## GANSpiration: Balancing Targeted and Serendipitous Inspiration in User Interface Design with Style-Based Generative Adversarial Network

Mohammad Amin Mozaffari mohammad-amin.mozaffari@polymtl.ca Polytechnique Montreal Montreal, QC, Canada

> Jinghui Cheng jinghui.cheng@polymtl.ca Polytechnique Montreal Montreal, QC, Canada

## ABSTRACT

Inspiration from design examples plays a crucial role in the creative process of user interface design. However, current tools and techniques that support inspiration usually only focus on example browsing with limited user control or similarity-based example retrieval, leading to undesirable design outcomes such as focus drift and design fixation. To address these issues, we propose the GAN-Spiration approach that suggests design examples for both targeted and serendipitous inspiration, leveraging a style-based Generative Adversarial Network. A quantitative evaluation revealed that the outputs of GANSpiration-based example suggestion approaches are relevant to the input design, and at the same time include diverse instances. A user study with professional UI/UX practitioners showed that the examples suggested by our approach serve as viable sources of inspiration for overall design concepts and specific design elements. Overall, our work paves the road of using advanced generative machine learning techniques in supporting the creative design practice.

## **CCS CONCEPTS**

- Human-centered computing  $\rightarrow$  User interface design.

## **KEYWORDS**

User interface design, inspiration, StyleGAN, creativity support

#### **ACM Reference Format:**

Mohammad Amin Mozaffari, Xinyuan Zhang, Jinghui Cheng, and Jin L.C. Guo. 2022. GANSpiration: Balancing Targeted and Serendipitous Inspiration in User Interface Design with Style-Based Generative Adversarial Network. In *CHI Conference on Human Factors in Computing Systems (CHI '22), April* 29-May 5, 2022, New Orleans, LA, USA. ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3491102.3517511

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 978-1-4503-9157-3/22/04...\$15.00 https://doi.org/10.1145/3491102.3517511 Xinyuan Zhang xinyuan.zhang@polymtl.ca Polytechnique Montreal Montreal, QC, Canada

Jin L.C. Guo jguo@cs.mcgill.ca McGill University Montreal, QC, Canada

## **1** INTRODUCTION

User interface designers take on daily challenges in creating effective, usable, and innovative design work. Inspiration from existing design examples plays an essential role in this creative process [12]. Previous studies have observed that designers often actively seek, combine, and transform design examples to draw an analogy from existing full-fledged designs embedded with similar yet provocative ideas [16].

Current tools that support this inspirational activity usually fall into one of the two categories. First, the proliferating design gallery platforms, such as Dribbble [10] and Behance [1] support a bottomup, serendipitous inspirational process where the designer examines a collection of designs, seemingly without a goal, in order to find "interesting" examples helpful to their work. Second, some recent work proposed design inspirational tools focused on suggesting examples based on certain types of design input (e.g. in the form of a sketch or an existing design), usually leveraging algorithms to determine image similarity [2, 23, 27, 30]; this represents a topdown, targeted process where the designer has a concrete idea in mind and seeks examples that implemented the idea.

While providing important inspirational support, both types of approaches have limitations. Our preliminary interaction with designers reveals that seeking inspiration from the design galleries can sometimes be an overwhelming experience and result in "design drift" (i.e. shifted design ideas from the original focus). On the other hand, over-exposure to examples with similar styles might cause design fixation (i.e. "a blind adherence to a set of ideas or concepts" [19]) that hinders the novelty of the design work [16, 24].

To address the limitations of existing design supporting tools, we attempt to seek a balance between targeted and serendipitous inspirations in this work. Particularly, we propose *GANSpiration*, a set of approaches that use a style-based generative adversarial network (StyleGAN), trained with a large dataset of existing interface designs, to generate a diverse and yet focused set of examples based on a preliminary design input. StyleGAN is built on the Generative adversarial network (GAN), a machine learning framework that is comprised of two neural networks trained jointly (i.e. a generative network and a discriminative network) [13]. The generative network in StyleGAN, in particular, takes "style inputs" at different stages during the process of image synthesis, controlling the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

style of the generated images at different levels of granularity and details [20].

GANSpiration-based approaches leverage the StyleGAN technique to perform style transfer and generate new design artifacts based on existing design artifacts and, therefore, provide a targeted and serendipitous inspiration for user interface designers. These approaches take a preliminary user interface design as input, merge the input image with a random set of existing designs using Style-GAN, and output representative examples that are either directly synthesized or from real UI screenshots that resemble the synthesized examples. During the style merge, the GANSpiration-based approaches alter the layout and/or the details of the original input, leveraging StyleGAN's architecture. Note that although user interface design artifacts can be represented in many formats (e.g., a tree of UI components [34]), we considered the UI screenshot images in our approach since they are one of the most frequent types of inspirational sources used by the designers. When investigating and evaluating GANSpiration, we pose the following research questions:

- **RQ1**: How do GANSpiration-based approaches compare with random examples and similarity-based examples in quantitative metrics indicating the ability of inspiration support?
- **RQ2**: How do UI/UX practitioners perceive the output of GAN-Spiration-based approaches in comparison to random examples and similarity-based examples?

To answer these questions, we first developed our approach which extends the StyleGAN architecture trained on a large-scale dataset including 58,040 screenshots of Android applications. We then proposed two quantitative measurements for evaluating the ability of inspiration support of a set of UI images: (1) similarity of the images to the input image and (2) diversity of the set of UI images. We found that the GANSpiration-based methods provide much more diverse design examples than a similarity-based method, and at the same time they provide more similar examples to the input image than a random example selection approach, indicating a balance between diversity and relevance. Through a user study with five professional UI/UX practitioners, we found that the participants perceived the GANSpiration-based methods as a viable way to gain inspiration to modify a UI design. Overall, our work contributes a novel and promising approach in which a style-based generative machine learning technique is applied in the context of inspiration and creativity support in user interface design. We believe that the ideas presented in our approach will encourage and influence more research efforts towards the pragmatic use of generative machine learning models in the creative, yet constraint, design tasks.

## 2 RELATED WORK AND BACKGROUND

Our work is most closely related to previous studies that focused on (1) design inspiration, (2) techniques for managing design artifacts, and (3) generative machine learning models and StyleGAN in particular. We briefly review each group of literature in the following sections.

## 2.1 Design Inspiration

Thrash et al. [31, 32] were among the first who empirically studied inspiration as a psychological construct. They have identified that human inspiration is categorized by motivation (i.e. goal-oriented self-initiation), evocation (i.e. an impulsive reaction to stimuli), and transcendence (i.e. feeling of gaining superior ideas that are "more elegant or novel than those generated willfully").

The problem of inspiration has been then investigated in a wide design community, beyond user interaction design. These previous studies were mostly conducted from the perspectives of how designers get access to and use existing design artifacts. For example, focusing on knitwear design, Eckert and Stacey [11] have identified that designers used a wide variety of sources of inspiration, including artifacts with intriguing shapes, patterns, and colours, as well as their own previous design, to not only concretize the otherwise abstract design ideas, but also to create "shortcuts" to help them recall and communicate using these visuospatial "chunks;" i.e. inspirational sources served as "a language of design."

In the HCI community, researchers have explored ways how industrial and user interaction designers get inspired by existing design artifacts. For example, Bonnardel [5] has identified that, in the context of product design, "the emergence of new ideas results from analogy-making." From an in-depth interview study with web, graphic, and product designers, Herring et al. [16] identified the common approaches they used and the challenges they faced when retrieve, store, and disseminate design examples. Based on a glossary of design ideation methods, Gonçalves et al. [12] have also conducted a survey with students and professional industrial designers to understand their sources and methods of inspiration. They found that, comparing to students, professional industrial designers adopted a wider variety of inspirational approaches.

The literature has also identified several problems and issues about the common inspirational methods. Notably, many studies have pointed to the fact that over-exposure to a homogeneous set of design examples may result in "design fixation," which will limit the inspirational power of the examples and result in less creative ideas [16, 19, 24]. Particularly, Marsh et al. [24] identified that exposing to a greater number of examples that share common critical characteristics would increase the fixation issue. The timing of the example exposure can affect the quantity and quality of ideas as well. For example, Siangliulue et al. [28] found that receiving examples when their participants seemed to have run out of ideas have allowed the participants to produce a larger number of ideas, whereas explicitly requesting examples when needed have allowed the participants to produce more novel ideas. In this study, we build on this body of literature to investigate techniques for supporting effective inspiration in user interface design, while avoiding design fixation.

## 2.2 Managing UI Design Artifacts

While abundant recent work focused on extracting UI elements, including their hierarchical design information, from design artifacts such as mockups (e.g. [3, 7, 8, 25, 30, 35]), they are not directly related to the objective of providing inspirational design examples. So we omit the detailed review of this body of literature here. In this section, we focus on reviewing related work that investigates the management of UI design artifacts for the purpose of design inspiration.

Towards this direction, some previous studies have focused on techniques that retrieve UI design examples based on an input UI screenshot. For example, Lee et al. [23] have proposed an "Adaptive Ideas" web design tool, which allows users to view examples similar to their current design work. In their tool, the users could control the dimensions (including background color, primary font, number of columns, and visual density) used to compare design similarities. Similarly, Behrang et al. [2] proposed a technique that combined keyword search and image-based search to retrieve apps (along with their code) with similar screenshots as an input design. Hashimoto et al. [14] have also introduced a technique that aims to help inexperienced designers retrieve similar design examples based on an input in the form of sketch or wireframe. Ritchie et al. [27] have proposed a design exploration tool that allows its users to query design examples by descriptive text including color keywords or style terms; the tool can also search by style similarity.

The recent development of deep neural networks has enabled more powerful techniques for similarity-based design example retrieval. For example, Huang et al. [18] introduced Swire, a sketchbased neural network-driven technique for retrieving user interface designs. The core component of Swire is a deep convolutional neural network (CNN) [22] that calculates the "embedding" (i.e. a numerical representation) of a design artifact (e.g. a sketch or a screenshot). Once trained, Swire could retrieve UI design artifacts that are similar to an input sketch or screenshot. More recently, Bunian et al. [6] proposed VINS to retrieve the most structurally similar UI screenshots to the input using object detection models to identify the UI components of the UI screenshots or wireframes. Based on the components and their layout, an image retrieval model helps to find similar UI screenshots in the reference dataset.

Notably, most previous studies relied on similarity when retrieving design artifacts, which can result in design fixation and may hinder the creative design process. Our study addresses this issue by focusing on a style-based generative approach that balances the targeted and serendipitous aspects of design artifact retrieval for inspiration.

### 2.3 StyleGAN and its Application on UI Design

StyleGAN, or Style-Based Generative Adversarial Network [20], extends the traditional GAN [13] architecture on a style-based generator model inspired by the style transfer literature. GANs typically include two machine learning models that are trained at the same time: a *generator* trained to synthesize data points (e.g., images) that resemble those in the original dataset and a *discriminator* that learns to classify if an input image is synthesized by the generator. Once trained, the generator can be used to synthesize images from an input vector in the latent space. Based on the GAN architecture, StyleGAN proposes a style-based generator that focuses on explicitly transferring 'styles' on an image at different resolution levels during the synthesis process. This results in a synthesized image with one input vector, or "latent code", dominating its overall features and the other latent code contributing mostly to the details of the image [20].

While GAN and StyleGAN have obtained great attention given their capacity for generating high-resolution and realistic-looking images, their application for UI design is still in its infancy. The only previous work that used generative models for providing UI examples is a very recent study done by Zhao et. al. [37]. They developed a technique to generate UI structures and reused UI components collected from existing mobile apps to fill in the generated structure in order to create UI examples. Their study only focused on the quality of the generated UIs, in terms of metrics such as color harmony and structure rationality. Additionally, their evaluation was not done with professional designers who are familiar with real-life design practices. Instead, we aim to understand the ability of the generative models in providing inspirational design support. Our techniques also address style-based design transformation, which is not explored in the literature.

StyleGAN has been improved since its publication. While our work is based on the original StyleGAN work, we expect the performance of our approach can be enhanced by using more advanced generative models, such as StyleGAN2 [21]. Our contribution, however, is not on using the most recent models, but instead on illustrating the potential of applying this line of work for a novel but important problem, i.e., generating design examples with high diversity and relevance for effective inspiration.

## **3 GANSPIRATION SYSTEM ARCHITECTURE**

The interaction between the designer and GANSpiration is initiated when the designer has a preliminary design artifact (e.g., a UI mockup image) at hand, related to their design task. The designer sends this image as an input to GANSpiration, which will first be encoded into a latent code (i.e., a high-dimension vector in the latent space). This latent code is then merged with other latent codes, either randomly generated or obtained from other UI images, to synthesize a unique set of new example images. From this set, the system selects the most representative example images and displays them to the designer. The designer can also configure GANSpiration to return the real UI screenshots from the database that are the closest to the generated results. As such, the GANSpiration architecture is primarily comprised of three components to achieve its key functionality: (1) a latent code search component, (2) a new examples synthesizer, and (3) a representative examples selection component. The overall architecture is illustrated in Figure 1. Below, we first describe each component in detail. Then we describe the process used to train the StyleGAN that supports these components.

#### 3.1 Component 1: Latent Code Search

StyleGAN relies on a condensed representation of images called latent codes, i.e., high dimensional vectors (512 dimensions in our case), to perform image synthesis. The input images of GANSpiration, therefore, need to be first encoded from their original format to the latent space that corresponds to a trained StyleGAN model. This is done through Latent Code Search, which is built upon the work of StyleGAN-Encoder [26]. This component searches the optimal latent code of an input image through a gradient descent update based on the difference between the synthesized image and the input image evaluated using a perceptual model [36]. The latent code obtained is then returned to represent the input image.

#### CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

Mohammad Amin Mozaffari, Xinyuan Zhang, Jinghui Cheng, and Jin L.C. Guo

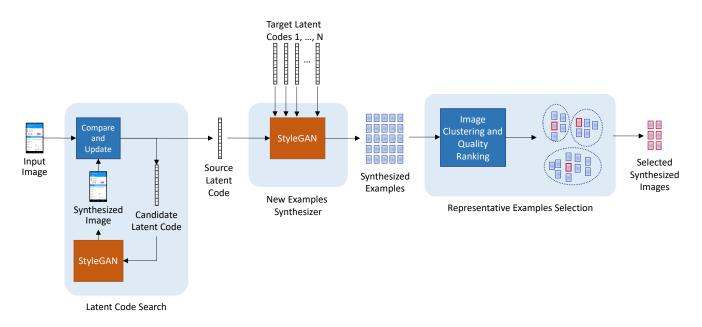


Figure 1: Overview of GANSpiration. GANSpiration takes a preliminary design image as input then transforms it into a latent code. The style of the original input is then merged with a set of target latent codes to synthesize new examples. Image encoding and clustering methods are then used to find the representative examples as output. While not illustrated in this diagram, the output of the synthesized examples can be additionally used to search real UI screenshots for inspiration.

## 3.2 Component 2: New Examples Synthesizer

New Examples Synthesizer merges the style of the input image as source with a set of target images or latent codes to produce a new set of examples. In this process, the source image latent code is obtained from the previous step, while the target latent codes can be any vectors from the latent space depending on the configuration of GANSpiration. For example, it can use the latent codes obtained from a set of existing target images or latent codes that are directly sampled from the latent space.

We set StyleGAN to include eight levels of image resolutions during the synthesis process [20]; each resolution level is comprised of two style input locations into which we feed the source and target latent codes. GANSpiration merges images by replacing the latent code of the source image for certain style input location(s) with the target latent code. We iterate all the consecutive style input locations, resulting in 136 images generated for each pair of source and target inputs. This set of output covers all possible granularity levels of influence of the target image (i.e., from the structural level to the detail level) on the source image.

## 3.3 Component 3: Representative Examples Selection

In the previous step, we synthesized a large number of new examples for each source-target pair, which can be overwhelming for the designers to examine. Moreover, since our style merging process is very fine-grained, it also introduces redundancies in the set of synthesized examples. In this component, we apply a clustering method to pick a smaller sample of representative images from the set of generated images. In particular, we adopt the DBSCAN method [4] (with threshold  $\epsilon$  as 0.9) to cluster the images through perceptual similarity [36] calculated between any two images from the synthesized example set. Within each cluster, we selected the image that the discriminator of the trained StyleGAN considered as the closest to the real UI screenshot as the representative example. The output of this component is a collection of synthesized images that are sufficiently different from each other and contain either coarse or detailed design aspects of the input image.

## 3.4 Training StyleGAN for GANSpiration

To use the capacity of StyleGAN in the context of UI design, we need to retrain the StyleGAN model with a dataset of user interface design artifacts. In this section, we describe the dataset we used and our training process of StyleGAN for GANSpiration. The trained StyleGAN model was then used in the components described above.

3.4.1 Dataset. We used the Rico dataset [9] for StyleGAN training. It contains 66,261 unique UI screenshots of Android apps and serves as one of the largest repositories of mobile app designs to date. The dataset includes a diverse set of UI screenshots with varied complexity that contains various types of UI components. In order to obtain a high-quality dataset for inspiration purposes, we removed UI screenshots that have only one or two unique component labels from the Rico dataset. This is because we found in a manual inspection that the UIs with less than three unique component types usually do not contain enough interaction elements to support inspiration or benefit from our approach. They also often introduce unnecessary visual complexity that only exists in singular instances (e.g., a splash screen, video screenshot, advertisement, or full-screen image, see Figure 2) that affects the training performance; in other

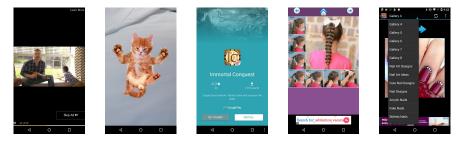
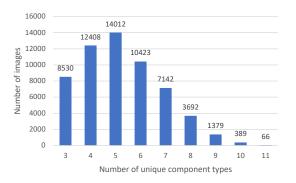


Figure 2: Samples of UI screenshots removed because they only have one or two unique component types. Most of the screenshots in this category contain splash screen, video screenshot, full-screen image, and web view in their design, adding unnecessary visual complexity.



### Figure 3: The distribution of the number of unique component labels in the UI screenshots in our dataset.

words, these instances make it difficult for StyleGAN to generate similar components and even if generated, the visual presentation of those components are too uncommon to be useful for inspiration. We used the UI view hierarchy data in the Rico dataset to calculate the number of unique component types in each screenshot. In total, we removed 8,221 images that had only one or two unique component types, resulting in 58,040 images in the dataset. Figure 3 shows the distribution of the number of unique component labels in the UI screenshots in the dataset. We then resized each image into  $1024 \times 1024$  for the training purpose.

3.4.2 Training process. We built upon the official TensorFlow implementation of StyleGAN<sup>1</sup> and used our preprocessed dataset to train the model. While we did not perform a formal hyperparameter search, we explored a few changes of hyperparameters reported in Karras et al. [20], including the initial learning rates of the generator and the discriminator, the number of times the discriminator was trained per generator iteration, and the number of minibatches to run before adjusting the training parameters. We eventually used the following hyperparameters because they achieved the best performance according to the Fréchet inception distance (FID) [17] in our exploration: for both the generator and the discriminator, the learning rate was set to 0.0015 for resolution levels equal or less than  $128 \times 128$ , 0.002 for resolution levels  $256 \times 256$  and  $512 \times 512$ , and 0.003 for  $1024 \times 1024$ ; the discriminator was trained at the

same frequency as the generator; and training parameters were adjusted after every four minibatches. Additionally, mirror data augmentation used in the original StyleGAN work was not enabled during training due to the unsymmetrical nature of UI images. We used a server that contained four NVidia V100SXM2 GPUs to train the model. Training terminates when the FID value increases (i.e., the generation quality deteriorates) for three consecutive iterations. The entire training process lasted 162 hours. The best performance measured with FID was achieved at 42.91. Although this performance is not ideal compared to the typical face generation tasks, we found that the generated images can already represent certain layout features and visual details that can help design inspiration. This difference in performance may due to the extraordinary complexity and diversity of screenshots of UI design. It is worth noting that our focus and contribution in this paper are not to achieve a higher performance in the generative model. Instead, we focus on evaluating the potential of this line of techniques for supporting the challenging task of design inspiration, even with less-than-perfect image generation.

## **4 SAMPLE USAGE SCENARIO**

The main purpose of GANSpiration-based approaches is to inspire designers with a set of UI design examples that are diverse enough to help avoid design fixation, while relevant to the designer's work at hand. To illustrate the design process with and without the support of GANSpiration, we describe a sample usage scenario from a UI designer's perspective. This usage scenario is created based on our informal discussion with several designers and is used to guide the design of our evaluation studies.

Daphne worked as a UI/UX designer for six years at a large company. She was recently positioned to lead the design of a revamped version of the company's main product. Before Daphne adopted the GANSpiration-based tool, she used to rely on two main ways to draw inspiration and reference from other systems and design examples during the ideation phase (sometimes with the users and/or the development team). First, Daphne frequently visited her favorite design gallery platforms, Dribbble [10] and Behance [1], either searching for examples for a certain type of functions and components or simply fumbling through the galleries until ideas hit her. These galleries contain collections of screenshots of design examples in various application domains, with drastically different design styles. So, although she felt that they are useful to help her

<sup>&</sup>lt;sup>1</sup>https://github.com/NVlabs/stylegan

think out of the box, she sometimes found those examples distracting and the whole process ineffective. Second, Daphne also had access to a similarity-based design example retrieval tool that her company procured. This tool would provide her with a collection of UI design screenshots that looks similar to a preliminary mockup that she created. She found this tool particularly useful in helping her to be focused on the design elements presented in her mockups. However, the design examples suggested by the tool often look very similar. As a result, Daphne often found herself stuck in a certain way that a UI is 'supposed' to look like and lost her creative power.

Daphne found the GANSpiration-based tool achieving a good balance between the two inspirational approaches she used before. At the beginning of the design revamping work, Daphne fed the screenshots of the old design to the tool and received a collection of synthesized design examples that resemble either the structure or the details of the original design but each had unique aspects that Daphne found interesting. She was able to create an initial revamped design mockup by combining several elements and aspects from the synthesized images. Daphne found that GANSpiration can effectively support communicating design ideas too. During a co-design session with two power users of their product, Daphne received a suggestion of changing the presentation of a list of items on the mockup. Examining the GANSpiration output of the mockup, Daphne got the idea of either adding an icon to each list item (i.e., a detailed change) or changing the list into a wall of cards (i.d., a structural change). Daphne picked a few examples synthesized by the tool to show to the users to illustrate her ideas and eventually settled on the wall-of-cards concept. Overall, Daphne was glad that she could easily obtain a collection of design examples in various occasions that were not only diverse enough to allow her to explore different design ideas but also relevant to her work at hand.

## **5** QUANTITATIVE EVALUATION

We conducted an experiment that focused on evaluating the generative methods against two other techniques that suggest design examples (i.e., random suggestion and similarity-based suggestion). The evaluation is based on the usage scenario we described above. The random suggestion baseline resembles the serendipity-based inspirational process that relies on exploration and encountering, while the similarity baseline reflects the similarity-based design example retrieval approach. While we used two quantitative metrics (i.e., similarity and diversity) for measuring the ability of inspiration support of the system outputs, we also focus on discussing our observations and insights on how the outputs might have contributed to the metrics.

## 5.1 Data sampling

To sample a diverse set of UI screenshots as inputs to GANSpiration, we divided the Rico dataset according to the number of unique component types on the UIs. We considered the number of unique component types as an indicator of the complexity of the UI, thus the complexity of the design task. Our dataset contained UI screenshots that contain 3 to 11 unique component types, resulting in nine unique groups (see Figure 3). We randomly selected three UI screenshots from each group, resulting in 27 screenshots as inputs in the evaluation scenario; these images are shown in Figure 4. This strategy ensures the coverage of different levels of complexity in the UI inputs, thus the coverage of the complexity of the design task.

## 5.2 Quantitative metrics

We derived two metrics to measure the ability of inspiration support of a set of UI images. These metrics are inspired by previous work about inspiration in design [15, 23] and collectively focused on both targeted and serendipitous inspiration. Both metrics rely on a measure of distance between images, particularly a measure based on a perceptual distance model aimed to approximate human visual perception; this measure is the same as the one used in the clustering task in the representative examples selection component of GANSpiration (see section 3.3).

In the following equations, dist(A,B) denotes the perceptual distance between images A and B, ranging from 0 to 1 inclusively, calculated using Zhang et al.'s technique [36];  $E_i^{(D)}$  (i = 1, 2, ..., n) denotes the *i*th output example image for an input design D; and  $O^{(D)} = \{E_1^{(D)}, E_2^{(D)}, ... E_n^{(D)}\}$  denotes the set of output images for an input D. The two metrics we used are:

• Similarity of the suggested examples to the input design. This metric is double-sided. A sufficient similarity may indicate the relevance of the suggested examples, which is important to provide targeted inspiration. However, a high similarity indicates the potential of design fixation. We use the mean similarity (i.e., the complement of the perceptual distance) between the output images and the input image to evaluate the overall similarity of the output examples to an input design:

$$Sim(O^{(D)}) = 1 - \frac{1}{n} \sum_{i=1}^{n} dist(D, E_i^{(D)})$$

• **Diversity** of the suggested examples indicates how varied the outputs are, given an input design. It plays an important role in preventing fixation. We use the mean of pairwise distances among all output examples of an input design to evaluate the diversity of the output set:

$$Div(O^{(D)}) = \frac{1}{n(n-1)/2} \sum_{j=i+1}^{n} \sum_{i=1}^{n-1} dist(E_i^{(D)}, E_j^{(D)})$$

#### 5.3 Experimental design

In the experiment, we considered the following conditions for suggesting a set of images for inspiration. Particularly, conditions 1, 2, 3, and 4 are four variants of the GANSpiration method. Figure 5 summarizes these conditions.

**Condition 1**. In this condition, we used the trained StyleGAN model to merge the input image with five random latent codes, each generated a set of examples according to the GANSpiration architecture (see Section 3). We then directly used these examples as the output.

**Condition 2**. This condition is similar to Condition 1; however, instead of using five random latent codes, we randomly selected five images from the prepossessed dataset and obtained their latent codes for style merging. The examples generated were then used as

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

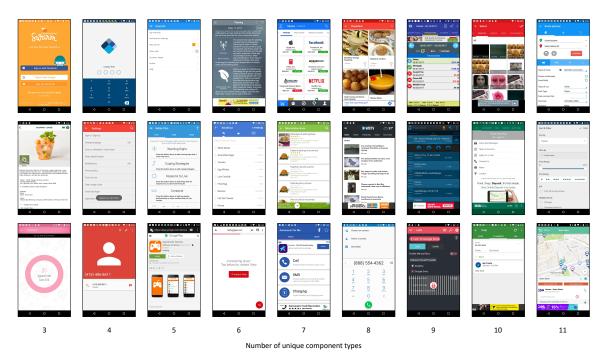


Figure 4: Sample of UI screenshots used in the evaluation study, ordered by the number of unique component types.

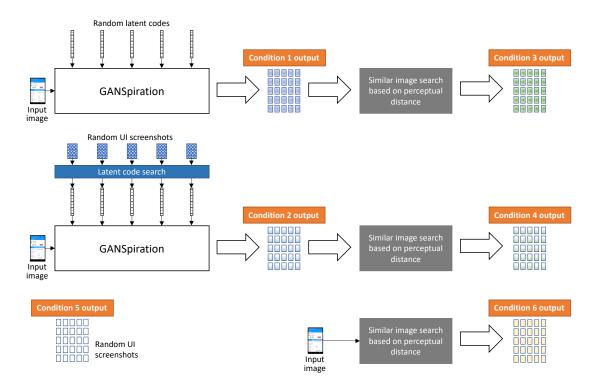


Figure 5: The six experimental conditions. The condition outputs were then used to calculate the metrics and in the user studies.

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

Mohammad Amin Mozaffari, Xinyuan Zhang, Jinghui Cheng, and Jin L.C. Guo

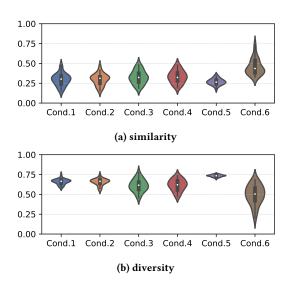


Figure 6: Distribution of the similarity and diversity metrics on the experimental conditions.

the output. Conditions 1 and 2 are created to evaluate two different variants for using directly generated images for inspiration.

**Condition 3.** In this condition, we first obtained the output of Condition 1. Then for each generated output image, we searched for the most similar image, using the perceptual distance described in section 5.2, from a dataset of real UI screenshots. Because of the computational cost of this search, we used a smaller dataset created by Huang et al. [18] as the search space; it contained 2201 high-quality UI screenshots sampled from the Rico dataset. Real UI screenshots obtained from this search were then used as the output.

**Condition 4**. This condition is similar to Condition 3 except that the initial generated images for search were obtained from Condition 2 instead of Condition 1. Conditions 3 and 4 are created to examine the effects of how realistic the design examples are on inspiration.

**Condition 5**. In this condition, we randomly selected 25 images from the prepossessed dataset as the output.

**Condition 6**. In this condition, we performed a search on a subset of Rico dataset created by Huang et al. [18] to obtain the most similar images to the input image, based on the perceptual distance described in section 5.2. The top 25 similar images were used as the output.

## 5.4 Results

Among all the sampled inputs, both Condition 1 and Condition 2 have resulted in an average of 19 output images (SD = 7.2 and 5.15, respectively). A comparison of samples of five outputs from each experimental condition for one input image is shown in Figure 7. Figure 6 shows the distributions of the similarity and diversity metrics on the six conditions. Kruskal-Wallis tests indicated statistically significant differences among the experimental conditions with respect to both metrics (p < 0.001). We then conducted posthoc pairwise analyses using the Mann-Whitney U test with

Bonferroni correction to identify the conditions that contributed to the difference; Bonferroni correction was used to address multiple comparisons. The posthoc analyses revealed the following results.

- The GANSpiration-based methods (Conditions 1, 2, 3, and 4) have resulted in significantly lower similarity (p < 0.001) than Condition 6 (i.e., search-based approach). Conditions 3 and 4 (i.e., generation + search) also resulted in significantly higher similarity (p < 0.05), thus relevance, than random examples (Condition 5).
- While the GANSpiration-based methods (Conditions 1, 2, 3, and 4) have resulted in significantly lower diversity (p < 0.001) than Condition 5 (i.e., random examples), they have also achieved significantly higher diversity (p < 0.01) than similarity-based approach (Condition 6).

To understand how the complexity of the input UI design can influence the similarity and diversity metrics of the six experiment conditions, we separated the 27 sampled inputs into three groups: (1) Low complexity inputs contain less than six unique component types. In our sample, they often represent login screens, setting menu screens, or screens that communicate a single piece of information (see Figure 4). (2) Medium complexity inputs contain between six to eight unique component types. In our sample, they often represent screens that contain heterogeneous information (usually presented in lists or cards) or screens that provides multiple options to users. (3) High complexity inputs contain more than eight unique component types. In our sample, they often represent screens that contain complex interaction mechanisms, including tabs, multiple options on list items, maps or complex forms. Figure 8 presents the distributions of the similarity and diversity metrics in each input complexity group. Kruskal-Wallis tests indicated statistically significant differences among the experimental conditions with respect to both metrics in all three groups (p < 0.05). Table 1 and Table 2 present the posthoc analysis results based on pairwise Mann-Whitney U tests with Bonferroni correction. Results indicated that certain GANSpiration-based approaches (particularly Conditions 1 and 2) achieved a significantly lower similarity than Condition 6 (i.e., search-based approach) for low and high complexity inputs, but not in medium complexity inputs. Additionally, for high complexity inputs, the GANSpiration-based approaches can achieve a similar level of diversity to random suggestions (Condition 5), and significantly higher diversity than the search-based suggestions (Condition 6).

We manually inspected the outputs of the approaches used in the six experimental conditions to identify their risks and potential to support design inspiration. We found the following themes through our inspection.

- Conditions 1 and 2 (i.e., generation-only): Examples generated in these two conditions resembled real UI screenshots, but contained a lot of blurry and noisy components. However, some examples generated in these conditions contained interesting variations and alternatives to the input UI. For example, Figure 9a shows that the directly generated examples suggested alternatives on the color scheme, layout, and item details.
- Conditions 3 and 4 (i.e., generation + search): Examples generated in these two conditions contained real images that

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

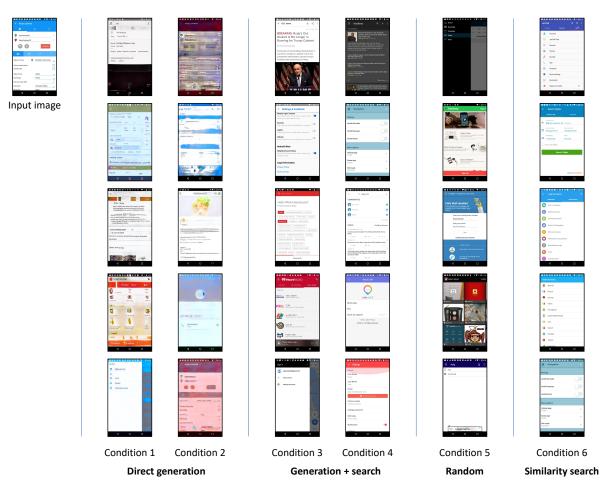
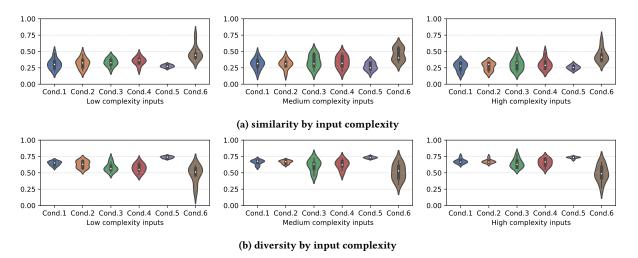
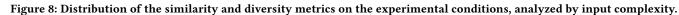


Figure 7: Samples of five outputs from each experimental condition for the input image. Samples from Condition 3 are matched with those from Condition 1, and samples from Condition 4 are matched with those from Condition 2.





## Table 1: Differences of medians (row minus column) on the similarity metric among the six conditions, by input complexity.

(a) Low complexity inputs						(b) Medium complexity inputs							(c) High complexity inputs				
	C.2	C.3	C.4	C.5	C.6		C.2	C.3	C.4	C.5	C.6		C.2	C.3	C.4	C.5	C.6
Cond.1	-0.03	-0.03	-0.07	0.03	*-0.14	Cond.1	0.00	0.00	0.00	0.07	-0.09	Con	l.1 -0.02	-0.05	-0.01	0.02	**-0.13
Cond.2		-0.00	-0.04	0.05	*-0.12	Cond.2		0.00	-0.01	0.06	-0.09	Con	1.2	-0.02	0.01	0.05	**-0.11
Cond.3			-0.04	0.06	*-0.11	Cond.3			-0.01	0.06	-0.09	Con	1.3		-0.04	0.07	-0.09
Cond.4				0.09	-0.08	Cond.4				0.07	-0.08	Con	l.4			0.03	-0.12
Cond.5					**-0.17	Cond.5					*-0.15	Con	l.5				**-0.16

Using pairwise Mann-Whitney U test with Bonferroni correction: \* p < 0.05, \*\* p < 0.01

## Table 2: Differences of medians (row minus column) on the diversity metric among the six conditions, by input complexity.

(a	) Low	comp	lexitv	inputs
----	-------	------	--------	--------

(b) Medium complexity inputs

## (c) High complexity inputs

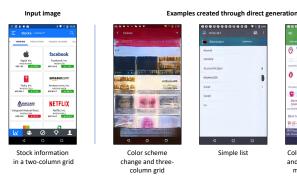
	C.2	C.3	C.4	C.5	C.6		C.2	C.3	C.4	C.5	C.6		C.2	C.3	C.4	
Cond.1	0.02	0.09	0.10	**-0.09	**0.14	Cond.1	0.01	0.04	0.06	**-0.06	0.15	Cond.1	0.00	0.04	0.00	
Cond.2		0.07	0.08	**-0.11	*0.12	Cond.2		0.04	0.05	**-0.06	0.14	Cond.2		0.04	0.00	-
ond.3			0.01	**-0.18	0.05	Cond.3			0.01	*-0.10	0.11	Cond.3			-0.04	-0
Cond.4				**-0.19	0.04	Cond.4				**-0.11	0.09	Cond.4				*-0
Cond.5					**0.23	Cond.5					**0.20	Cond.5				

Using pairwise Mann-Whitney U test with Bonferroni correction: \* p < 0.05, \*\* p < 0.01

Color scheme change

and enriched list with

more item details





Screen with tags

(c) Search results sometimes do not reflect the intention of the style-based generation results.

Textual screen

(a) Directly generated examples involves noises but can provide useful insights and suggestions.



(b) Searched examples based on generated images are cleaner and can provide useful insights and suggestions.

Figure 9: Examples demonstrating insights gained from manual inspection of outputs of GANSpiration-based approaches.

addressed the noise issues in Conditions 1 and 2. Some of the returned examples also contained interesting variations to the input image (see Figure 9b). However, some of the searched images do not reflect directly the intention of the originally generated images. For example, in Figure 9c, the generated example seems to suggest a different color scheme and a layout change, but the search resulted in somewhat irrelevant screens.

 Condition 6 (i.e., search-based approach): The outputs of this condition commonly contained visually similar UI screenshots with the input, both in terms of color and layout (see Figure 7, Condition 6).

## **6 USER EVALUATION**

To understand the potential of the GANSpiration-based techniques from the perspective of professional UI/UX practitioners, we conducted a user study focusing on the practitioners' opinions on how well the results of GANSpiration may help in their design practice.

#### 6.1 Methods

The user studies were conducted in June and July 2021. In this section, we describe our participants, the procedure of the study sessions, the materials used, and the analysis methods. The user study protocol is approved by the ethics committee at our institutions.

*Participants.* We conducted the user study with five UI/UX professionals. The participants had varying levels of experience, ranging from one to ten years as either UX researcher or UI/UX designer. They worked in different types of organizations, including two freelancers, one in a small start-up company, one in a more established medium-sized company, and one in a large multinational company. Table 3 summarizes the participants' characteristics.

Procedure and material. The user studies were conducted online via the Zoom platform. Each participation took about one hour to complete. Each study session began with a short interview in which the participants were asked about their professional background and their experience of using design examples. We then presented one input UI screenshot from the sample used in the quantitative evaluation (the one that is shown in Figure 7) to the participants. We told the participants to consider a scenario in which they want to get inspiration from examples to modify the design of this UI. This UI screenshot contained 11 different types of components, representing a complicated design task. After the participants familiarized themselves with this UI screenshot, we then presented the corresponding output images from all six conditions as the design examples, one after another; a sample of these materials was shown in Figure 7. The participants were asked to examine each set of the design examples and to provide feedback on (1) their relevance to the design task, (2) the diversity of the design examples, (3) the effectiveness of the design examples for inspiration, and (4) general positive and negative perceptions of the examples. The order of these six conditions was randomized among the participants to mitigate the order effects; the participants were also not aware of the conditions in which the examples were obtained when examining the examples. After the participants examined all six

conditions, we described the GANSpiration techniques and asked the participants about their general perception of the generated design examples.

Analysis. Two researchers watched recordings of all the study sessions and took detailed notes regarding the participants' comments and reactions. We then conducted an inductive thematic analysis [33] on the combined notes. Particularly, two researchers first independently conducted open coding to consolidate different aspects discussed by the participants. Then we held three two-hour meetings to discuss our observations and extract common themes from our codes. During this analysis, we focused on identifying the positive and negative aspects that the participants mentioned about the design examples resulting from the four categories of the experimental conditions: direct style-based generation of UI images (Conditions 1 and 2), style-based generation then search of real UI examples (Conditions 3 and 4), random selection (Condition 5), and similarity-based search (Condition 6).

### 6.2 Results

All of the participants indicated that they frequently looked for examples in their design projects. A common reason that the participants search and examine examples was to learn from the examples and borrow elements from them. For example, P3 mentioned "Every single time when I need to design an interface, I will go to all these different reference websites. I always start from there to see what a possible solution is out there. Because it saves my time not to reinvent the wheel." P2 talked about the granularity in which the design examples may help, saying "Even for a specific project, every part of the project, I need to assess what exists right now and what I can learn from what is already existing." P4 also emphasized the importance of examples by mentioning the efforts she often spends on it, saying "Every project I have to search for examples for inspiration. I will spend lots of time on it."

6.2.1 Perceptions on the directly generated examples (Conditions 1 and 2). The negative aspects of these generated examples mentioned by the participants were most concerned with the quality of the images and the noises included in the UI. For example, P3 mentioned: "It looks a bit strange to me because all the UI elements are not real UI elements. They just mimic the shapes. It doesn't provide a lot of realistic details for me to get a record for my design." P5 also declared, "If there is not any noise in the images, it is a good idea to have this technique."

Participants also discussed the potential of the generated examples. Particularly, these generated images provided enough abstraction and omitted unnecessary details to convey the design concept while eliciting creativity; this is similar to the effect of using lowfidelity prototypes such as sketches and wireframes. For example, P2 mentioned: "The idea is pretty close to what I am looking for, overall, conceptually... Since they are all pretty abstract, they are already helping me more, so that I can see the shapes and stuff ... to think out of the box." Additionally, the collections of generated images were perceived as a gateway to gather ideas of a wider range of design styles, presented in the context of the designer's work. For example, P3 said: "If it can generate a lot of good information online that we don't have time to look for, and it manages to make

ID	Job title	Organization description	Years of experience	Projects contributed to
P1	UX researcher	Freelancer	1 year	1 project
P2	UI/UX designer	A small-sized company developing software solutions that allow the creation of printable personalized products	8 years	>10 projects
P3	UI designer	A medium-sized company developing a cloud-based computer-aided design (CAD) software	3 years	10 projects
P4	UI designer	Freelancer	10 years	>10 projects
P5	UX researcher	A large multinational company developing business management software	2 years	2 projects

#### Table 3: Characteristics of participants in the user study

a merge of all the relevant interface, I can use it as a reference to see the trend on color palette and layouts."

6.2.2 Perceptions on the examples created from generation and then search (Conditions 3 and 4). All but one of the participants agreed the examples can be effective for inspiration. Echoing our quantitative results, participants appreciated that the examples are both diverse and relevant. For instance, P1 mentioned: "It is diverse, we have different designs... They are also relevant since I can see some of the same patterns.." Focusing on the diversity of the examples on the level of both the overall UI structure and the specific components, P5 said, "These are effective for inspiration as example images. Since they include both text and big image and different kinds of components inside." Participants also noticed some of the components in the images could be inspiring; those components are not in the original input image but are relevant to the design task. P4 mentioned, "Image number 14 at the top right there is an alert icon which I can use in my design and Number 3 also has a search icon." P3 also stated, "There are some images which have tabs, which I can use as references for the tab of the main image, and also the toggle switch, radio buttons, ..."

The participant (P2) who thought the examples are not effective for inspiration was mainly concerned that the overall design style of the output examples is homogeneous and a bit outdated. She mentioned: "It is all the same [Google] Material Design style and nothing exciting - it is very similar to what I would work on already [in the input image]." This comment pointed out the limitations of using UI screenshots from the same platform for inspiration. Some participants also voiced other concerns related to the current GANSpiration-based techniques. For example, P3 mentioned that she wished to have the examples contain more UI screenshots from the same page type and the same application domain: "I did not find any image that included extended forms or any transportation apps." P1 and P4 also would like to see more recent and modern designs in the examples; this comment, along with P2's concerns, highlighted the important role of the dataset used for generating the examples and searching for existing designs.

6.2.3 Perceptions on the randomly selected examples (Condition 5). All participants were impressed by the obvious diversity of the examples. For instance, P2 stated, "I like that the color schemes are getting different. It has different layouts from different stuff." However, they also noticed the randomness of the examples. P3 said, "There are a lot of things that are not relevant here. Some of the examples are definitely noises." P2 also said, "I do think that those layouts, that are different, are not related to the design I am looking for."

6.2.4 Perceptions on the examples selected from similarity-based search (Condition 6). All participants agreed that the examples are similar to the source image, but not diverse enough for inspiration. P3 said, "The first thing that I noticed is the similar color palette and elements." P2 also said, "All of them are lists and have the same colors, not very helpful but relevant." P4 also mentioned, "Although the information is clear and easy to use, [the examples are] too boring."

Overall, the participants considered the GANSpiration-based approaches, particularly the ones that output full-fledged UI screenshots (i.e., conditions 3 and 4), as viable ways to gain inspiration in their practical design workflows. Participants mentioned that the output examples from these approaches can help them widen their perspective and obtain immediate access to relevant and diverse inspirational sources. For example, P1 indicated "This tool would help in the competitive analysis part. If it has more trendy images it could be really helpful." Some participants also commented on the ability of our approaches in helping the designers think differently and unconventionally. For example, P2 stated that "I could use this technique... to think out of the box."

## 7 DISCUSSION

In both the quantitative evaluation and the qualitative user study, we found that GANSpiration-based approaches are able to generate design examples that are both diverse and relevant to the screenshot of the input design. The quantitative results indicated that our approach can generate more relevant examples than random suggestion, while providing more diverse examples than similarity-based suggestion. In the user study, our participants commented on the balance between diversity and relevancy as a desirable attribute. Participants generally preferred the GANSpiration-based approaches, particularly the ones that suggests full-fledged UI screenshots based on the generated examples, over random examples and similaritybased examples. The user evaluation also indicated that our approach is able to help designers broaden their horizons and get inspired. While the user study focused on an in-depth examination of one input image (representing a complicated design task), the quantitative evaluation covered a wider range of design complexity and indicated that our approach may work particularly well

for low complexity and high complexity inputs. Overall, these results demonstrated that GANSpiration can help achieve an intricate balance between design drift and design fixation when providing examples for inspiration in the challenging context of user interface design. In the rest of this section, we discuss the implications of our study results for designing tools that leverage generative techniques such as GANSpiration for design inspiration.

## 7.1 Style-based image generation provides inspiration at different granularity levels

Our user study results indicated that the style-based generation technique adopted by GANSpiration is able to provide design inspiration on three levels: (1) the coarse level that provides ideas for layout or structural changes, (2) the middle level that suggests component design alternatives, and (3) the fine level that proposes different aesthetics such as color schemes. This is made possible by the ability of StyleGAN to alter the input image based on different granularity levels (i.e., different spatial resolutions) of the target 'style images'. The three aspects of inspiration were all appreciated by the participants during the user study. The tool design that leverages GANSpiration-based approaches can explicitly incorporate these levels of design inspiration. Particularly, a design inspiration tool can indicate and explain the intention of the suggested examples by checking the granularity level of the style image used for style merge. Based on this information, a descriptive label of inspiration granularity can be assigned to each example. This way, if the designers have a particular concern when searching for inspiration (e.g., need to find a different layout but using the same color scheme), they would be able to better focus on such examples. Additionally, it is also possible to give the designers control over the granularity level of style merge.

# 7.2 The visual quality of the generated image is an important factor for inspiration

Our results revealed that the participants preferred full-fledged UI screenshots over directly generated images that typically include noises, although the direct generation and the generation-thensearch conditions achieved the same level in the diversity and relevance metrics. While some participants were impressed that the directly generated images look like a UI, our results indicated that the visual quality of the generated images does affect how well the designers perceive the examples. In this study, we retrieved the most visually similar UI screenshots to the directly generated images to address this issue. However, our manual inspection revealed that the search results based on the generated images do not always match the intention of the directly generated results. To further resolve these issues, techniques for identifying components on generated images could be investigated. With such techniques, the quality of the individual components can be improved to make the generated images look more similar to real ones. Further, if components were identified, the directly generated images can be converted to wireframes in order to provide layout or structural suggestions. Color schemes of the components can also be matched to the input image to provide direct alternatives.

## 7.3 A diverse and relevant training dataset would help generate more insightful examples

Our study relies on the Rico dataset, which is created in 2017. Some of our participants voiced that the examples retrieved do not reflect the most recent design trends, thus limit the inspirational power of the examples. This result indicates that the dataset used for training the generative model, as well as the dataset used for retrieving real UI screenshots, are important factors to consider. Potential solutions to this problem include using only the newest apps in the dataset and collecting datasets from the most recent design sharing platforms (e.g., dribbble.com). Additionally, using a merged dataset including UI screenshots from different platforms (e.g., Android, iOS, desktop app, web app, etc.) with different design frameworks/systems [29] would help avoid platform or frameworkspecific design stereotypes.

# 7.4 Combine generative models with other techniques

In our experimental design, conditions 3 and 4 (i.e., generation + search) are our first attempts to combine the pure generative approach with other techniques to provide examples for effective inspiration. These attempts can be further enriched and expanded. Particularly, our participants seemed to desire examples from the same application domain (e.g., route planning) and/or focusing on the same type of page (e.g., configuration page) as the input UI page. Thus, it could be useful to perform the style-merging generation using the style images from the same application domain and/or page type as the input image. Moreover, combining the generative technique with textual thematic specifications that describe characteristics of the desired examples (e.g., the ones similar to [27]) would give designers more control over the returned examples and support a more effective inspiration. Finally, techniques that can highlight interesting areas in the examples related to the suggested alternatives to the input image would also facilitate a more efficient exploration for serendipitous and targeted inspiration.

## 8 THREATS TO VALIDITY AND LIMITATIONS

As mentioned before, the dataset we used in this study (i.e., the Rico dataset) is published about five years ago and only included screenshots of Android applications. Although it is a large dataset that is frequently used in studies involving UI design artifacts, we do not know the inspirational power of GANSpiration-based approaches if a newer dataset or a dataset on another platform was used. We recognize that building a large-scale UI dataset is a non-trivial task. Even with the old dataset, our study demonstrated the potential of GANSpiration in supporting both serendipitous and targeted inspirations.

Moreover, our approach treated user interface design as static images. We acknowledge that the design artifacts can be presented in many other formats (e.g., a tree of UI components [34]) and may include other information such as animations and page transitions. We chose to focus on static images because they are one of the most frequent types of inspirational sources currently used by the designers, supported by tools such as design galleries (e.g., Dribbble). CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

Mohammad Amin Mozaffari, Xinyuan Zhang, Jinghui Cheng, and Jin L.C. Guo

Future work may investigate how similar approaches can be developed to incorporate more structured and richer representations of user interface design artifacts.

Additionally, our user study only included five participants. Moreover, while the evaluation was based on a realistic scenario, it only included one input UI and may not be able to incorporate all the real-world inspirational challenges related to UI design. The small sample size is partially due to the challenges we experienced in recruiting participants during the pandemic. However, our participants were all professional practitioners and represented diverse UI/UX-related expertise. With their professional experiences, they also reflected on their practice when examining the examples during the user study, providing a real-world perspective.

Finally, we acknowledge that UI/UX design involves complex workflows, diverse tasks, and various considerations. We do not claim that our work can possibly address all these aspects. Instead, we targeted the particularly challenging problem of design inspiration and demonstrated the potential of a creativity support approach for this context that leverages a style-based generative machine learning technique. We encourage future work to investigate the applicability of this line of techniques in other tasks and aspects of UI/UX design.

## 9 CONCLUSION

Our proposed GANSpiration-based approaches aim to provide concrete creativity support by balancing both targeted and serendipitous inspiration in user interface design. The evaluation studies highlighted the capacity of GANSpiration in generating examples that are both relevant to the designers' work at hand and diverse for avoiding design fixation. Professional UI/UX practitioners appreciated such techniques as viable support in their day-to-day design practice. Our results also revealed possible improvements and design implications when such a generative technique is used for supporting design inspiration. Overall, our work demonstrates the potential of applying style-based generative machine learning techniques in the challenging context of design inspiration and creativity support. It opens a new direction and paves the road for future efforts in using advanced intelligent technology for supporting the creative, but at the same time constraint, design practice.

## ACKNOWLEDGMENTS

We thank the participants for their time and valuable insights. The project is partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant Program [RGPIN-2018-04470] and Fonds de Recherche du Québec – Nature et technologies (FRQNT) Team Research Project Grant [2022-PR-299099].

## REFERENCES

- [1] Behance. 2022. Behance Search Projects: Photos, videos, logos, illustrations and branding. https://www.behance.net
- [2] Farnaz Behrang, Steven P Reiss, and Alessandro Orso. 2018. GUlfetch. In Proceedings of the 5th International Conference on Mobile Software Engineering and Systems - MOBILESoft '18. ACM Press, New York, New York, USA, 236–246. https://doi.org/10.1145/3197231.3197244
- [3] Tony Beltramelli. 2018. Pix2Code: Generating Code from a Graphical User Interface Screenshot. In Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems (Paris, France) (EICS '18). ACM, New York, NY, USA, Article 3, 6 pages. https://doi.org/10.1145/3220134.3220135

- [4] Derya Birant and Alp Kut. 2007. ST-DBSCAN: An algorithm for clustering spatial-temporal data. Data & knowledge engineering 60, 1 (2007), 208–221.
- [5] Nathalie Bonnardel. 1999. Creativity in design activities. In Proceedings of the third conference on Creativity & cognition - C&C '99. ACM Press, New York, New York, USA, 158–165. https://doi.org/10.1145/317561.317589
- [6] Sara Bunian, Kai Li, Chaima Jemmali, Casper Harteveld, Yun Fu, and Magy Seif Seif El-Nasr. 2021. VINS: Visual Search for Mobile User Interface Design. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (Yokohama, Japan) (CHI '21). ACM, New York, NY, USA, Article 423, 14 pages. https://doi.org/10.1145/3411764.3445762
- [7] Chunyang Chen, Ting Su, Guozhu Meng, Zhenchang Xing, and Yang Liu. 2018. From UI Design Image to GUI Skeleton: A Neural Machine Translator to Bootstrap Mobile GUI Implementation. In Proceedings of the 40th International Conference on Software Engineering (Gothenburg, Sweden) (ICSE '18). ACM, New York, NY, USA, 665–676. https://doi.org/10.1145/3180155.3180240
- [8] Chunyang Chen, Ting Su, Guozhu Meng, Zhenchang Xing, and Yang Liu. 2018. From UI Design Image to GUI Skeleton: A Neural Machine Translator to Bootstrap Mobile GUI Implementation. In Proceedings of the 40th International Conference on Software Engineering (Gothenburg, Sweden) (ICSE '18). ACM, New York, NY, USA, 665–676. https://doi.org/10.1145/3180155.3180240
- [9] Biplab Deka, Zifeng Huang, Chad Franzen, Joshua Hibschman, Daniel Afergan, Yang Li, Jeffrey Nichols, and Ranjitha Kumar. 2017. Rico: A Mobile App Dataset for Building Data-Driven Design Applications. In Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (Québec City, QC, Canada) (UIST '17). Association for Computing Machinery, New York, NY, USA, 845–854. https://doi.org/10.1145/3126594.3126651
- [10] Dribbble. 2022. Dribbble Discover the World's Top Designers and Creative Professionals. https://dribbble.com/
- [11] Claudia Eckert and Martin Stacey. 2000. Sources of inspiration: a language of design. Design Studies 21, 5 (2000), 523-538. https://doi.org/10.1016/S0142-694X(00)00022-3
- [12] Milene Gonçalves, Carlos Cardoso, and Petra Badke-Schaub. 2014. What inspires designers? Preferences on inspirational approaches during idea generation. *Design Studies* 35, 1 (2014), 29–53. https://doi.org/10.1016/j.destud.2013.09.001
- [13] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative Adversarial Nets. In Proceedings of the 27th International Conference on Neural Information Processing Systems Volume 2 (Montreal, Canada). MIT Press, Cambridge, MA, USA, 2672–2680.
- [14] Yasunari Hashimoto and T Igarashi. 2005. Retrieving web page layouts using sketches to support example-based web design. In Proceedings of 2nd Eurographics Workshop on Sketch-Based Interfaces and Modeling. Eurographics Association, Goslar, DEU, 10.
- [15] S.R. Herring, B.R. Jones, and B.P. Bailey. 2009. Idea Generation Techniques among Creative Professionals. In Proceedings of the 42nd Annual Hawaii International Conference on System Sciences, HICSS. IEEE Computer Society, Los Alamitos, CA, USA, 1–10. https://doi.org/10.1109/HICSS.2009.241
- [16] Scarlett R. Herring, Chia-Chen Chang, Jesse Krantzler, and Brian P. Bailey. 2009. Getting Inspired! Understanding How and Why Examples Are Used in Creative Design Practice. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Boston, MA, USA) (CHI '09). Association for Computing Machinery, New York, NY, USA, 87–96. https://doi.org/10.1145/1518701.1518717
- [17] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. 2017. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. In Proceedings of the 31st International Conference on Neural Information Processing Systems (Long Beach, California, USA). Curran Associates Inc., Red Hook, NY, USA, 6629–6640.
- [18] Forrest Huang, John F. Canny, and Jeffrey Nichols. 2019. Swire: Sketch-based User Interface Retrieval. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (Glasgow, Scotland Uk) (CHI '19). ACM, New York, NY, USA, Article 104, 10 pages. https://doi.org/10.1145/3290605.3300334
- [19] David G. Jansson and Steven M. Smith. 1991. Design fixation. Design Studies 12, 1 (1991), 3–11. https://doi.org/10.1016/0142-694X(91)90003-F
- [20] T. Karras, S. Laine, and T. Aila. 2019. A Style-Based Generator Architecture for Generative Adversarial Networks. In 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, Los Alamitos, CA, USA, 4396–4405. https://doi.org/10.1109/CVPR.2019.00453
- [21] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila. 2020. Analyzing and Improving the Image Quality of StyleGAN. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, Los Alamitos, CA, USA, 8107–8116. https://doi.org/10.1109/CVPR42600.2020.00813
- [22] Yann LeCun and Yoshua Bengio. 1998. Convolutional Networks for Images, Speech, and Time Series. MIT Press, Cambridge, MA, USA, 255–258.
- [23] Brian Lee, Savil Srivastava, Ranjitha Kumar, Ronen Brafman, and Scott R. Klemmer. 2010. Designing with interactive example galleries. In Proceedings of the 28th international conference on Human factors in computing systems - CHI '10. ACM Press, New York, New York, USA, 2257. https://doi.org/10.1145/1753326.1753667

- [24] Richard L. Marsh, Joshua D. Landau, and Jason L. Hicks. 1996. How examples may (and may not) constrain creativity. *Memory & Cognition* 24, 5 (sep 1996), 669–680. https://doi.org/10.3758/BF03201091
- [25] Kevin Moran, Carlos Bernal-Cárdenas, Michael Curcio, Richard Bonett, and Denys Poshyvanyk. 2020. Machine Learning-Based Prototyping of Graphical User Interfaces for Mobile Apps. *IEEE Transactions on Software Engineering* 46, 2 (2020), 196–221. https://doi.org/10.1109/TSE.2018.2844788
- [26] Dmitry Nikitko. 2019. StyleGAN-Encoder. https://github.com/Puzer/styleganencoder.
- [27] Daniel Ritchie, Ankita Arvind Kejriwal, and Scott R. Klemmer. 2011. d.tour: style-based exploration of design example galleries. In Proceedings of the 24th annual ACM symposium on User interface software and technology - UIST '11. ACM Press, New York, New York, USA, 165. https://doi.org/10.1145/2047196.2047216
- [28] Pao Siangliulue, Joel Chan, Krzysztof Z. Gajos, and Steven P. Dow. 2015. Providing Timely Examples Improves the Quantity and Quality of Generated Ideas. In Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition -C&C '15. ACM Press, New York, New York, USA, 83–92. https://doi.org/10.1145/ 2757226.2757230
- [29] Marco Suarez, Jina Anne, Katie Sylor-Miller, Diana Mounter, and Roy Stanfield. 2022. Design Systems Handbook. InVision, New York, NY, USA.
- [30] Amanda Swearngin, Mira Dontcheva, Wilmot Li, Joel Brandt, Morgan Dixon, and Andrew J. Ko. 2018. Rewire. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18. ACM Press, New York, New York, USA, 1–12. https://doi.org/10.1145/3173574.3174078
- [31] Todd M. Thrash and Andrew J. Elliot. 2003. Inspiration as a Psychological Construct. Journal of Personality and Social Psychology 84, 4 (2003), 871–889.

https://doi.org/10.1037/0022-3514.84.4.871

- [32] Todd M. Thrash, Emil G. Moldovan, Victoria C. Oleynick, and Laura A. Maruskin. 2014. The psychology of inspiration. *Social and Personality Psychology Compass* 8, 9 (2014), 495–510. https://doi.org/10.1111/spc3.12127
- [33] Mojtaba Vaismoradi, Hannele Turunen, and Terese Bondas. 2013. Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. Nursing & Health Sciences 15, 3 (2013), 398–405. https://doi.org/10.1111/ nhs.12048 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/nhs.12048
- [34] Jason Wu, Xiaoyi Zhang, Jeff Nichols, and Jeffrey P Bigham. 2021. Screen Parsing: Towards Reverse Engineering of UI Models from Screenshots. Association for Computing Machinery, New York, NY, USA, 470–483. https://doi.org/10.1145/ 3472749.3474763
- [35] Mulong Xie, Sidong Feng, Zhenchang Xing, Jieshan Chen, and Chunyang Chen. 2020. UIED: A Hybrid Tool for GUI Element Detection. In Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering (Virtual Event, USA) (ESEC/FSE 2020). ACM, New York, NY, USA, 1655–1659. https://doi.org/10.1145/3368089.3417940
- [36] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang. 2018. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, Los Alamitos, CA, USA, 586–595. https://doi.org/10.1109/CVPR.2018.00068
- [37] Tianming Zhao, Chunyang Chen, Yuanning Liu, and Xiaodong Zhu. 2021. GUIGAN: Learning to Generate GUI Designs Using Generative Adversarial Networks. In 2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE). IEEE Computer Society, Los Alamitos, CA, USA, 748–760. https://doi.org/10.1109/ICSE43902.2021.00074