known as the reachability distance. This measurement involves calculating the reachability distance between two objects, ensuring that it does not fall below the k distance of the second object, as defined by

$$\mathrm{RD}\left(X_{gen},X_i\right) = \max\left\{distance\left(\left(X_{gen},X_i\right),k-distance\left(X_i\right)\right\}. \quad (1)$$

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

The local reachability distance (LRD) of a point is then defined as the inverse of the average reachability distance from its neighbors, calculated using

$$\mathrm{LRD}\left(X_{gen}\right) = \frac{1}{\frac{\Sigma_{X_i \in N_k} RD\left(X_{gen}, X_i\right)}{N_k}}, \tag{2}$$

where N_k represents the set of k neighbors of the generated sample X_{gen}. Having the LRD defined, the LOF score is formulated as the ratio of the average LRD of neighboring points to the LRD of the generated sample itself, expressed as

$$\mathrm{LOF}\left(X_{gen}\right) = \left(\frac{\Sigma_{X_i \in N_k} LRD(X_i)}{N_k}\right)\left(\frac{1}{LRD\left(X_{gen}\right)}\right). \tag{3}$$

LOF values below 1 generally indicate inliers, representing data points within denser regions, while values significantly greater than 1 signal outliers, indicating points that are distinct from their neighbors. The utilization of LOF as part of our evaluation framework provides a mechanism to effectively capture and quantify the novelty of generated design concepts.

4.2. Diversity: CRUB loss and stratified noise vector sampling

To foster diversity within the generative process, we introduce a twofold approach aimed at guiding the generator toward more varied outcomes. The first facet of our technique centers on the diversification of latent codes, which serves as the foundation for sample generation. In contrast to generating an equal number of latent codes as the batch size, we initiate a pool of random codes, from which we selectively extract the most diverse ones, corresponding in quantity to the batch size. This strategy incentivizes the generator to explore uncharted regions of the latent space. To execute this diversified sampling from the pool, we adopt a stratified sampling method.

Stratified sampling is particularly suited to our task, wherein the latent space is partitioned into distinct hypercubes. This methodology ensures that each subgroup contributes proportionately to the overall diversity, avoiding potential biases that might arise with alternative sampling techniques. The partitioned space facilitates the allocation of each latent code to a unique subgroup, from which we draw an equal number of points. This step enhances both the sampling process and the subsequent promotion of diversity. This technique was utilized in both the training and inference phases of the experiments to ensure consistency in the diversity of samples generated by the model. This consistency helps stabilize the training process and ensures that the diversity observed during training is reflected in the inference phase.

The second facet of our approach incorporates an evaluator into the loss function. Within this loss framework, we gauge the diversity of the batch, resulting in a single score that extends to all samples within that batch, thus guiding the calculation of the loss function. For assessing diversity, we follow the procedure explained in Algorithm 1. As explained in Section 3.1, we utilize the diversity score option of the algorithm for loss function integration purpose for which CRUB method is adapted. Due to the efficiency of CRUB, we can use the original extracted feature vector without the need for dimensionality reduction. The CRUB is

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

determined by computing the maximal distance between any Voronoi vertex and its nearest neighbor in the set of points. The covering radius of a generated point set $X_{gen} = \left\{\vec{X}_1, \ldots, \vec{X}_N\right\} \subset S$ is calculated as follows:

$$CR\left(X_{gen}, S\right) = \sup_{\vec{x} \in S}\left\{ \min_{1 \le i \le N} \left\| \vec{x} - \vec{x}_i \right\| \right\}. \tag{4}$$

Then, the upper bound of CR is obtained by $\overline{CR}\left(X_{gen}, S\right) := \max\left\{CR\left(\vec{x}_i, S_i\right) i = 1, \ldots, N\right\}$. Here, S_i represents the i^{th} strata where $\cup_{i=1}^{N} S_i = S$ and $x_i \in S_i$ represents the sample point within those strata.

CRUB plays a pivotal role in global optimization by bounding the worst-case error approximation of the global optimum. A distinctive attribute of CRUB lies in its ability to maximize the shortest distance between all samples within the batch, rather than relying on the average distance that encompasses all pair-wise distances. This approach effectively encourages the dispersion of all samples away from one another, thus contributing to greater diversity among the generated outcomes.

4.3. Geometrical constraint: SSIM loss

Through the authors' consultations with a prominent sneaker design firm, it was determined that product designers are often tasked with generating concepts that maintain visual consistency with their respective product lines, adhering to a defined silhouette. Most consumer product brands have specific design elements such as shapes and logos that are important intellectual property assets. To be useful, it is important that these very specific features be captured during model training. As such, specialized loss components become increasingly valuable. Adherence to a specific silhouette is a critical aspect in various design domains, including logo design, fashion design, product packaging, character design for animation and games, icon design for applications and websites, architectural design, vehicle design and stencil design (Giannini & Monti 2002; Hyun *et al.* 2015; Huang & Hsu 2019; Chujitarom & Panichruttiwong 2020). This specific loss term is applicable across these domains and any other areas where silhouette conformity is a key requirement. Silhouette is a two-dimensional black image that illustrates the body shape of a group of products, leaving out the details. Thus, the appearential geometry of the concepts must be preserved with respect to a contour provided in the DCG process. This objective does not seem attainable using traditional GAN architecture, since (1) GANs generate samples based on a noise vector from the latent space allowing us no control over the features of the final concept, and (2) we aim to enlarge the input dataset for diversity enhancement, resulting in hundreds of silhouettes existing in the dataset. Consequently, a generated concept is likely to be of a non-target contour or a combination of them.

We propose a novel technique to quantitatively assess the extent of preservation of geometric constraints within the context of DCG. Preserving apparent geometry with respect to a provided contour is imperative to maintain the outer shape consistency of products within a designated product line. In this approach, silhouettes serve as reference contours to preserve constraints. The integration process follows a structured sequence. Within each iteration, for every generated

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

image in the batch, we start by extracting the image's contours. These contours are then compared to the corresponding contours of a designated set of silhouettes. This comparison yields a set of similarity scores, each indicating how closely the generated image's contours match those of the silhouettes. Importantly, we utilize the highest similarity score divided by other similarity scores to serve as the geometrical constraint score in the broader loss function. This choice is guided by the core principle that a generated sample should bear a noticeable similarity to only one of the predefined silhouettes.

To overcome these challenges, we introduce a dedicated framework for evaluating and preserving geometrical constraints within the GAN-generated concepts. In each training iteration and for every generated image in the batch, the contours of the generated images are extracted. These extracted contours are then compared with the contours derived from a set of benchmark silhouettes. The comparison process involves the utilization of the SSIM metric described earlier, which is well suited for our purpose because of its ability to quantify the compatibility of structural changes between images. SSIM's emphasis on the overall image structure is in harmony with our aim of extracting detail-free body shapes from the generated concepts. Moreover, its incorporation of perceptual phenomena, such as luminance masking and contrast masking, enhances its capability to precisely identify structural disparities between images. SSIM's focus on interdependencies among neighboring pixels effectively captures vital information about object structure, particularly in terms of luminance masking's influence on dissimilarities in bright regions and contrast masking's ability to detect differences in textured or active areas.

In light of recent advancements, the ControlNet (Zhang, Rao & Agrawala 2023) model, with its constraint mechanism currently applied to diffusion models, presents a potential for integration into GAN frameworks. Exploring this integration represents a promising avenue for future research to further enhance GAN capabilities.

4.4. Desirability: DMDE loss

Following the philosophy of our proposed generic framework, we also integrate the DMDE model into DCG-GAN architecture as a desirability evaluator. Drawing from the favorable outcomes achieved by the DDE-GAN (Yuan *et al.* 2023) model in assessing user satisfaction with generated samples, this inclusion strengthens the DCG-GAN's ability to ensure desirability in the produced visual designs. For a GAN model to generate unique and desirable design concepts, it must be equipped with a design evaluator so that the model is trained not only according to the realism of the outputs but also according to their desirability. Deep multimodal design evaluation (DMDE) is an AI-driven model created by the authors (Yuan, Marion & Moghaddam 2022) to provide an estimate of the desirability of a design concept without having to release the product and aggregate market results. DMDE is capable of performing design evaluation at the general level, attribute level or both in any field that requires the provision of images, textual descriptions and user reviews. The training workflow on this platform consists of four main parts: attribute-level sentiment analysis, image feature extraction, description feature extraction and multimodal predictor. First, attribute-level sentiment intensities are extracted from online reviews, which serve as ground truth for training.

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

Subsequently, visual and textual features are simultaneously extracted using a fine-tuned ResNet50 model and a fine-tuned BERT language model, respectively. Finally, the extracted features are processed by a multimodal deep regression model to predict desirability. The DMDE model is one example of an AI-driven tool that can help guide the GAN generator toward creating samples that are user-centered and desirable. Designers or mathematical methods can use DMDE to predict the performance of a new design concept from the perspective of end users simply by feeding the renderings and descriptions to the model. This platform eases the process of evaluating design concepts, which is one of the most challenging tasks in the design of competitive design concepts. In our research, we specifically utilized the image mode of DMDE due to its relevance in visual concept generation. However, it must be noted that, as demonstrated in the paper (Yuan *et al.* 2022), the single image mode yields a lower prediction accuracy rate of 76.54%, in contrast to the multimodal version's 99.14%. This disparity underscores the potential benefits of integrating the multimodal version in our future work to enhance accuracy.

4.5. Integrated DCG-GAN architecture

In each iteration, the outputs generated by the aforementioned evaluators yield vectors of uniform size, aligned with the batch size (i.e., $B \times 1$). Each row within these vectors corresponds to a distinct sample within the batch. The outputs of the five evaluators, including the discriminator, undergo normalization using min–max normalization techniques. Subsequently, the weighted summation of these normalized vectors is computed, preserving the order of samples (or rows), which is then backpropagated through the generative network to adjust the generator's parameters. As static algorithms such as LOF and SSIM are employed, except for the discriminator, the other evaluators do not necessitate updates via specific loss functions during the training process. The augmented loss function is formulated as follows:

$$\begin{aligned}\mathscr{L}_{\mathrm{DCG-GAN}} = {} & \mathbb{E}_{z\sim p_{gen}}[D(z)] - \mathbb{E}_{x\sim p_{real}}[D(x)] \\ & + \lambda_{GP}\mathbb{E}_{\hat{x}\sim p_{\hat{x}}}\left[\left(\nabla_{\hat{x}}D(\hat{x})-1\right)^2\right] + \lambda_{\mathrm{LOF}}\mathbb{E}_{z\sim p_{gen}}[\mathscr{L}_{\mathrm{LOF}}(z,s)] \\ & + \lambda_{\mathrm{CRUB}}\mathbb{E}_{z\sim p_{gen}}\left[\mathscr{L}_{\mathrm{CRUB}}\left(z,p_{real}\right)\right] + \lambda_{\mathrm{SSIM}}\mathbb{E}_{z\sim p_{gen}}[\mathscr{L}_{\mathrm{SSIM}}(z)] \\ & + \lambda_{\mathrm{DMDE}}\mathbb{E}_{z\sim p_{gen}}[\mathscr{L}_{\mathrm{DMDE}}(z)],\end{aligned} \tag{5}$$

where s represents the silhouette. The associated loss functions are defined as follows:

$$\mathscr{L}_{\mathrm{LOF}} = \frac{1}{N}\Sigma_{g=1}^{N}\left(LOF\left(z_g,p_{real}\right) - \max(LOF)\right)^2, \tag{6}$$

$$\mathscr{L}_{\mathrm{CRUB}} = \frac{1}{N}\Sigma_{g=1}^{N}\left(CRUB\left(z_g,s\right) - \max(CRUB)\right)^2, \tag{7}$$

$$\mathscr{L}_{\mathrm{SSIM}} = \frac{1}{N}\Sigma_{g=1}^{N}\left(SSIM\left(z_g,s\right) - \max(SSIM)\right)^2, \tag{8}$$

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

$$\mathscr{L}_{\mathrm{DMDE}} = \frac{1}{N}\Sigma_{g=1}^{N}\left(DMDE(z_g) - \max(DMDE)\right)^2. \quad (9)$$

Overall, our proposed framework serves as a bespoke platform for automated design concept recommendation, ensuring the quality of generated samples in terms of realism, outer shape geometry, novelty, diversity and desirability.

5. Experiments

In this section, we present a case study conducted on a large-scale dataset extracted from multiple online footwear stores, along with details of the implementation and results. Subsequently, we present and analyze the results in depth to benchmark DCG-GAN against Style-GAN2 in terms of diversity and novelty.

5.1. Dataset and training

To evaluate and validate the performance of DCG-GAN versus Style-GAN2 in generating novel and diverse concepts, a large-scale dataset comprising 6,745 images was scraped from a major online footwear store. The images included were side view of several brands of footwear such as Adidas, ASICS, Converse, Crocs, Champion, FILA, PUMA, Lactose, New Balance, Nike and Reebok to avoid mode collapse and increase the diversity of the dataset. The neural network models were trained using the Pytorch implementation of Style-GAN2 from scratch. The training process employed four Tesla V100-SXM2 GPUs, PyTorch version 1.8, and Python version 3.7. The configuration settings remained consistent throughout, including a latent code represented by both z and w of dimension 512, the use of 8 fully connected layers in the mapping network, activation functions employing leaky ReLU with a slope parameter $\alpha = 0.2$, bilinear filtering in all up/down-sampling layers, equalized learning rates for all trainable parameters, incorporation of a minibatch standard deviation layer at the conclusion of the discriminator, implementation of an exponential moving average for generator weights, utilization of a non-saturating logistic loss function with R1 regularization, and Adam optimizer with specific hyperparameters: $\beta_1 = 0.5$, $\beta_2 = 0.9$, $\epsilon = 10^{-8}$ and a minibatch size of 64.

5.2. Results

The low-novelty samples from both DCG-GAN and Style-GAN2 presented in Figures 2 and 3, respectively, bear a strong resemblance to established shoe models and brands. Specifically, the depicted concepts closely mirror existing Adidas,

Figure 2. High novelty examples of Style-GAN2 generated concepts.

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

Figure 3. Low novelty examples of Style-GAN2 generated concepts.

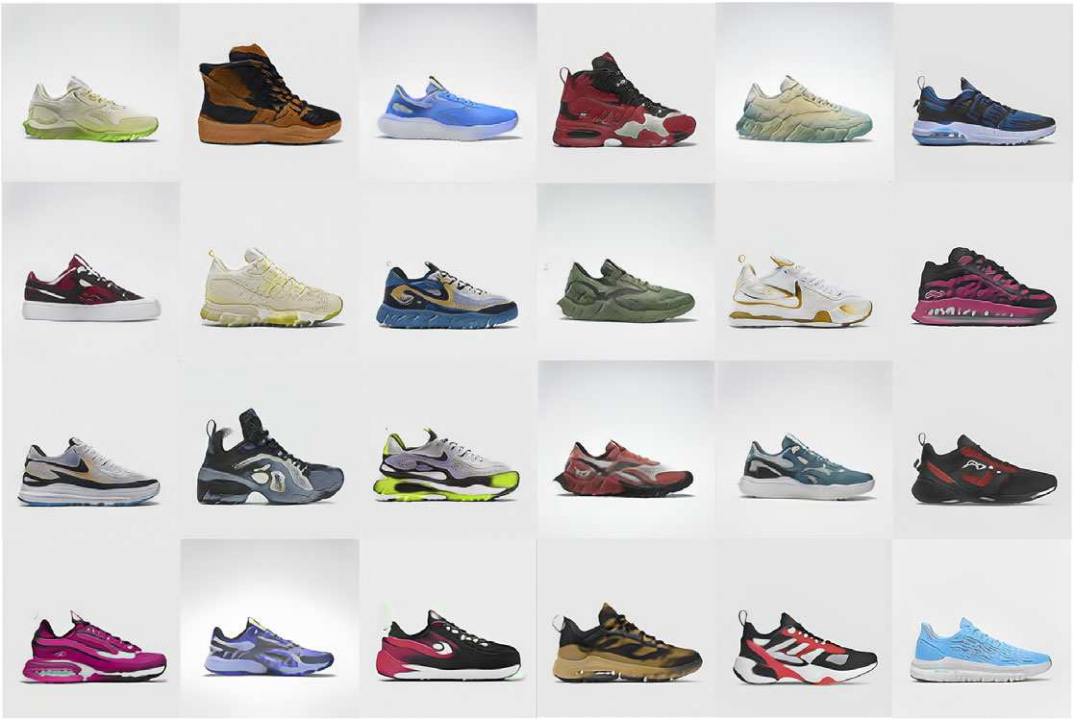

Figure 4. High-novelty examples of generated concepts using DCG-GAN.

Asics, Nike and Reebok shoe designs available on the market. This outcome aligns with the inherent limitation of GAN generators, which tend to replicate patterns and characteristics present in the training dataset. For further insights into the mechanism and outcomes of StyleGAN, refer to (Ghasemi *et al.* 2023).

Figures 4 and 5 illustrate several high-novelty and the most novel examples of design concepts generated by DCG-GAN, respectively, demonstrating various visually discernible features:

- *Novel features.* The DCG-GAN architecture excels in generating design concepts imbued with abundant novel attributes. This stark contrast becomes more pronounced compared to the concepts generated from the baseline. Although the latter typically mirror existing products from the training set, lacking distinctive and unique features, the DCG-GAN consistently introduces innovative elements.

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

Figure 5. Low-novelty examples of generated concepts using DCG-GAN.

Figure 6. Most novel examples of generated concepts using DCG-GAN.

- *Higher diversity*. The efficacy of DCG-GAN in fostering diversity is evident through the wide-ranging spectrum of generated designs, which span different styles, patterns and structural arrangements.
- *Esthetic appeal*. The generated design concepts boast captivating forms and harmonious color palettes. The model's ability to produce visually appealing outcomes is indicative of its capacity to capture intricate design esthetics.
- *Minimal brand-specific features*. Intriguingly, a notable observation is the limited inclusion of logos within the generated sneaker concepts. The model leans toward designs that prioritize other design elements over logo placement.
- *High visual quality*. Despite reducing the discriminator's weight to 20% in the loss function (with all evaluators sharing equal weights), the samples generated in DCG-GAN demonstrate remarkable quality. This outcome is particularly surprising, as a diminished discriminator weight might suggest a compromise in realism; however, the generated designs retain high-quality characteristics. In fact, the DCG-GAN's generated samples exhibit a pronounced reduction in unrealistic appearances (e.g., lacking recognizable attributes characteristic of sneakers) compared to the baseline model.
- *Compatibility with geometrical constraints*. The generated design concepts align with a predetermined set of silhouettes, serving as a benchmark for geometrical constraints.

A quantitative comparison of novelty distributions for the illustrated concepts in Figures 2–6 shows the corresponding template-matching confidence scores (Figure 7). A noteworthy finding arising from the juxtaposition of high-novelty instances generated by the DCG-GAN and Style-GAN2 models is the discernible trend wherein DCG-GAN examples, despite exhibiting comparable levels of novelty as measured by template matching scores, manifest a higher degree of visual realism, that holds true even for the most unrealistic instances (three right most concepts in Figure 5).

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

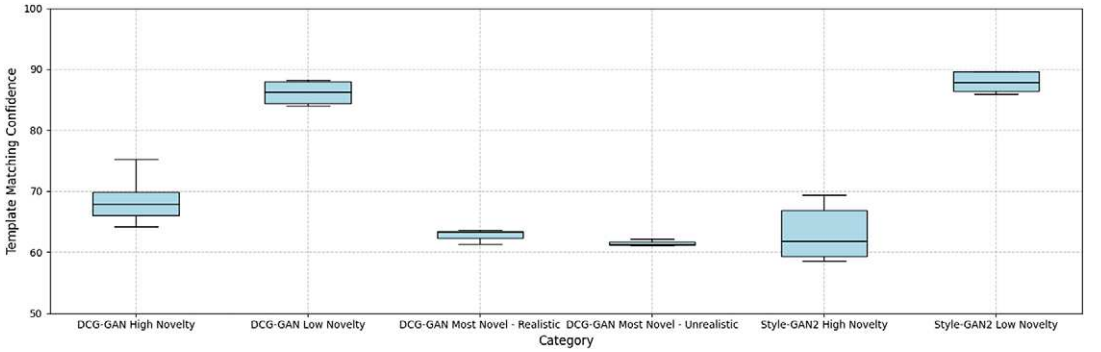

Figure 7. Comparison of novelty distribution for generated concepts with different degrees of novelty.

5.3. Quantitative validation

We used the methods presented in Sections 3.1 and 3.2 to quantitatively assess the performance of DGC-GAN and the baseline GAN in generating diverse and novel samples, respectively.

5.3.1. Diversity assessment

We generated two sets of 1,500 design concept images using DCG-GAN and the baseline model. To ensure diversity and avoid redundancy, we introduced a random seed as a numerical element by creating two separate sets of 1,500 random variables, each of which was individually injected into the model. To evaluate the diversity of the generated samples, we initiated the process by extracting essential visual attributes from the images using VGG16. This process produced a vector of 1,000 features for each image. Given the complexity of interpreting a 1,000-dimensional space, we then applied PCA to these vectors, reducing their dimensionality to a 2D space. This transformation enabled a more accessible comparison of the solution space coverage achieved by each model. For the feature extraction step, we utilized the VGG16 model due to its capability to embed images while excluding the top output layer. To ensure that the model could better identify design-specific features rather than general ones from broader datasets like ImageNet, we initially trained it on our combined dataset, consisting of both the original and generated images. The VGG16 model was trained on RGB images with dimensions of 224×224 and included three fully connected layers at the top of the network, without pooling layers, and employed a softmax activation function.

In Figure 8, the 2D representation of the mapped data points is presented, where the red points represent the original data set, and the green points represent the generated data set. This visualization illustrates that the Style-GAN2 model (and other traditional GAN architectures) did not explore the entirety of the original data space, resulting in designs that are constrained to a specific and incomplete range of styles. Notably, the green points cover only a subset of the space occupied by the red points, highlighting a limitation in GANs' ability to learn the complete distribution of the dataset. Additionally, the scatter plot reveals the model's inadequacy, especially in regions where there are fewer original samples, indicating that the model's capacity to learn a subspace is reliant on the presence of an adequate number of data samples from that specific subspace. These findings

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

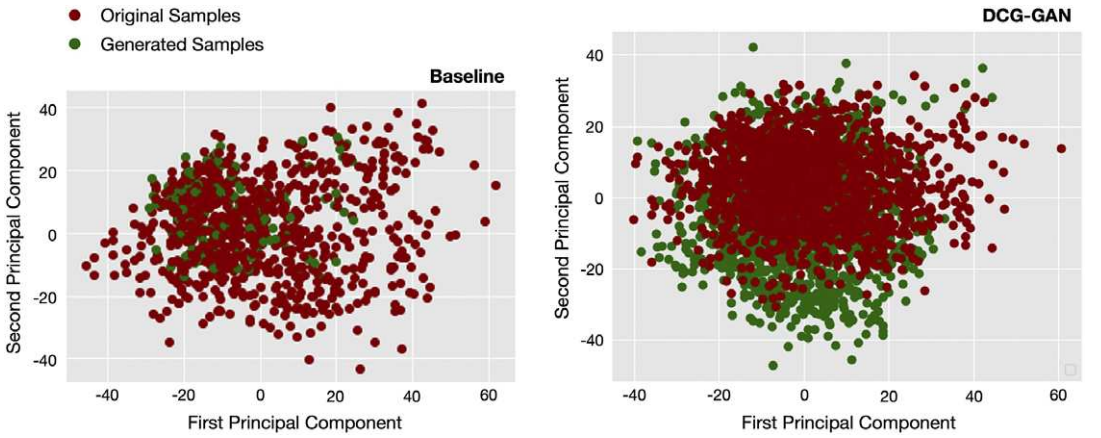

Figure 8. Covered areas of the design space by the original (red) and generated (green) samples. Left: The baseline (Style-GAN2) results. Right: The DCG-GAN's results.

provide partial validation of the central hypothesis concerning the limited diversity of GAN-generated design samples.

As depicted in Figure 8, the visualization results indicate that the images generated by DCG-GAN possess the capacity to traverse uncharted areas within the solution space. Remarkably, this exploration is not only superior to the baseline model's performance, but also extends beyond the confines of the original dataset. The DCG-GAN model effectively expands the boundaries of the solution space that the original dataset occupies from several directions. Notably, an intriguing observation demonstrates that, unlike the baseline model, DCG-GAN adeptly learns and captures areas within the solution space that lacked adequate representation within the original dataset.

5.3.2. Novelty assessment

To quantify the novelty of GAN-generated sets, we employed our novelty evaluation algorithm, incorporating template matching. The process involved identifying the most similar design instance from the original dataset for each generated sample within each generated set. Subsequently, these results were consolidated into a similarity distribution function, enabling a statistical assessment of novelty in the generated outputs. The template matching method operates by searching a source image to identify areas that closely resemble a provided template image. Confidence scores for different areas within the template image are computed and compared using a sliding window approach. The algorithm gauges the similarity between the source image and the template image by considering the confidence within the rectangle whose dimensions match those of the source image and whose top-left corner corresponds to a specific point. To ensure that template matching aligns with our objectives, we utilized source (i.e., created) and template (i.e., original) images of identical dimensions, ensuring equal weighting for all sections of the images in the confidence score calculation. The confidence score in this context operates within a numerical range of 0 to 1, where a score of 0 indicates an absence of similarity, while a score of 1 signifies complete identity.

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

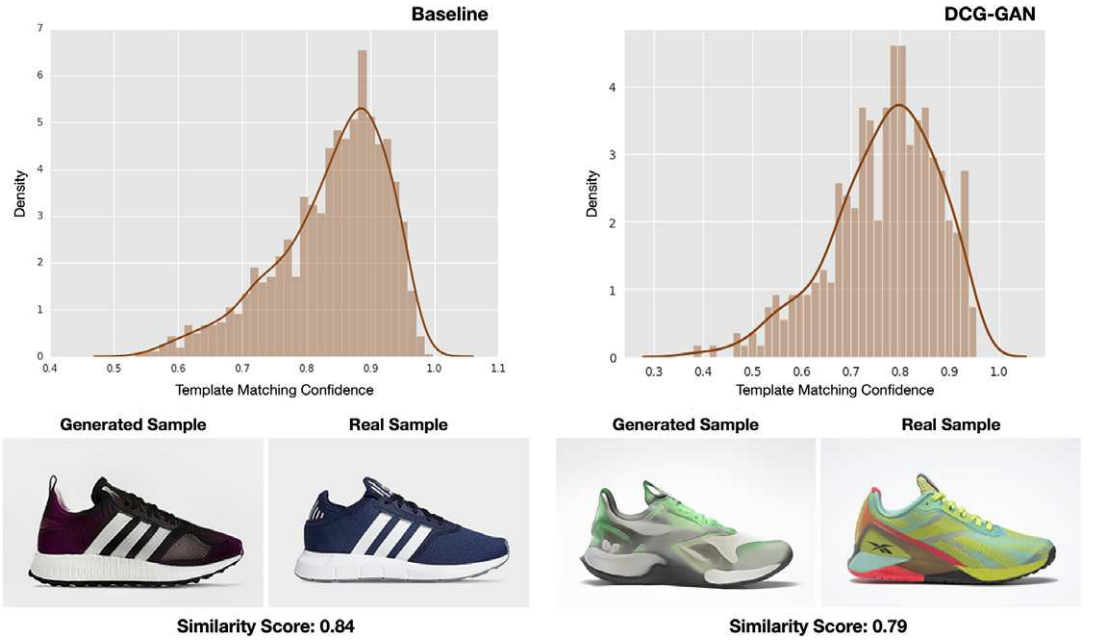

Figure 9. Top: Distribution function and correlating semi-Gaussian function of the template matching confidence scores based on the generated-real comparisons. The left and right side figures represent the baseline and DCG-GAN, respectively. Bottom: Example of a generated sample and the most similar real sample from the training dataset.

Figure 9 presents a distribution graph accompanied by a semi-Gaussian function, which has been fitted with a mean of 0.8385 and a variance of 0.0075. This distribution suggests that the majority of the design samples generated by Style-GAN2 bear a very close resemblance to those found in the original dataset. This outcome aligns with our central hypothesis, affirming that GANs have limitations in generating novel design concepts. However, it is important to note the presence of samples with relatively low confidence scores, which can be attributed to two factors: (1) Some generated images exhibit an unrealistic nature that they cannot be readily recognized as sneakers. (2) Template matching tends to differentiate between sneakers with the same pattern but different colors, resulting in lower similarity scores. Figure 9 illustrates the most similar pair from the original dataset for a generated image, and it is visually apparent that the generated image closely resembles an item already present in the original dataset, devoid of any discernibly novel attributes. To comprehensively assess the novelty of traditional GAN models, we additionally computed the distribution function for the confidence scores pertaining to the generated samples.

The evaluation results reveal DCG-GAN's proficiency in generating design concepts with increased novelty. The dissimilarity between a sample generated by the DCG-GAN model and its closest counterpart from the original dataset is prominently evident, emphasizing DCG-GAN's ability to generate design concepts that deviate significantly from existing instances. This contrast in similarity is supported by the confidence scores, with the DCG-GAN pair exhibiting a significantly lower similarity score (0.79) compared to the baseline pair (0.84),

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

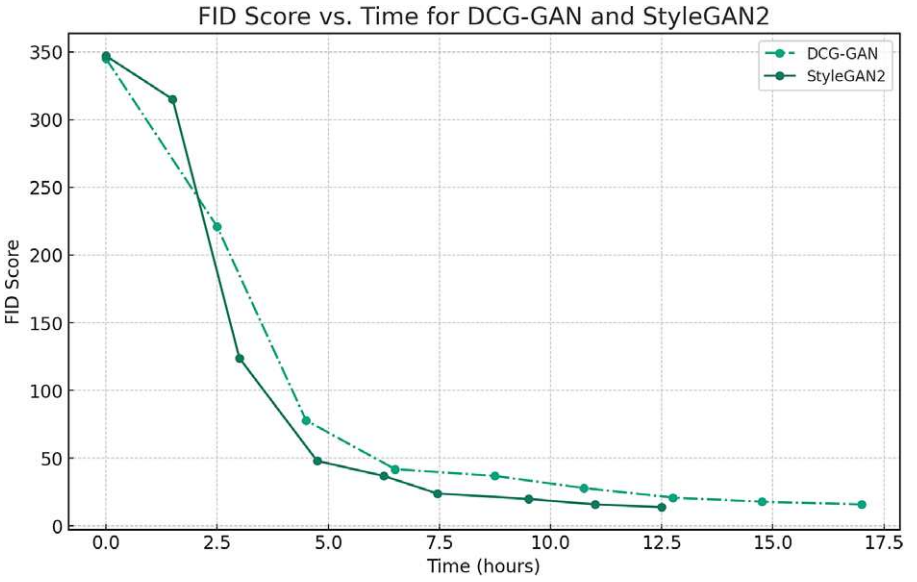

Figure 10. FID as a function of wall-clock time for DCG-GAN versus the baseline.

reinforcing the enhanced novelty of DCG-GAN-generated concepts. The distribution of confidence scores depicted in Figure 2 further underscores these findings. Notably, this distribution shows a considerable leftward shift of approximately 10% compared to the baseline, indicative of a substantial 10% enhancement in novelty for the median, most novel and least novel samples. The mean and variance of the DCG-GAN distribution stand at 0.7680 and 0.0116, respectively. This mean value substantiates an average improvement of 7% in terms of novelty, while the reduced variance indicates a higher prevalence of generated concepts by DCG-GAN that possess novel features.

In summary, the above findings indicate that there is a lack of novelty in the GAN-generated samples according to different perspectives toward similarity evaluation and different levels of feature spaces, which validates our central hypothesis.

5.3.3. Quality and computational efficiency assessment

The DCG-GAN model demonstrates substantial improvement in image generation quality as measured by the Fréchet inception distance (FID) scores (Figure 10), decreasing from 345 to 16 over the training period of 0 to 17 hours. This metric indicates a significant enhancement in the model's ability to produce high-quality images that closely resemble the distribution of real images. Similar improvements are observed with the Style-GAN2, with its FID score reducing from 347 to 14, suggesting that both models achieve comparable advancements in image quality over time. For a more detailed analysis of the quality metrics and the integration of the DMDE model with GAN, interested readers are encouraged to refer to Yuan *et al.* (2023). This article delves into the nuances of combining DMDE with GAN architecture.

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

Regarding computational efficiency, despite the inclusion of four additional loss function components in the DCG-GAN, which could potentially increase computational demand, the model maintains a competitive level of computational efficiency. The average processing time for DCG-GAN is noted to be 177 seconds per tick, compared to Style-GAN2's 108 seconds per tick, indicating that the enhanced functionality of DCG-GAN does not disproportionately impact its performance efficiency.

5.4. Qualitative validation

To comprehensively evaluate the performance of the DCG-GAN model, an additional layer of qualitative assessment was introduced through a quantitative survey. This survey aimed to garner human insights into the novelty and diversity of the design concepts produced by the DCG-GAN model, contrasting them with those generated by the baseline model. The summative evaluation process centered on the rating of two distinct sets, each comprising 20 sneaker design concepts. Within these sets, 10 designs were generated by the DCG-GAN model, while the remaining 10 were generated by the baseline model. Participants were instructed to provide ratings for each concept's novelty and diversity using a scale ranging from 0 to 10, where 0 represented "no novelty/diversity," 5 indicated a "neutral" viewpoint and 10 signified "high novelty/diversity." Importantly, participants were kept unaware of the generative model linked to each concept.

To ensure a diverse array of perspectives and minimize potential biases, the survey was extended to individuals from diverse professional and demographic backgrounds. The pool of participants included sneaker designers, designers specializing in other diverse applications, engineering students, and individuals with limited familiarity with design, engineering and AI models. A total of 89 individuals actively participated in the survey. The survey comprised three main sections:

1. Demographic questions: Participants responded to inquiries about their age group, gender, highest level of education, occupation, ethnicity/race and familiarity with generative AI models.
2. Novelty assessment: Following the presentation of both academic and simplified definitions of novelty, participants were tasked with individually rating the novelty of the 20 randomly ordered concepts, generated by the DCG-GAN and baseline models.
3. Diversity assessment: Similar to the novelty assessment, participants were provided with a definition of diversity before rating two sets of concepts. One set contained concepts generated by DCG-GAN, while the other featured concepts generated by the baseline model. Importantly, participants were kept unaware of the models tied to each set.

Quantitative outcomes depicted in Figure 11 show the minimum, first quartile, median, third quartile and maximum values, as well as the average derived from the participant ratings. These values represented the novelty assessment for each individual design concept and the diversity assessment for each set of concepts. The survey results on novelty indicated an average novelty assessment of 5.5257 for DCG-GAN, marking a 15% improvement compared to the baseline average score of 4.0757. Notably, the minimum novelty score for DCG-GAN surpassed the

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

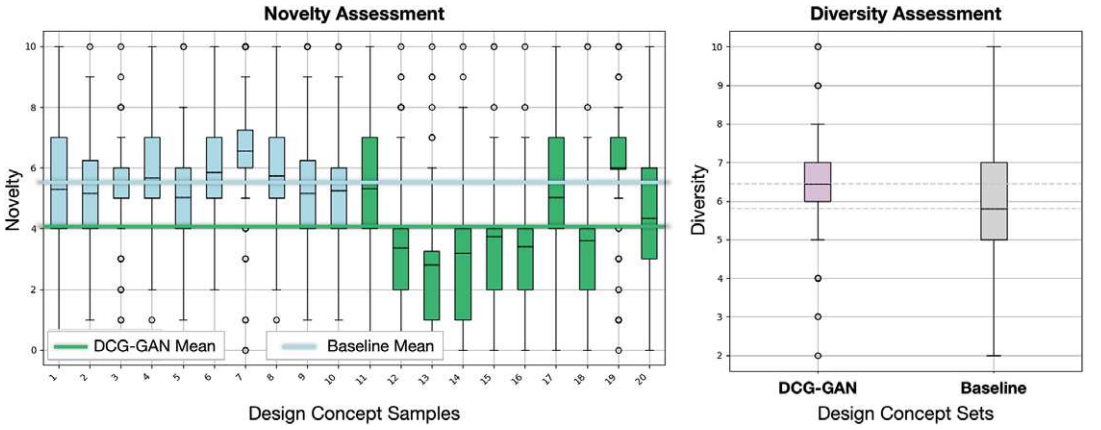

Figure 11. Qualitative assessment results of the blinded experiments. Left: Novelty ratings pertaining to 20 generated samples by DCG-GAN (1–10) and the baseline (11–20). Right: Diversity ratings for two sets of samples generated by DCG-GAN and the baseline.

minimum for the baseline. Additionally, the standard deviation for DCG-GAN (0.4583) was considerably reduced compared to the baseline (1.0395), which signifies enhanced consistency and suggests that not a subset of DCG-GAN generated samples but all exhibited increased novelty. In the diversity assessment, DCG-GAN achieved an average score of 6.4492, indicating a 7% improvement over the baseline average score of 5.7971. A hypothesis test conducted on the results revealed significant differences. The p-value for novelty was less than 0.001, indicating a statistically significant improvement in novelty with the DCG-GAN model compared to the baseline. Similarly, the p-value for diversity was 0.03, also suggesting a statistically significant enhancement in diversity with the DCG-GAN model. These survey results consistently affirm the findings from both objective and visual analyses, collectively showcasing the effectiveness of the DCG-GAN model in generating design concepts that are not only visually appealing but also novel and diverse.

6. Discussion and conclusions

There are several key limitations that GANs face in the context of DCG and generative design. First, GAN variants necessitate large training datasets, and when dealing with diverse design images, the generator's ability to capture various modes may suffer, particularly when the dataset is insufficient, leading to overlooked modes with fewer data. Second, mode collapse, a common issue, arises during training, where the generator tends to produce a narrow set of samples associated with a limited distribution subset, especially problematic in high-dimensional inputs like images and videos. The generator's lack of reward for diversification during training exacerbates this issue, causing over-optimization on a single output. Third, GANs can struggle with diversity, novelty and desirability, as their objective encourages mimicry of input data, potentially leading to an overemulative generator. Pushing for greater diversity and creativity may compromise sample quality and utility. Fourth, evaluating GAN performance remains challenging, with a lack of standardized methods to assess generated versus real

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

distributions regarding different criteria. Among widely recognized evaluation metrics (e.g., image quality, stable training and image diversity) in the literature (Borji 2019), the focus primarily centers on image quality and training stability, leaving room for improvement in other evaluation criteria such as image diversity. In this work, we discuss the limitations inherent in traditional GAN architectures, with a particular focus on Style-GAN2 chosen for quantitative validation. It is important to note that the findings derived from this analysis are applicable to all traditional GAN models, as they share a common evaluation framework.

In the context of image generation, GANs present certain challenges that are addressed more effectively by other generative models such as variational auto-encoders (VAEs), diffusion models and transformers. A notable limitation of GANs is their susceptibility to mode collapse and training instability, which can lead to a lack of diversity in generated images and difficulties in model convergence. Additionally, GANs often require careful tuning of hyperparameters to achieve balance between the generator and discriminator (Dhariwal & Nichol 2021), a complexity less prominent in the more stable training dynamics of VAEs and Diffusion models. Unlike Transformer-based models, GANs might also struggle with capturing long-range dependencies in data, potentially limiting their effectiveness in certain contexts. Despite these challenges, GANs offer distinct advantages. They are particularly celebrated for their ability to generate images of high fidelity and remarkable realism, outperforming VAEs, which tend to produce blurrier images due to their reconstruction loss function. GANs also have an edge in computational efficiency, especially when compared to Diffusion models that require iterative refinement processes and thus longer generation times, enhancing their usefulness in industry application. The GPU demands of generative models are extremely high, leading to high costs of cloud-based services to train and run the models and excessive usage of electricity to power these data centers. As such, the efficiency of GANs makes them attractive for real-time design optimization in real-world use cases. Moreover, GANs demonstrate a comparative advantage over diffusion models concerning the propensity for overfitting. Diffusion models are notably susceptible to memorizing training images, especially when the original dataset is limited in size (Akbar *et al.* 2024). GANs also exhibit a broader spectrum of applicability compared to VAEs, particularly in scenarios involving non-sequential tasks or imbalanced datasets. Notably, GANs possess the capability to balance diversity against fidelity. This trade-off allows for the production of high-quality samples, though it may result in not fully capturing the entire distribution spectrum (Dhariwal & Nichol 2021). GANs typically encounter challenges in applications that require general-domain adaptability (Wang *et al.* 2023). For DCG, the necessity shifts toward a model that is specialized and informed by domain expertise. This specialization entails a model adept at generating designs for specific products or items, without the imperative for broad generalizability. The focal requirement for such a model is to exhibit both diversity and novelty, albeit within a narrowly defined context. This characteristic underscores the superiority of GANs in scenarios where robustness against overfitting is critical. The array of characteristics and limitations inherent to various models led us to choose GANs as the backbone model for customization to cater to multi-objective scenarios with a specific context.

In this study, we conducted an extensive and systematic examination of GANs as a potential tool for DCG, with a specific focus on assessing creativity in terms of

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

novelty and diversity. Our approach began with the mathematical modeling of diversity and novelty, as defined within the design literature, addressing the fourth limitation of GANs. Subsequently, we applied these models to evaluate the output of traditional GAN architectures, revealing a deficiency in creativity. To overcome this limitation, we introduced a novel and versatile multi-criteria GAN architecture, adaptable to various domains. Specifically, for the generation of design concepts, we extended this architecture to encompass four additional criteria alongside realism: diversity, novelty, desirability and geometrical proportionality. To facilitate the assessment of these criteria, we devised dedicated evaluation algorithms and recommended appropriate implementation techniques for efficient training. Through visual inspection and both quantitative and qualitative assessments, our model demonstrated a significant enhancement in terms of diversity and novelty. This outcome confirmed it as a valuable tool for DCG. Furthermore, our approach addressed the initial challenges of GANs, demonstrating its capacity to explore design spaces with limited real samples, comprehensively explore the real dataset distribution and produce outputs that excel across multiple criteria. Lastly, this research helps advance the transition of emerging technologies into useful tools for the designer. While it is unclear the exact form these technologies will ultimately take (Moghaddam *et al.* 2023), this study could enable large numbers of novel and diverse concepts to be presented to the human designer as well as fast concept evaluation frameworks in terms of diversity and novelty, leveraging the speed and efficiency of computer-generated design knowledge while maintaining the critical eye and decision making of the human. This augmented approach may enable the "best of both worlds," with the ultimate outcomes being significantly improved design efficiency and quality.

Although DCG-GAN demonstrates substantial advances in visual DCG, it has some key limitations that serve as opportunities for future research in this field. First, DCG-GAN focuses primarily on visual esthetics and does not consider the functional aspects of design concepts. Incorporating functionality criteria into the generative process would enable the production of design concepts that are not only visually appealing but also practical and functional. Second, DCG-GAN is currently limited to generating 2D design concepts. Future work could extend the model's capabilities to encompass 3D design generation, broadening its applicability across various design domains. Third, DCG-GAN involves a homogeneous real dataset, requiring clear backgrounds, aligned objects and consistent picture angles. Future research can focus on approaches that can improve the resolution of datasets in domains where data is scarce or difficult to obtain. To this end, a mixture model sampled from a chosen Gaussian can be developed to reparameterize the latent generative space, which can be particularly useful when the dataset is small and limited. Lastly, DCG-GAN operates as a non-interactive model, generating numerous novel concepts randomly within a short timeframe. Future research could explore ways to introduce interactivity into the generative process, allowing users to guide and fine-tune the output according to specific requirements and preferences, thereby increasing its practical utility. In this regard, a layer-wise decomposition approach can also be used to identify potential control spaces and manipulate images from high-level properties in either the latent or feature space. The focus on shoe design as the sole case study in this article represents another limitation due to constraints in data collection, annotation and domain-specific expertise. Future research will aim to expand the application of our framework

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

across diverse design domains, addressing the challenges associated with acquiring and annotating relevant datasets and incorporating expert knowledge on domain-specific constraints. Moreover, ablation studies are required to better demonstrate and assess the decoupled effects of different components of the model. Additionally, in this work, DCG-GAN is designed to balance multiple objectives in DCG, focusing on novelty, diversity, desirability and geometrical integrity through the integration of various loss terms. While our approach prioritizes a comprehensive multi-objective framework, we recognize the importance of analyzing and potentially minimizing overlaps between these metrics. Future research can delve into optimizing these interplays, enhancing the model's efficiency while preserving its capability to meet diverse design criteria. These limitations offer opportunities for further progress in this domain.

Financial support

This material is based on work supported by the National Science Foundation under the Engineering Design and System Engineering grant #2050052. Any opinions, findings or conclusions expressed in this material are those of the authors and do not reflect the views of the NSF.

References

Ahmed, F. 2019 *Diversity and novelty: Measurement, learning, optimization*. PhD dissertation, University of Maryland.

Akbar, M. U., **Larsson, M.**, **Blystad, I.** & **Eklund, A.** 2024 Brain tumor segmentation using synthetic MR images – A comparison of GANs and diffusion models. *Scientific Data* **11**, 259.

Basulto-Lantsova, A., **Padilla-Medina, J. A.**, **Perez-Pinal, F. J.** & **Barranco-Gutierrez, A. I.** 2020 Performance comparative of OpenCV Template Matching method on Jetson TX2 and Jetson Nano developer kits. In 2020 *10th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 812–816, doi:10.1109/CCWC47524.2020.9031166.

Bloch, P. H. 1995 Seeking the ideal form: Product design and consumer response. *Journal of Marketing* **59**, 16–29.

Borji, A. 2019 Pros and cons of Gan evaluation measures. *Computer Vision and Image Understanding* **179**, 41–65.

Breunig, M. M., **Kriegel, H.-P.**, **Ng, R. T.** & **Sander, J.** 2000 LOF: Identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data (SIGMOD '00)*, pp. 93–104. Association for Computing Machinery; doi:10.1145/342009.335388.

Buonamici, F., **Carfagni, M.**, **Furferi, R.**, **Volpe, Y.** & **Governi, L.** 2020 Generative design: An explorative study. *Journal of Computer-Aided Design and Applications* **18**, 144–155.

Burnap, A., **Liu, Y.**, **Pan, Y.**, **Lee, H.**, **Gonzalez, R.** & **Papalambros, P. Y.** 2016 Estimating and exploring the product form design space using deep generative models. In *Proceedings of the ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. Volume 2A: 42nd Design Automation Conference*. V02AT03A013. ASME; doi:10.1115/DETC2016-60091.

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

Carlini, N., **Hayes, J.**, **Nasr, M.**, **Jagielski, M.**, **Sehwag, V.**, **Tramèr, F.**, **Balle, B.**, **Ippolito, D.** & **Wallace, E.** 2023 Extracting training data from diffusion models. In *32nd USENIX Security Symposium*, Article 294, pp. 5253–5270. USENIX Association.

Chen, W. & **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**, p. V11AT11A010). American Society of Mechanical Engineers.

Chujitarom, W. & **Panichruttiwong, C.** 2020 Animation - AR Silhouette model to create character identity innovation for computer art pre-production. *TEM Journal* **9**, 1740–1745.

Dhariwal, P. & **Nichol, A.** 2021 Diffusion models beat GANs on image synthesis. In *Advances in Neural Information Processing Systems*, Article 672, pp. 8780–8794. Curran Associates Inc.

Dieter, G. E., **Schmidt, L. C.** & **Azarm, S.** 2009 *Engineering Design*, Vol. **4**. McGraw-Hill Higher Education.

Dow, S. P., **Glassco, A.**, **Kass, J.**, **Schwarz, M.**, **Schwartz, D. L.** & **Klemmer, S. R.** 2010 Parallel prototyping leads to better design results, more divergence, and increased self-efficacy. *ACM Transactions on Computer-Human Interaction (TOCHI)* **17** (4), 1–24.

Dubinski, J., **Deja, K.**, **Wenzel, S.**, **Rokita, P.** & **Trzcinski, T.** 2022 Diversity balancing generative adversarial networks for fast simulation of the Zero Degree Calorimeter in the ALICE experiment at CERN. In *Machine Learning and the Physical Sciences Workshop*.

Elgammal, A., **Liu, B.**, **Elhoseiny, M.** & **Mazzone, M.** 2017 CAN: Creative adversarial networks, generating "Art" by learning about styles and deviating from style norms. In *International Conference on Computational Creativity (ICCC)*, arXiv:1706.07068.

Gaut, B. 2010 The philosophy of creativity. *Philosophy Compass* **5** (12), 1034–1046.

Ghasemi, P., **Yuan, C.**, **Marion, T.** & **Moghaddam, M.** 2023 Are generative adversarial networks capable of generating novel and diverse design concepts? An experimental analysis of performance. *Proceedings of the Design Society* **3**, 633–644.

Giannini, F. & **Monti, M.** 2002 An innovative approach to the aesthetic design. In Durling, D. and Shackleton, J. (eds.), *Common Ground—DRS International Conference 2002*, 5–7 September, London, United Kingdom. https://dl.designresearchsociety.org/drs-conference-papers/drs2002/researchpapers/29.

Goodfellow, I., **Pouget-Abadie, J.**, **Mirza, M.**, **Xu, B.**, **Warde-Farley, D.**, **Ozair, S.**, **Courville, A.** & **Bengio, Y.** 2014 Generative adversarial nets. In *Proceedings of the 27th International Conference on Neural Information Processing Systems—Volume 2 (NIPS'14)*, pp. 2672–2680. MIT Press.

Gui, Y., **Zhou, B.**, **Xie, X.**, **Li, W.** & **Zhou, X.** 2021 GAN-based method for generative design of visual comfort in underground space. In *IOP Conference Series: Earth and Environmental Science*; doi:10.1088/1755-1315/861/7/072015.

Guzdial, M. & **Riedl, M.** 2019 Combinets: Creativity via recombination of neural networks. In *Proceedings of the 10th International Conference on Computational Creativity*, ArXiv 1802.03605.

Heyrani Nobari, A., **Rashad, M. F.** & **Ahmed, F.** 2021 CreativeGAN: Editing generative adversarial networks for creative design synthesis. In *Proceedings of the ASME 2021 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. Volume 3A: 47th Design Automation Conference (DAC)*. V03AT03A002. ASME; doi:10.1115/DETC2021-68103.

Hsiao, K. A. & **Chen, L. L.** 2006 Fundamental dimensions of affective responses to product shapes. *International Journal of Industrial Ergonomics* **36**, 553–564.

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

Huang, Y.-C. & **Hsu, C.-C.** 2019 Network virtual reality clothing silhouette design influencing factors. In *IEEE 4th International Conference on Signal and Image Processing*, pp. 707–711, doi:10.1109/SIPROCESS.2019.8868778.

Hyun, K. H., **Lee, J. H.**, **Kim, M.** & **Cho, S.** 2015 Style synthesis and analysis of car designs for style quantification based on product appearance similarities. *Advanced Engineering Informatics* **29**, 483–494.

Jagtap, S. 2019 Design creativity: Refined method for novelty assessment. *International Journal of Design Creativity and Innovation* **7** (1–2), 99–115.

Jiang, Z., **Wen, H.**, **Han, F.**, **Tang, Y.** & **Xiong, Y.** 2022 Data-driven generative design for mass customization: A case study. *Advanced Engineering Informatics* **54**, 101786–101786.

Johnson, T. A., **Cheeley, A.**, **Caldwell, B. W.** & **Green, M. G.** 2016 Comparison and extension of novelty metrics for problem-solving tasks. In *Proceedings of the ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. Volume 7: 28th International Conference on Design Theory and Methodology*. V007T06A012. ASME; doi:10.1115/DETC2016-60319.

Karras, T., **Laine, S.** & **Aila, T.** 2019 A style-based generator architecture for generative adversarial networks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4396–4405, doi:10.1109/CVPR.2019.00453.

Karras, T., **Laine, S.**, **Aittala, M.** & **Hellsten, J.** 2020 Analyzing and improving the image quality of stylegan. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8107–8116, doi:10.1109/CVPR42600.2020.00813.

Kim, S. & **Lee, S.** 2023 Self-supervised augmentation of quality data based on classification-reinforced GAN. In *International Conference on Ubiquitous Information Management and Communication (IMCOM)*, pp. 1–7, doi:10.1109/IMCOM56909.2023.10035575.

Landwehr, J. R., **Labroo, A. A.** & **Herrmann, A.** 2011 Gut liking for the ordinary: Incorporating design fluency improves automobile sales forecasts. *Marketing Science* **30**, 389–564.

Landwehr, J. R., **Labroo, A. A.** & **Herrmann, A.** 2013 Product design for the long run: Consumer response to typical and atypical designs at different stages of exposure. *Journal of Marketing* **77**, 92–107.

Li, X., **Su, J.**, **Zhang, Z.** & **Bai, R.** 2021 Product innovation concept generation based on deep learning and Kansei engineering. *Journal of Engineering Design* **32**, 559–589.

Linsey, J. S., **Tseng, I.**, **Fu, K.**, **Cagan, J.**, **Wood, K. L.** & **Schunn, C.** 2010 A study of design fixation, its mitigation and perception in engineering design faculty. *Journal of Mechanical Design* **132**(4), 041003.

Liu, H., **Wan, Z.**, **Huang, W.**, **Song, Y.**, **Han, X.** & **Liao, J.** 2021a PD-GAN: Probabilistic diverse GAN for image inpainting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9371–9381.

Liu, R., **Ge, Y.**, **Choi, C. L.**, **Wang, X.** & **Li, H.** 2021b DivCo: Diverse conditional image synthesis via contrastive generative adversarial network. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16372–16381, doi:10.1109/CVPR46437.2021.01611.

Liu, Y., **Li, K. J.**, **Chen, H.** (A.) & **Balachander, S.** 2017 The effects of product design on demand and marketing effectiveness: An empirical analysis. *Journal of Marketing* **81**, 83–102.

Macedo, L. & **Cardoso, A.** 2002 Assessing creativity: The importance of unexpected novelty. *Structure* **1**. https://api.semanticscholar.org/CorpusID:15674655.

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

Mao, Q, **Lee, H.-Y.**, **Tseng, H.-Y.**, **Ma, S.** & **Yang, M-H**. 2019 Mode seeking generative adversarial networks for diverse image synthesis. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1429–1437.

Marion, T. J. & **Fixson, S.** 2018 *The Innovation Navigator: Transforming Your Organization in the Era of Digital Design and Collaborative Culture*. University of Toronto Press.

Meng, Y. & **Xu, L.** 2020 Diversity augmented conditional generative adversarial network for enhanced multimodal image-to-image translation. In *International Conference on Learning Representations*.

Moghaddam, M., **Marion, T.**, **Holtta-Otto, K.**, **Fu, K.**, **Olechowski, A.** & **McComb, C.** 2023 Emerging technologies and methods for early-stage product design and development. *Journal of Mechanical Design* **145** (4), 040301.

Norman, D. 2002 Emotion design: Attractive things work better. *Interactions* **9**, 36–42.

Norman, D. A. 2013 *The Design of Everyday Things*. The MIT Press.

Oh, S., **Jung, Y.**, **Kim, S.**, **Lee, I.** & **Kang N.** 2019 Deep generative design: Integration of topology optimization and generative models. *Journal of Mechanical Design* **141** (11), 111405.

Osborn, A. F. 1953 *Applied Imagination*. Scribner.

Qian, C., **Tan, R. K.** & **Ye, W.** 2022a An adaptive artificial neural network-based generative design method for layout designs. *International Journal of Heat and Mass Transfer* **184**, 122313.

Qian, C., **Tan, R. K.** & **Ye, W.** 2022b Design of architectured composite materials with an efficient, adaptive artificial neural network-based generative design method. *Acta Materialia* **225**, 117548.

Roozenburg, N. F. M. & **Eekels, J**. 1995 *Product Design: Fundamentals and Methods*. Wiley.

Sarkar, P. & **Chakrabarti, A.** 2011 Assessing design creativity. *Design Studies* **32** (4), 348–383.

Shah, J. J., **Smith, S. M.** & **Vargas-Hernandez, N.** 2003 Metrics for measuring ideation effectiveness. *Design Studies* **24**, 111–134.

Shu, D., **Cunningham, J.**, **Stump, G.**, **Miller, S. W.**, **Yukish, M. A.**, **Simpson, T. W.** & **Tucker, C. S.** 2020 3D design using generative adversarial networks and physics-based validation. *Journal of Mechanical Design* **142** (7), 071701.

Simonyan, K. & **Zisserman, A.** 2014 Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning Representations (ICLR 2015)*, pp. 1–14. Computational and Biological Learning Society. arXiv:1409.1556.

Singh, V. & **Gu, N.** 2012 Towards an integrated generative design framework. *Journal of Design Studies* **33**, 185–207.

Siriwardane, E. M. D., **Zhao, Y.**, **Perera, I.** & **Hu, J.** 2022 Generative design of stable semiconductor materials using deep learning and density functional theory. *npj Computational Materials* **8**, 164.

Toh, C. A., **Miller, S. R.** & **Okudan Kremer, G. E.** 2014 The impact of team-based product dissection on design novelty. *Journal of Mechanical Design* **136** 4, 041004.

Ulrich, K., **Eppinger, S.** & **Yang, M. C.** 2020 *Product Design and Development* (7th ed.), McGraw-Hill.

Vasconcelos, L. A., **Cardoso, C. C.**, **Sääksjärvi, M.**, **Chen, C. C.** & **Crilly, N.** 2017 Inspiration and fixation: The influences of example designs and system properties in idea generation. *Journal of Mechanical Design* **139** (3), 031101.

Wang, B., **Zhu, Y.**, **Chen, L.**, **Liu, J.**, **Sun, L.** & **Childs, P.** 2023 A study of the evaluation metrics for generative images containing combinational creativity. *Artificial Intelligence*

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press

for Engineering Design, Analysis and Manufacturing. **37**, e11. doi:10.1017/S0890060423000069.

Wang, P., **Wang, S.**, **Peng, D.**, **Chen, L.**, **Wu, C.**, **Wei, Z.**, **Childs, P.**, **Guo, Y.** & **Li, L.** 2020 Neurocognition-inspired design with machine learning. *Design Science* **6**, e33.

Wang, Z., **She, Q.** & **Ward, T. E.** 2021 Generative adversarial networks in computer vision: A survey and taxonomy. *ACM Computing Surveys (CSUR)* **54** (2), 1–38.

Wu, J.-L. & **Huang, S.** 2022 Application of generative adversarial networks and Shapley algorithm based on easy data augmentation for imbalanced text data. *Applied Sciences* **12** (21), 10964. doi:10.3390/app122110964.

Wu, Q., **Liu, Y.**, **Miao, C.**, **Zhao, B.**, **Zhao, Y.** & **Guan, L.** 2019 PD-GAN: Adversarial learning for personalized diversity-promoting recommendation. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19)*, pp. 3870–3876. AAAI Press.

Xu, J., **Ren, X.**, **Lin, J.** & **Sun, X.** 2018 Diversity promoting GAN: A cross entropy-based generative adversarial network for diversified text generation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 3940–3949. Association for Computational Linguistics.

Yang, D., **Hong, S.**, **Jang, Y.**, **Zhao, T.** & **Lee, H.** 2019 Diversity-sensitive conditional generative adversarial networks. In *International Conference on Learning Representations*.

Yean, S., **Somani, P.**, **Lee, B. S.** & **Oh, H. L.** 2021 GAN+: Data augmentation method using generative adversarial networks and dirichlet for indoor localisation. In *IPIN-WiP*. https://api.semanticscholar.org/CorpusID:247387224.

Yuan, C., **Marion, T.** & **Moghaddam, M.** 2022 Leveraging end-user data for enhanced design concept evaluation: A multimodal deep regression model. *Journal of Mechanical Design* **144** (2), 021403.

Yuan, C., **Marion, T.** & **Moghaddam, M.** 2023 DDE-GAN: Integrating a data-driven design evaluator into generative adversarial networks for desirable and diverse concept generation. *Journal of Mechanical Design* **145**, 041407.

Yuan, C. & **Moghaddam, M.** 2020 Attribute-aware generative design with generative adversarial networks. *IEEE Access* **8**, 190710–190721.

Zhang, L., **Rao, A.** & **Agrawala, M**. 2023 *Adding conditional control to text-to-image diffusion models*, . In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3813–3824. IEEE; doi:10.1109/ICCV51070.2023.00355.

https://doi.org/10.1017/dsj.2024.13 Published online by Cambridge University Press