ARE GENERATIVE ADVERSARIAL NETWORKS CAPABLE OF GENERATING NOVEL AND DIVERSE DESIGN CONCEPTS? AN EXPERIMENTAL ANALYSIS OF PERFORMANCE

Ghasemi, Parisa;
Yuan, Chenxi;
Marion, Tucker;
Moghaddam, Mohsen

Northeastern University

ABSTRACT

Generative Adversarial Networks (GANs) have shown stupendous power in generating realistic images to an extend that human eyes are not capable of recognizing them as synthesized. State-of-the-art GAN models are capable of generating realistic and high-quality images, which promise unprecedented opportunities for generating design concepts. Yet, the preliminary experiments reported in this paper shed light on a fundamental limitation of GANs for generative design: lack of novelty and diversity in generated samples. This article conducts a generative design study on a large-scale sneaker dataset based on StyleGAN, a state-of-the-art GAN architecture, to advance the understanding of the performance of these generative models in generating novel and diverse samples (i.e., sneaker images). The findings reveal that although StyleGAN can generate samples with quality and realism, the generated and style-mixed samples highly resemble the training dataset (i.e., existing sneakers). This article aims to provide future research directions and insights for the engineering design community to further realize the untapped potentials of GANs for generative design.

Keywords: Machine learning, Artificial intelligence, User centred design, Generative design, Generative adversarial networks

Contact:
Moghaddam, Mohsen
Northeastern University
United States of America
mohsen@northeastern.edu

Cite this article: 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', in *Proceedings of the International Conference on Engineering Design (ICED23),* Bordeaux, France, 24-28 July 2023. DOI:10.1017/pds.2023.64

1 INTRODUCTION

Design is a complex process that requires designers to develop a network connection among their knowledge of different areas to establish and define solutions and pertinent structures for problems not solved before, or new solutions to problems that have previously been solved in a different way (Dieter et al., 2009). Exploring the uncharted areas of the design space increases the chances of the success of candidate solutions. However, this exploration can be time-consuming and prone to biased and fixated initial conditions and existing ideas for which we either want to blindly explore the design space uniformly or optimize it according to the requirements. In either case, computational technologies, more specifically AI methods, seem essential for enabling high-speed search and expansion of the solution space. Generative design approaches have been proposed to improve design quality and speed by automating all or parts of the design process using computational power. The advantages of generative design include efficient solution creation given a specific period of time, reduced costs, better optimized and accurate solutions, and consistent design instances. Existing generative design approaches fall into five main categories, namely: shape grammars, L-systems, cellular automata, genetic algorithms, and swarm intelligence (Singh and Gu, 2012). These approaches generally optimize the design generator through mathematical functions or physics-based simulators (Shu et al., 2020). For mechanical design purposes, a powerful generative design tool is Autodesk's Generative Design workspace with a major focus on structural optimization by optimizing a cost function to solve mathematically framed design problems (Buonamici et al., 2020). Yet, there is a lack of studies on data-driven generative design approaches that optimize the design generator by training it on past product and user feedback data.

AI and more specifically machine learning algorithms are a more general approach and do not necessarily fall into one of the five categories mentioned above. Instead, AI is a means of automating any process. The strategy behind each of those categories can theoretically be implemented in an AI framework. For example, a genetic algorithm can be turned into a stochastic AI model. AI-driven generative design can serve as a transformative assistive tool to augment designers' ability to create more innovative and desirable concepts faster due to the efficiency with which it can process vast amounts of product and user feedback data, learn complex patterns, generate novel concepts, and evaluate them based on data. The designer can then choose, synthesize, and modify the concepts generated. Deep generative models have recently been adopted for design automation to improve the performance of designers through co-creation with AI. Specifically, generative adversarial networks (GANs) (Goodfellow et al., 2014) have shown tremendous success in a variety of generative design tasks. GANs can generate images from random noise and do not require detailed information or labels from existing samples to start the generation. GANs have been applied to engineering design generation, such as generating 3D aircraft models in native format for complex simulation (Shu et al., 2020), numerous wheel design options optimized for engineering performance (Oh et al., 2019), and realistic samples from the paired fashion clothing distribution and providing real samples to pair with arbitrary fashion units for style recommendation (Yuan and Moghaddam, 2020).

GANs offer a novel generative modeling architecture that enables a new form of learning algorithms with indirect supervision, where a discriminator network serves as a source of feedback for a generator network tasked with learning the complex distribution of a training dataset. In the sense of game theory, GANs can be viewed as a two-person zero-sum game between the generator and the discriminator, where the gains of each competitor exactly equal the loss of the other competitor. The main focus of GAN models is to generate realistic outputs with the ability to mimic the latent space of the input dataset. As a result, GANs could potentially be of great use for generative design because of their ability to produce visually feasible design concepts. However, generative design was initially proposed to encourage divergent thinking and creativity, as aesthetics and creativity are of great importance in the field of design (Buonamici et al., 2020). Despite the potential of GANs in creating feasible design solutions, it is not yet clear how they can enable creativity, since existing GAN architectures inherently tend to mimic the training dataset with the same statistics without expressing much creativity. The rationale behind the lack of creativity is that during the training process, the GAN generator is encouraged to generate samples close to the training data distribution to fool the discriminator in a minimax game, which inevitably results in limited creativity, especially in terms of diversity and novelty.

https://doi.org/10.1017/pds.2023.64 Published online by Cambridge University Press

1.1 Knowledge gaps

In engineering design, creativity is highly valued as one of the most important elements in evaluating the effects and performance of design tasks. Creativity is often defined as a function of novelty and usefulness (Shah et al., 2003). Usefulness is correlated and measured with the quality of the design; therefore, some studies define novelty as a metric of how unusual or unexpected an idea is compared to other ideas (Shah et al., 2003), and some studies measure the creativity, quantity, quality, and diversity of the generated designs (Toh et al., 2014). Creativity has been seen as the success mostly; however, it is hard to assess and enhance design creativity effectively and efficiently because of its intangible and subjective nature. Currently, a significant amount of research is focused on studying engineering design tools that address these design aspects. In this article, in addition to the common attributes of creativity, quality, and novelty, another term diversity is proposed and used to better represent the creativity of the generative design community (Wang et al., 2021). The authors observed that certain features and contours of the shoe models were more repetitive than others in the generated samples and that the coverage of the solution space depends on the density of input samples from a particular area. By reducing the dimensionality of the design space and preserving the two most contributing features, a 2D design space is defined in which only some areas are covered by the generated samples. Therefore, this paper defines diversity as the extent of coverage of the solution space.

This paper proposes GANs as a design concept recommendation system for designers rather than a means of completely automating the concept generation process. Hence, the goal is to expand the range of the generated samples, allowing for enrichment of the designer's ideation as the consumer of the generated outputs. The objective of a generative design process for early-stage concept generation is not to generate "market-ready" designs, but rather to provide examples that would stimulate a designer's mental faculties, imagination, and creativity. Evidence suggests that more diverse solution spaces can increase the likelihood of producing a successful design instance. There are two main categories of methods in the design literature according to which the diversity of a design space can be measured: subjective rating and genealogical tree approach. As an example of subjective rating of design space diversity, we can categorize a set of design ideas into various idea pools based on intuitive categories. Despite being efficient in terms of required time and effort, the results may not be as valid and reliable, since the inferences are based on the rater's mental models. However, 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, their main shortcomings are the lack of sensitivity and accuracy, since they use the same set of formulae for all types of design problems (Ahmed, 2019). This paper aims to address two critical knowledge gaps concerning the effectiveness of state-of-the-art GANs for generative design:

- **Lack of diversity.** Several recent studies have attempted to improve the diversity of GAN samples. Shmelkov et al. (2018) proposed an image classification-based evaluation method for GANs that considers the diversity of the generated outputs. Xu et al. (2018) developed a cross entropy-based GAN that reinforces diversity in a textual latent space by assigning higher rewards for non-repeated output. Wu et al. (2019) proposed a GAN-based recommendation system that samples recommendations from a determinantal point process kernel matrix of two learnable components associated with evaluating the diversity of samples and generating diverse samples. Liu et al. (2021) proposed a perceptual diversity loss function for GAN models that improves the diversity of the generated contents. The aforementioned developments aim to expand the diversity of generated outputs using GANs; however, to the best of our knowledge, no attempts have been made to analyze the extent of diversity in GANs quantitatively.
- **Lack of novelty.** One of the most challenging tasks in the design process is to evaluate the novelty of concepts and, ideally, to distinguish instances with the highest probability of success. In the design literature, novelty indicates the uniqueness of a design idea compared to other concepts that fall within the same class of design problems. This uniqueness does not have to be from a particular aspect (e.g., appearance); instead, it can stem from the otherness of any characteristic within the concept or the design process. A natural and convenient approach to evaluate the novelty of design instances is to assess their similarity to the existing concepts. When measured by human judges, this evaluation occurs by developing mental connections between various knowledge sets to score dissimilarities, usually resulting in subjective and hard-to-explain decisions. On the contrary, when

https://doi.org/10.1017/pds.2023.64 Published online by Cambridge University Press

mathematical methods are used for novelty evaluation, predefined rules based on design attributes are applied, the main drawback being the lack of generalizability (Ahmed, 2019).

1.2 Objectives and outline

A successful GAN-based concept recommendation system should be able to generate a large set of solutions for the intended design problem (in our case, the exterior form of the sneaker) that are diverse, novel, and desirable. This paper explores the diversity and novelty of GANs by adopting a state-of-the-art GAN architecture, StyleGAN2 (Karras et al., 2020), to generate early-stage design concepts based on a large training dataset of sneakers scraped from multiple online footwear stores. In the development of a new product like sneakers, the design currently begins with 2D sketches of the side view of the shoe. Color, material, sole shape, and other attributes are important aspects. Once the appearance direction is defined, considerations for structure and manufacturing are then determined. That happens later in the design process. Our analysis right now is focused on the very front-end of the process, which begins with visualization of concepts. This is not unlike car design, which begins in the design studio with visual concepts. Once a direction is defined, more detailed conceptual engineering is performed. The preliminary results reported indicate that although the trained generator can generate realistic images of sneakers, the generated samples highly resemble existing products (i.e., the training dataset). The consequence of the generator focusing solely on fooling the discriminator by generating samples that resemble the training dataset is the lack of novelty and diversity, which limits its applicability in generative design. The contributions of this paper are as follows:

1. We conducted design concept generation and style-mixing experiments and conducted results that empirically illuminate the aforementioned knowledge gap.
2. We designed an experiment that quantifies the novelty of generated design concepts using the template matching technique by computing the similarity of generated and training samples. The results indicate a lack of novelty in the generated dataset.
3. We designed an experiment for diversity evaluation by visualizing the diversity of the output samples in a 2D design space by applying PCA on the features extracted from the generated images using VGG16. The results certify the lack of diversity in the generated samples.
4. We also propose several directions for future research and exploration in data-driven generative design.

The remainder of this paper is organized as follows. Section 2 presents the methods for generating samples and mixing them to create more diverse results, the Template Matching method that was used to quantitatively analyze the limitation of GANs in the generation of novel solutions, and the PCA algorithm that was adopted to illustrate the limitation of GANs in terms of diversity. Section 3 discusses and analyzes the results, and Section 4 concludes the paper and provides future research directions.

2 METHODS

This section discusses the overall structure of our proposed method to evaluate the creativity of GAN models by quantifying their diversity and novelty, along with a brief technical overview of the methods used. To this end, we first trained a StyleGAN2 model on our dataset to generate a set of sneakers and style-mix them, resulting in 1750 design concepts. The preliminary results of the experiments, presented in Section 3, demonstrate the lack of diversity and novelty in the generated samples, as the generated and style-mixed concepts resemble one or a combination of the original dataset. However, we expanded our evaluation by adopting techniques that detect the existence of new or different features in a design concept. For the evaluation of diversity, we proposed using PCA on both the original and generated datasets to better visualize and compare the areas of the design space covered by each dataset. VGG16 was also used to extract higher-level features from the images before applying PCA resulting in a more informative mapping of the samples by PCA. For novelty evaluation, we adopted template matching so that each genetic sample is compared with the entire original dataset to find the corresponding most similar original sample according to a similarity score. Moreover, we calculate the distribution function of the similarity scores to show that the generated samples take into account the features of the existing design instances.

2.1 Preliminary experiments: GAN, StyleGAN2, and style-mixing

This section briefly introduces GANs and StyleGAN (Karras et al., 2020), a state-of-the-art GAN architecture that can generate highly realistic samples. StyleGAN is used in this paper as a benchmark to evaluate the limitations of GANs for generative design in terms of diversity and novelty. A standard GAN architecture comprises two neural networks: a generator G and a discriminator D, which are interactively trained by competing against each other in a minimax game. The generator attempts to produce realistic samples, while the discriminator attempts to distinguish the fake samples from the real ones. In the standard GANs model, there is no control over the modes of the data being generated. GANs are notoriously difficult to train and often unstable due to mode collapse, one of the main problems in the generative model (Oh et al., 2019). In this way, it is not a good choice to use this approach for generating realistic designs especially considering the significant developments of GANs, which established a new state-of-the-art in generated images with high-quality and high-resolution. This work builds on a cutting-edge GAN architecture for artificial image generation, called StyleGAN2 (Karras et al., 2020). StyleGAN, created by NVIDIA, produces facial images in high resolution with unprecedented quality and is capable of synthesizing and mixing non-existent photorealistic images (Karras et al., 2019). Style-GAN2 is a variation of StyleGAN with minor quality developments such as removed blob-like artifacts, stabilized high-resolution training, and reduced computational cost. See Karras et al. (2020) for details of the StyleGAN2 model.

2.2 Evaluating the novelty of GAN samples: template matching

Humans often find themselves drawn to the new, be it different items, unknown environments, or sudden modifications and unforeseen outcomes. According to Intepat IP, designs can be considered original or novel when they have not been revealed to the public before. In other words, a design is novel when it is distinct from any existing designs or a merger of multiple designs.

Template matching, which is a similarity detection technique active primarily in the field of computer vision, was exploited to quantitatively assess the similarity of the generated images with the original image set. Template matching is capable of finding similar areas of a template image T (original images in our case) to the objects specified in a source image S (generated images in our work), sometimes referred to as the training image. Template matching exploits the sliding-window approach from top to bottom and from left to right to compare different areas of the template with the source. The comparison method depends on the content of the images and the intention (Basulto-Lantsova et al., 2020). The most commonly used similarity scoring methods for template matching include square difference, cross-correlation, and cosine coefficient, as well as their normalized versions, which generally have more accurate results. We tested the normalized version of all three methods and finally selected the normalized cross-correlation because the matched results were slightly more similar. The mechanism behind this technique is to store similarity scores associated with each area of the image in a two-dimensional result matrix R to find the highest/lowest value depending on the comparison method. Template matching can be used to find the most similar part or the location of the said area. However, we adopted this technique to find the image most similar to the source image from an image set. Template matching is a very simple to implement and computationally efficient method. The matching procedure for one source image and one template image is as follows:

Algorithm 1 Template matching algorithm to compare GAN-generated and original images.

$T \leftarrow$ *template image* (e.g., real images of sneakers)
$S \leftarrow$ *source image* (e.g., StyleGAN-generated images of sneakers)
$R \leftarrow two - dimensional\ matrix\ of\ \ S.width - T.width * S.height - T.height$
for $i = 1, \ldots, S.width - T.width$ **do**
  for $j\ = 1, \ldots, S.height - T.height$ **do**
    $R_{i,j} \leftarrow \mathrm{SUMDIFFS}(T, I, i, i + T.width, j\ , j\ + T.height)$
  end for
end for

https://doi.org/10.1017/pds.2023.64 Published online by Cambridge University Press

2.3 Evaluating the diversity of GAN samples: PCA & VGG16

The diversity of the output samples was evaluated by visualizing two principal components of the sample features using PCA. The features were extracted using the VGG16 model to provide a more informative input space for PCA. VGG16 was initially proposed as an image classification and object detection model that gained 92.7% accuracy on the ImageNet dataset. As a state-of-the-art convolutional neural network (CNN) model, VGG16 is a very powerful model for feature extraction and image coding. Therefore, we used VGG16 for the task of embedding 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 4096 each, and the last one depending on the number of classes. The top-most 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.

PCA (Abdi and Williams, 2010) is a multivariate statistical technique utilized in this paper to reduce the dimensionality of high-dimensional data from the intercorrelated feature space. As the dataset on which we used PCA was a high-dimensional set of dependent features extracted from an image set, using this method to assess the diversity of generated samples is convenient. PCA is used in this paper to analyze and interpret complex data by disentangling the most representative features. This task is carried out by computing values of the data table corresponding to a new set of orthogonal variables; thus, PCA can geometrically be viewed as the projection of the data samples onto the principal components' space. These variables, which are called principal components, are acquired as a linear combination of the original variables. The first principal component is computed so that it has the largest possible variance. The first weight vector, based on which the first principal component is calculated, satisfies the following expression: $w_1 = \arg\max_{\|w\|=1} \Sigma_i (x_i.w)^2 = \arg\max (w^T X^T X w)/(w^T w)$, where x_i is a row vector of the original data table X, and w is a coefficient vector set to be a unit vector. Equivalently, in a closed format, the first component's mapping weight vector w_1 can be calculated using the second part of the equation, where w is the eigenvector of the matrix that results in the largest corresponding eigenvalue. The k^{th} component is obtained under the constraint of being orthogonal to $k-1$ previous components as well as having the k^{th} largest possible variance. Thus, we first subtract $k-1$ previous components from X and then use it as the original matrix in the following equation: $w_k = \arg\max_{\|w\|=1} (\hat{X}_k w)^2 = \arg\max \frac{w^T X^T X w}{w^T w}$, where $\hat{X}_k = X - \Sigma_{j=1}^{k-1} X w_j w_j^T$. The number of principal components calculated depends on the data structure and how much dimension reduction we require. For diversity evaluation, since we need to compare the areas of the design space that are explored by the original and generated datasets, a two-dimensional representation of the samples is the most informative for visual analysis.

3 RESULTS AND DISCUSSIONS

This section elaborates on the implementation of the experiments and the reason for their usage. Then, the results of the experiment are illustrated and analyzed in detail to validate the initial hypothesis that the GAN models lack both diversity and novelty.

3.1 Dataset and training

To test and validate the performance of StyleGAN2 in generating realistic and diverse images, a large-scale dataset was scraped from a major online footwear store to conduct numerical experiments. To avoid mode collapse and increase the diversity of the dataset, several brands of footwear are included in the dataset including Adidas, ASICS, Converse, Crocs, Champion, FILA, PUMA, Lactose, New Balance, Nike, and Reebok. A total of 6745 images were collected and cleaned from an online retail store where footwear images have only two orthographic perspectives: a side view and a 3/4 view. The

https://doi.org/10.1017/pds.2023.64 Published online by Cambridge University Press

neural network models were trained in the Pytorch implementation of StyleGAN2 [1] and performed on 4 Tesla V100-SXM2 GPUs with PyTorch 1.8 and Python 3.7. Most configurations remain unchanged, where the dimensionality of the latent code z and w is 512 and the mapping network architecture is 8 fully connected layers. For the style-based generator, leaky ReLU activation was used with $\alpha = 0.2$, bilinear filtering in all up/down-sampling layers, and equalized learning rate for all trainable parameters. Other settings include minibatch standard deviation layer at the end of the discriminator, an exponential moving average of generator weights, style mixing regularization, and nonsaturating logistic loss with R1 regularization. The optimizer used is Adam with hyperparameters $\beta_1 = 0.5$, $\beta_2 = 0.9$, $\epsilon = 10^{-8}$, and minibatch = 64.

3.2 Preliminary results

Figure 1 presents examples of images generated or synthesized by mixing two latent codes at various scales. The five images in the first left column are generated images from random noise, named source A, and the set of four images at the top are generated from random noise, named source B. The rest of the images, called style-mixing images, were generated by copying a specific subset of styles from source B and taking the rest of the styles from source A. Figure 1.1 shows the images synthesized as a result of copying the styles corresponding to the coarse spatial resolutions (4^2-8^2). The images show high-level aspects from source B, while finer features resemble those from source A. Figure 1.2 shows the synthesized sneaker images resulting from copying the styles corresponding to the middle-level spatial resolutions (16^2-32^2), in which on a smaller scale the aspects of the shoes of source B are extracted, while the overall shape and color of source A are briefly reserved. In Figure 1.3, the higher resolutions (32^2-64^2) are used to extract the styles from source B to mix with source A. Figure 1.4 shows the images generated by mixing more finer styles (64^2-1024^2) from source B with source A.

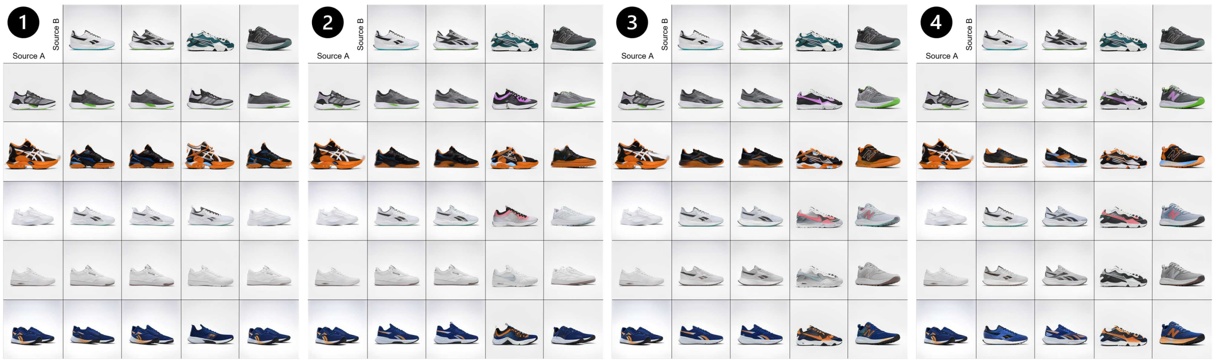

Figure 1. Examples of images generation and style mixing at various scales.

Four important observations from these examples are as follows. First, the style mixing images in Figure 1.1 were synthesized with more similarity to source A and few features from source B, indicating that the model copies the styles corresponding to coarse spatial resolutions from source B and brings high-level aspects such as general sneaker style, overall shape, color, midsole, and orientation from source A. Second, the mixed-style images in Figures 1.2 and 1.3 preserve the middle style of both source A and source B, where some features of source A can be observed and some features of source B are easily recognized in different shoes. Furthermore, with an increase in the style scale of source B added to source A, the images in Figure 1.3 display more distinct styles of B compared to Figure 1.2. Third, the synthesized images listed in Figure 1.4 contain a finer style variation from source B, such as color, outlook patterns, heel counter, midsole, and outsole tread. At the same time, they still remain small-scale styles of source A and it is prominent that the synthesized sneakers are most style-mixed and novel. We can see that each subset of styles controls meaningful high-level attributes of the image. Fourth, it can be visually inspected that both the generated samples and the synthesized samples resemble highly existing shoe models and brands. For example, as shown at the intersection of the third row

[1] https://github.com/rosinality/stylegan2-pytorch

https://doi.org/10.1017/pds.2023.64 Published online by Cambridge University Press

and the last column of Figure 1.4, the sneaker synthesized by mixing the styles of Asics and New Balance is finally a New Balance. This concurs with the aforementioned limitation of GAN generators to the training dataset.

Frechet inception distance (FID) is applied to evaluate the quality of the generated images. FID measures the discrepancy between two sets of images by comparing the distributions of randomly sampled real images from a training set and the generated images (Heusel et al., 2017). FID values are calculated for every pair of images. The average value and lower scores have been shown to correlate well with higher-quality images. In contrast, a higher score indicates a lower-quality image. In the results of the experiment reported in this article, the FID value decreased from 289.65 to 18.97 after 65,000 training steps and converged. The small FID value of the experiments shows good performance of the StyleGAN2 model and confirms that the generated samples are realistic and high quality.

GANs generate new sample using random noise vector, meaning one can think although they follow the input distribution, they may sometimes generate distict and new ideas. However, our quantitative study by formulating metrics such as novelty and diversity shows that the only case that GANs produce distinct concepts is when the output is not identifiable as a sneaker. This suggests that in GANs, usefulness (in terms of the defined design problem) is against novelty since only one of them can be reached through the traditional architecture.

3.3 Diversity analysis

In this section, we present the results of diversity analysis using PCA on the generated samples by StyleGAN2. For this purpose, we produced 50 batches of generated and style-mixed images, each containing $5 + 5$ generated images as source A and source B, as well as 5×5 style-mixedimages generated from these sources. As a result, we built a set of 1,750 images in total to be evaluated in terms of diversity. To produce random noise for source instance generation from the latent space to be fed to StyleGAN2 as an input, one numerical element is required to be set as the random seed that prevents redundancy in the generated set. To further increase the diversity of the output images, we produced 50 sets of 2×5 random variables that were then fed to the model one set at a time. We adopted a PCA model to assess the diversity of the outputs by first quantitizing it through calculating the most informative aspects of the data samples and then, visualizing the most effective features. PCA, however, works better on tabular data format with dependent features of a high-dimensional space compared to the raw image. As a result, we first extracted higher-level features from the original images and generated them to further compare the areas of the output space that the two sets cover. For feature extraction, we used the VGG16 model, because of its capability to embed images simply by removing the top output layer. The model was first trained on the combination of our original and generated dataset to better identify the features from a design perspective rather than considering general features extracted from broader datasets such as ImageNet. The VGG16 model was trained on RGB images of size 224×224 with 3 fully connected layers at the top of the network, no pooling layers, and softmax activation function.

The extracted features were fed to the PCA model that was initialized by two principal components, so that we can visualize and compare the sets in a two-dimensional space. PCA is a beneficial method for diversity analysis in this context, as it is capable of compressing the size of the data while extracting the most important information, thus enabling structure analysis for the dataset by simplifying the descriptive features. Using PCA, each data sample is mapped to a point in the new coordinate system that allows the representation of the pattern of similarity of the data samples. The 2D representation of the mapped data points in Figure 2 with the red point and the green points representing the original and generated data sets, respectively, shows that the entire original space was not explored by the StyleGAN model, resulting

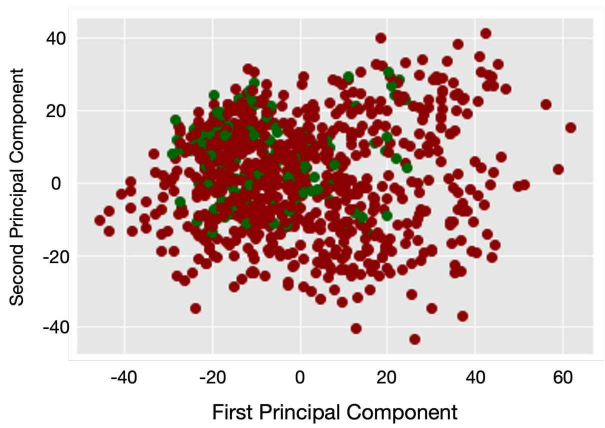

Figure 2. Principal components extracted from the original (red) and generated datasets (green).

https://doi.org/10.1017/pds.2023.64 Published online by Cambridge University Press

in design concepts limited to a specific and incomprehensive set of styles. The green dots cover a subset of the space covered by the red dots, suggesting the limitation of GANs in learning the entire distribution of the dataset. The scatter plot also suggests that the model is especially inadequate in areas where there are fewer original samples, demonstrating the dependence of learning a subspace on the presence of adequate data samples from the subspace.

3.4 Novelty analysis

In this section, details of the experiments and results of the novelty analysis are presented. To investigate the ability of StyleGAN2 to produce novel solutions, we used the template matching technique. The analysis was performed on the same set of 1750 generated images as described in the previous section. The procedure is first to find the most similar design instance from the original dataset to each one of the generated samples, and then to aggregate the results on a distribution function of similarity to statistically assess the extent of novelty in the outputs. Template matching is usually used to find the part of a template image that is most similar to a source image by computing and comparing the confidence scores of different areas in the template image according to a sliding window. The confidence of a point in this regard represents the algorithm's certainty of similarity between the source image and the template image's area within the rectangle whose top-left corner is at the said point and whose height and width are the same as those in the source image. However, we took advantage of its capability to calculate an overall similarity score for the entire template image to then rank the similarities of each generated image with all samples in the original dataset. In order for template matching to be aligned with our purpose, we used the source (i.e., generated) images and the template (i.e., original) images of the same size, so that all areas contribute the same in the calculation of the confidence score.
Figure 3.1 shows the distribution plot along with the semi-Gaussian function fitted with a mean of 0.8385 and a variance of 0.0075. Statistics show that most samples are very similar to those of a sample from the original dataset, which proves the theory that GANs are not capable of generating novel design concepts. However, there are samples with low confidence of similarity that can be due to two main reasons: 1) some of the generated images are as unrealistic as they cannot inherently be identified as sneakers; 2) On one hand, template matching treats the same shape with different colors as different shapes, thus, two sneakers with the exact same style but different colors have a low similarity score. On the other hand, one of the main changes that style-mixing applies when generating an image is altering the color. As a result, a considerable number of style-mixed images may not have a high confidence score despite having a parallel instance in the original dataset. An example of the most similar matches found for a generated image by iteratively calculating the confidence score on the original dataset is illustrated in Figure 3.2. It is visually inferable from the images that the generated image resembles an existing design instance from the original dataset and does not contain any novel features. Furthermore, we calculated the distribution function of the confidence scores for the generated samples to quantitatively assess the novelty of StyleGAN in an aggregated way.

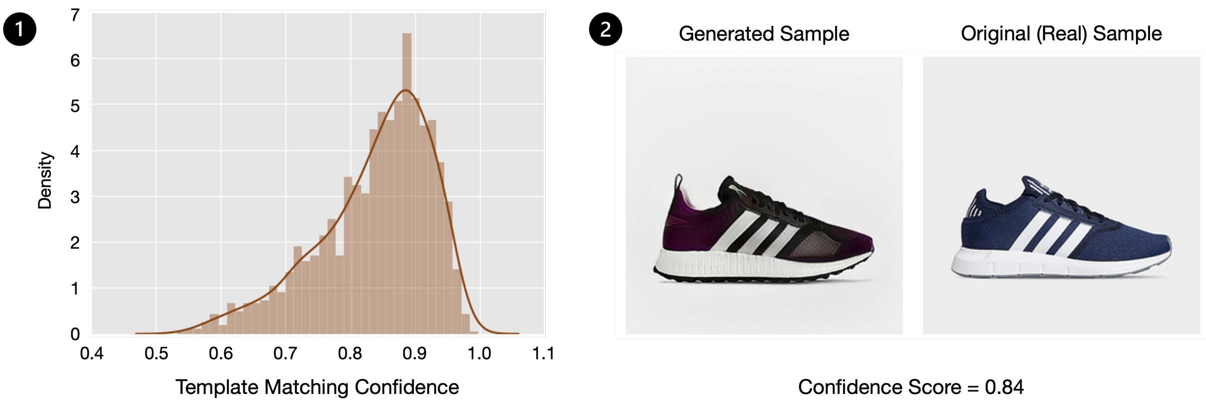

Figure 3. (1) Distribution function and correlating semi-Gaussian function of the template matching confidence scores based on the generated-real comparisons. (2) Example of a generated sample and the most similar real sample from the training dataset.

https://doi.org/10.1017/pds.2023.64 Published online by Cambridge University Press

4 CONCLUSIONS

This paper expands on StyleGAN2, one of the most cutting-edge variations of GAN to generate visual design concepts. The StyleGAN2 model has the ability to generate realistic images and then synthesize the images produced through style mixing to introduce some diversity to the generated samples. By mapping proportional styles to the source image, the architecture enables intuitive scale-specific mixing operations where coarse, medium, and fine styles can be observed. This approach can potentially serve designers by helping them in design ideation with multimodal concept generation, aggregation, and mixing. This chapter contributes to the vast literature on StyleGAN applications by investigating the challenges and limitations of this deep generative modeling approach in the field of engineering design. Larger-scale GAN models currently have a strong capability to generate high-quality and high-resolution images. However, the results of these models are shown to not fully capture the diversity of the true distribution. This is partly due to limited availability of training data in the engineering design domain, and more significantly, because of the generator-discriminator architecture and its sole emphasis on sample quality and realism. GANs also suffer from the lack of usefulness as it does not have any guidance to generate concepts that are not from the input distribution and still fulfill the requirements for a design problem. That is why we should provide guidance for the network to learn other criteria, so it generates samples out of the distribution that are relevant to the desired product. To this end, our next steps will include providing the model with guidance for design-specific objectives, for example, preserving the geometrical balance of the concept, which is an important aspect of usefulness. Future research should also explore approaches that can improve the resolution of dataset in domains where data are scarce or difficult to collect. Furthermore, it is necessary to devise new modifications to the architecture of StyleGAN2 to guide the generator towards generating more diverse and novel samples: (1) An enhanced mapping model can be developed that can reparameterize the latent generative space as a mixture model sampled from the chosen Gaussian and learn the parameters of the mixture model to enrich the model and generate diverse samples, particularly in cases where the dataset is small and limited. (2) A layer-wise decomposition approach can be devised to identify useful potential control spaces and allow manipulation of images from high-level properties on either the latent space or the feature space. The authors also propose to couple GAN-based, visual generative design with user-centered evaluation mechanisms (Yuan et al., 2022) for end-to-end design concept generation informed by large-scale user reviews. The proposed architecture combines the predefined evaluation method with an adversarial objective function to train the generator and the discriminator. It can measure the quality and diversity of individual generated samples, where each generated sample is updated based on an evolutionary adversarial training framework. By integrating an evolutionary optimization algorithm to the generative model where the fitness of the samples generated by the generator is measured by the DMDE model, the proposed architecture can serve as a comprehensive generative design tool that can not only ensure realism and quality, but also guarantee the desirability and performance of the generated samples.

ACKNOWLEDGMENTS

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

Abdi, H. and Williams, L.J. (2010), "Principal component analysis", *Wiley Interdisciplinary Reviews: Computational Statistics*.

Ahmed, F. (2019), *Diversity and Novelty: Measurement, Learning, Optimization*, PhD dissertation, University of Maryland.

Basulto-Lantsova, A., Padilla-Medina, J.A., Perez-Pinal, F.J. and Barranco-Gutierrez, A.I. (2020), "Performance comparative of opencv template matching method on jetson tx2 and jetson nano developer kits", *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*.

Buonamici, F., Carfagni, M., Furferi, R. and Governi, L. (2020), "Generative design: An explorative study", *Journal of Computer-aided Design and Applications*, Vol. 18, pp. 144–155.

Dieter, G.E., Schmidt, L.C. et al. (2009), *Engineering design*, Vol. 4, McGraw-Hill Higher Education Boston.

https://doi.org/10.1017/pds.2023.64 Published online by Cambridge University Press

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y. (2014), "Generative adversarial nets", in: *Advances in neural information processing systems*, pp. 2672–2680.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B. and Hochreiter, S. (2017), "Gans trained by a two time-scale update rule converge to a local nash equilibrium", *Advances in neural information processing systems*, Vol. 30.

Karras, T., Laine, S. and Aila, T. (2019), "A style-based generator architecture for generative adversarial networks", in: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4401–4410.

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J. and Aila, T. (2020), "Analyzing and improving the image quality of stylegan", in: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 8110–8119.

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

Oh, S., Jung, Y., Kim, S., Lee, I. and Kang, N. (2019), "Deep generative design: Integration of topology optimization and generative models", *Journal of Mechanical Design*, Vol. 141 No. 11.

Shah, J.J., Smith, S.M. and Vargas-Hernandez, N. (2003), "Metrics for measuring ideation effectiveness", *Design studies*, Vol. 24 No. 2, pp. 111–134.

Shmelkov, K., Schmid, C. and Alahari, K. (2018), "How good is my gan?", *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 213–229.

Shu, D., Cunningham, J., Stump, G., Miller, S.W., Yukish, M.A., Simpson, T.W. and Tucker, C.S. (2020), "3d design using generative adversarial networks and physics-based validation", *Journal of Mechanical Design*, Vol. 142 No. 7, p. 071701.

Singh, V. and Gu, N. (2012), "Towards an integrated generative design framework", *Journal of Design Studies*, Vol. 33, pp. 185–207.

Toh, C.A., Miller, S.R. and Okudan Kremer, G.E. (2014), "The impact of team-based product dissection on design novelty", *Journal of Mechanical Design*, Vol. 136 No. 4, p. 041004.

Wang, Z., She, Q. and Ward, T.E. (2021), "Generative adversarial networks in computer vision: A survey and taxonomy", *ACM Computing Surveys (CSUR)*, Vol. 54 No. 2, pp. 1–38.

Wu, Q., Liu, Y., Miao, C., Zhao, B., Zhao, Y. and Guan, L. (2019), "Pd-gan: Adversarial learning for personalized diversity-promoting recommendation", *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19)*, p. 3870–3876.

Xu, J., Ren, J., Lin, J. and Sun, X. (2018), "Diversity promoting gan: A cross entropy-based generative adverserial network for diversified text generation", *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 3940–3949.

Yuan, C., Marion, T. and Moghaddam, M. (2022), "Leveraging end-user data for enhanced design concept evaluation: A multimodal deep regression model", *Journal of Mechanical Design*, Vol. 144 No. 2.

Yuan, C. and Moghaddam, M. (2020), "Attribute-aware generative design with generative adversarial networks", *IEEE Access*, Vol. 8, pp. 190710–190721.

https://doi.org/10.1017/pds.2023.64 Published online by Cambridge University Press

https://doi.org/10.1017/pds.2023.64 Published online by Cambridge University Press