Producing 2D Asset Libraries with AI-powered Image Generators

Phillip Fernberg[1], Benjamin H. George[2], Brent Chamberlain[1]

[1] Utah State University, Utah/USA
[2] Utah State University, Utah/USA · benjamin.george@usu.edu

Abstract: A proliferation of Artificial Intelligence (AI) applications specific to landscape architecture has revealed potential disruptions to many aspects of the professional design process, including tasks that require creative skills but are very time-consuming. Creating 2D assets for design renderings is an example of one such task, requiring an inordinate amount of time to create just a few image cut-outs with little customization. Generative AI art tools offer the possibility to both reduce production time and improve the quality and customizability of asset libraries. In this paper, we present a comparative assessment of three image generators' (Dall-E 2, Midjourney, and Stable Diffusion) abilities to produce 2D asset libraries. The analysis includes the strengths and weaknesses of each generator in accuracy, usability, and artistic style. Recommendations for potential prompts and workflows to achieve desired results with each generator are also provided, along with a reflection on the greater implications of generative AI for landscape practice.

Keywords: Generative AI, landscape design, design representation, machine learning

1 Introduction

While Artificial Intelligence (AI) has been growing steadily for many decades, recent advancements in applications specific to the architecture, construction, and engineering (AEC) industry have revealed potential evolutions for many parts of the landscape design process. These include tasks such as site inventory, landform modeling, or conceptual urban design, among others (BARBARASH et al. 2022, BRÜTTING 2020, JOHNS et al. 2020, LIU & TIAN 2022, RAMAN et al. 2022). Many of these tasks require unique expertise, skills, cognition, and creative agency and can be an expensive, even if vital, portion of design services. On the other hand, there are many production tasks that are less impactful but typically still require the skills of a designer to complete. AI may be able to offload some of these secondary activities, improving the trade-off between their costs and quality. At the same time, integrating AI into design workflows may also be perceived as a disruption of traditional (creative) processes such as design representation. For instance, the recent release of descriptive image generators such as DALL-E 2, Midjourney, or Stable Diffusion could dramatically change the way designers obtain and inventory image assets. With descriptive prompts from a human user, these generators utilize machine learning models trained on millions of images to produce two-dimensional renderings in nearly any conceivable artistic style in less than a minute. While the ability to autonomously create entire artistic scenes for design ideation is striking, a more intriguing and perhaps pragmatic use for these generators is the production of 2D asset libraries to be used for typical design renderings.

Many design practices build internal entourage libraries to facilitate rapid conceptual rendering production using online resources such as ARTCUTOUT, pngimg.com, or SKALGUBBAR. While useful, these libraries are often limited in materials, styles, and diversity of assets. This is especially evident when trying to create bespoke representations that evoke a

Journal of Digital Landscape Architecture, 8-2023, pp. 186-194. © Wichmann Verlag, VDE VERLAG GMBH · Berlin · Offenbach. ISBN 978-3-87907-740-3, ISSN 2367-4253, e-ISSN 2511-624X, doi:10.14627/537740020.

localized sense of place, either with native plant materials, textures, implementations, or people. AI image generators offer the potential to overcome these limitations through ongoing, rapid, and near-infinite production of assets that can be refined with increasingly specific command prompts – at an increasingly lower price. In this paper, we present a comparative assessment of three image generators' abilities to produce 2D asset libraries. Our analysis includes the strengths and weaknesses of each generator in accuracy, usability, and artistic style. We also provide a reflection on potential prompts and workflows for most quickly achieving desired results with each generator.

2 Assessment of Three AI-image Generators

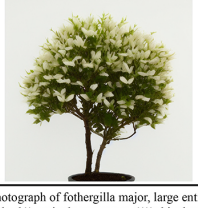
2.1 Methods

The three AI image generators assessed as part of this research were Dall-E 2, Midjourney, and Stable Diffusion. These were selected based on their proven high-quality outputs, large user base, and usability factors. Dall-E 2 and Midjourney are paid subscription services that operate on cloud-based servers, while Stable Diffusion is run locally. All three support both text-based and image-based prompts, though other features vary, such as inpainting (AI redrawing portions of an image), outpainting (AI generating imagery around the periphery of an image), program modifications (creation of supplemental features or AI generation models), and seamless tiling (repeatable patterns).

All three services were assessed for their usefulness in aiding the creation of 2D representations through the rapid creation of unique plants and entourage assets, which are discrete image elements that can be included in a larger image. The researchers attempted to create several near photo-realistic target assets in the different categories of trees, shrubs, perennials, people, and objects. We also attempted to create several assets in various artistic styles. Over 300 total assets in the above categories were generated for this study. We abandoned the process of creating an asset once 15 prompts had been used without a successful asset being generated, as at this point it typically required between 20-30 minutes and would not be considered feasible from a financial and resource perspective. Once a usable asset had been generated, there was further evaluation of the ability to reliably produce usable variations, the ease of accessing the image files, and the ease of sharing with limited post-production (<2 minutes of masking or background deletion). Finally, as a result of this work we supply best practices for efficiently generating assets in all three AI image generators.

Prompts are the textual guide for the AI to generate an image and should include specific details for the desired image. In addition to describing the target subject (ie. a specific plant species), the prompts can include contextual information to guide the AI in crafting the visual style of the image. For instance, adding '8k' will encourage the AI to create a highly detailed image, or "3d game asset" will encourage an image with a 3-dimensional, digital feel. Additionally, Midjourney and Stable Diffusion offer different functions within the prompts. For instances, in Stable Diffusion the addition of () bracketing will cause Stable Diffusion to assign greater weight to that word, and multiple parens can be added to further increase the weight of a word. In Midjourney, specific commands can be utilized to alter the AI model, for instance adding – test will invoke Midjourney's artistic mode. A full list of commands for each service can be found in their respective documentation and users will benefit from becoming familiar with them.

Table 1: Examples of text prompts, image results, and performance from each generator for typical assets used in 2D renderings (trees, shrubs, people, etc.)

		Dall-E 2	Midjourney	Stable Diffusion
Trees	Image Output			
	Text Prompt (steps)	photorealistic quercus gambelii on white background, autumn, realistic, detailed, 8k, full tree, portrait (3)	a photograph of a gambel oak tree in autumn on white background, realistic, detailed, 8k –test –upbeta (4)	photograph of gambel oak autumn color, 3d texture map atlases, quixel megascans, tree branch, white background (9)
	Observation	Reliably produced a photorealistic asset of most varieties. Prompts can be added to customize the image such as: autumn color, gnarled, no leaves, etc.	Reliably produced a useable asset within three or fourRiterations	Reliably produces a useable asset. Adding a rudimentary image prompt using the img2img feature is effective at producing a useable image quickly.
Shrubs	Image Output			
	Text Prompt	fothergilla major full shrub isolated (4)	photograph of fothergilla major, entire large plant, realistic, detailed, isolated on white background (15)	photograph of fothergilla major, large entire ((shrub)), quixel megascans, ((((white background)))) (10)
	Observation	Used very reliably with most shrubs although multiple iterations may be required to find a seed that produces an uncropped shrub image.	Midjourney was unreliable in producing isolated images of shrubs	Produced a reliably useable asset, though the accuracy was sometimes lacking
Perennials	Image Output			
	Text Prompt	penstemon cyananthus whole plant on white background (1)	photograph of wasatch blue penstemon, highly detailed, isolated white background –test –upbeta	photo of wasatch blue penstemon plant, isolated, ((white background))
	Observation	used reliably with many perennials. Utilizing inpainting produces more reliable results.	Midjourney was unreliable in producing isolated images of perennials. (15)	Stable Diffusion was unreliable in producing isolated images of perennials. (15)
People	Image Output			
	Text Prompt	portrait photograph of woman wearing red shirt walking small dog, white background, isolated (1)	highly detailed photorealistic render of woman walking small dog, octane render, white background, isolated (15)	photograph of a full body shot woman ((walking)) a small dog, (((isolated))), entourage asset, ((white background)) (15)
	Observation	use of portrait in the prompt will increase the likelihood of returning an uncropped image	Midjourney was unreliable in producing isolated images of people	Stable Diffusion was unreliable in producing isolated images of isolated people

2.2 Results

Based on the image outcomes (see Table 1), Dall-E 2 is overall the most reliable AI image generator for the creation of high-quality visual assets across all the categories, on average producing the highest quality images with the fewest prompts required (identified as steps in table 1). Additionally, Dall-E 2 reliably produced usable images utilizing very similar prompts for each asset and was excellent at creating variations (see Figure 1). Midjourney required a far greater number of iterations on prompts to produce useable assets and many times failed to produce a viable asset. Generally, it was effective at generating trees and objects, but underperformed in production of shrubs, perennials, and people, often returning close-ups of leaves and flowers, or overly complex artistic compositions of people.

However, Midjourney was excellent at producing stylized assets (e. g. urban sketcher, impressionist, etc.). Once a usable prompt was developed, all three platforms were effective at generating reliable variations, but Midjourney produced the widest variations in images. However, Midjourney also produced the highest quality images with a more effective built-in AI upscaling feature. Finally, it was determined that output images in all three platforms were equally accessible for download and use.

Fig. 1: *Juniperus scopulorum* image asset produced by Dall-E and four variations then geerated from the asset

For objects, we found that Dall-E was the most reliable at creating objects, but that both Midjourney and Stable Diffusion struggled to create isolated objects but tended to create the object within a larger artistic composition. For Dall-e, a simple prompt of “photograph of a <target object> on white background” reliably produced a usable image asset.

Fig. 2: *Q. gambelii* produced by Stable Diffusion without an image prompt (left), the image prompt (center), and result produced using the image prompt (right)

Image prompts were also found to be useful, especially in the case of Stable Diffusion. For instance, in generating a "*Quercus gambelii* with autumn color," the results in Stable Diffusion were significantly improved when given a basic image prompt of an orange circle atop a brown line on a white background (see Figure 2). While Dall-E 2 and Midjourney support similar functionality, they do not reliably produce as positive results as Stable Diffusion using image prompts. Stable Diffusion was typically effective at generating trees, and shrubs, but struggled with generating perennials, people, and objects. Notably, Stable Diffusion was the least effective at isolating the assets individually on a white background, which significantly hampers the use of the asset in a visualization workflow.

3 Discussion

The goal of this experiment was to assess the ability of AI image generators to produce reliable 2D assets that could be used for design renderings. While all three generators assessed were eventually able to achieve the desired results, they also exhibited some general distinctions in process and product. Given the above results, our observations are provided in the following subsections.

3.1 Recommendations for AI to 2D Workflows

We used AI image generation as a means of creating a specific, visually-isolated entourage that could be easily integrated into a 2d rendering using photo-editing software. Beyond the creation of an accurate and high-quality asset, the most important feature is that the asset is isolated on a clean, solid background color that can easily be removed to isolate the asset. In creating the assets, we tracked our prompts to try and identify particular prompts that would reliably create new assets with little modification of the prompts. Based on our research we identified several prompts that produced reliable results, and several keywords that could commonly be added to fine-tune a prompt. These prompts are included in Table 1.

3.2 Overall Assessment of Three Generators

***Dall-E 2**:* For the pure task of producing isolated, usable images, Dall-E 2 exhibited the most responsiveness to and greatest understanding of prompts. In other words, it was the most apt at intuiting the desired result from the language provided and thus the prompts should be clear, simple, and segmented for maximum efficacy. Dall-E 2 also seemed to produce the most usable photorealistic versions of people assets. The products might not have always been the highest quality of photorealism compared to those of its counterparts, but Dall-E 2 was the best at isolating assets for easy transfer into a 2D rendering task. Where Dall-E struggles is taking the desired products and reiterating them with different aesthetics. It does what you want but will not be initially responsive or flexible with how you want it. Knowing this, Dall-E 2 seems to be the superior choice as a sort of asset library workhorse.

Midjourney: Midjourney is by far the best generator for artistic representation at the time of writing. Its fluidity in iteration on anything from photorealism, to mid-century modern, to renaissance styles is, so far, ahead of the pack. However, Midjourney's initial inability to stray from the more artistic or illustrative into rawer, isolated, and usable imagery means that 2D asset creation will take more effort and might still fail to produce desired results. These points seem to suggest that Midjourney is best used for more exploratory or evocative design exercises like producing mood boards, ideating conceptual motifs, or storyboarding, activities in which the program renders some uncannily stunning results.

Stable Diffusion: Even if Stable Diffusion did not match the overall performance of Dall-E 2 or Midjourney, it excelled in utilizing image prompts in combination with text prompts, which could be reliably used to reduce the number of prompts needed. This is an important quality, as a hybridized image-text prompt tool that is responsive holds near unbounded potential. There are many design precedents in the world that are not yet textually or linguistically recorded, let alone integrated into a natural language model like those used by AI-generators, but that have a clear visual motif that can be replicated. If a user could use the prompt tool consistently and with effective description, Stable Diffusion could help them generate new, heretofore "unnamed" design vernaculars and then integrate them into their practice's asset library.

3.3 Possibilities (and Challenges) for Pedagogy and Practice

As we were conducting this research, the authors simultaneously experimented with integrating generative AI in the classroom. Combining the more structured approach of our study with anecdotal information from parallel teaching efforts offers some insights for issues, possibilities, or challenges with using AI-driven applications in both design pedagogy and practice. The results of both efforts suggest that AI image generators are a viable medium through which students and young practitioners can achieve rapid generation of large asset libraries in a comparatively shorter time than traditional methods and in many instances with higher quality results. Typical workflows include drawn out web searches and perusing social media sites like Pinterest for pre-made .png images that do not always fit aesthetic or functional needs of a project (e. g. specific plant species, objects, views or actions of people, etc.). A lack of fit would then cause students to grab "next best" images and spend extra time masking or isolating desired elements in post-production. AI image generators, Dall-E 2 in particular, seem to drastically reduce the disjointedness of these workflows allowing greater customizability, more rapid iteration if image needs are not immediately met, and fewer post-produc-

tion needs so long as the outputs properly isolate elements. The process of creating these libraries also has the potential to reinforce student knowledge of design details such as materials, plant species/genus/common name/structure/etc. or stylization (see Figure 3), and help them know how to describe said details in both broad and succinct "terms" as they iterate on the semantic abilities of the AI-driven language models that power the text prompters.

Fig. 3: Assets produced by students in Dall-E 2 using different stylizations

All of these same benefits can apply to professional practice, which for all intents and purposes is an extension of the academic learning environment with the added concreteness of using the image assets created for real-world, consultant-quality deliverables. After all, entry-level designers, often new grads, carry a disproportionate amount of production work that involves both the creation and implementation of 2D image assets in crafting renderings for conceptual or schematic packages. It is they who are at once tasked to learn as much as possible about the office's approach to design, materials use, project management, and then project those values to clients and the public in the form of two-dimensional imagery. Generative AI applications could help the industry's emerging professionals better achieve this task by mitigating cognitive load (Robinson 2019) and time expenditure – which in turn affords them more mental energy and critical thinking capacity to focus on the more complex design aspects of a given project. In other words, a dramatically reduced and simplified production workload can breed more energetic, more adept, and therefore happier and impactful young designers who will be more likely to stay in the landscape architecture workforce.

However, while there are many great possibilities for integrating AI image generators into teaching and practice, it is also important to highlight their possible limitations and challenges. First, the success of these generators in producing desirable and actionable images for the specific task of building 2D libraries is largely dependent on the quality and consistency of the input data – i. e. the text prompts and base images. As the programmer George Fuechsel's attributed proverb goes, "garbage in, garbage out." If a designer does not quickly get a sense of what verbiage elicits the best response from the AI or does not properly put themselves "on the loop" to help the AI see desired patterns, they may end up spending just as much time and energy producing assets as they would using traditional methods. Similarly, if the use of the generator is not purposeful or task-oriented, the user could also find themselves distracted from the more meaningful problems of a project by the alluring creative capabilities of the generator, exploring representational possibilities in the same way they

would a social media site instead of working to 'create' those possibilities through their mastery of the project's actual design needs. Social media algorithms like those of Pinterest are known to induce this type of "curatorial" labor and distraction (LABBAN & BIZZI 2022, SCOLERE & HUMPHREYS 2016). While AI image generators like those produced by OpenAI free users from the attention-seeking nature of recommender algorithms, they should be considered no different in their potential to be a sort of representational Potemkin village.

Moreover, the development of generative AI art tools is nascent, with most applications having come to market in the past two or three years. Such nascence has upsides – such as the fact that the underlying computer vision and machine learning models will only improve over time as more users provide feedback and the internet provides more open data – but it also has downsides, such as the potential for the models to build in bias or become lopsided in their outputs based on who is using them (e. g. if the vast majority of MidJourney users are fantasy book illustrators, Midjourney may begin to center its production patterns on aesthetics and functions of those users). Furthermore, the art and design worlds have yet to establish the commercial landscape of AI-produced imagery – which in the case of the generators in question is not technically original imagery but rather novel composition and stylization of open access images from the web – which leaves the legality of using AI-generated assets for design consulting contested and could in higher profile cases create problems for the more aggressive users of the medium (SMITS & BORGHUIS 2022, ZEILINGER 2021). While these problems are currently more abstract relative to the purposes of our study, they raise some important issues to reflect on as the use of AI in design practice continues to grow.

4 Conclusions and Future Research

AI image generation holds substantial promise for creating improved 2D representations of design projects. Even though the generators assessed in this study are in their infancy, they are already useable in existing projects. As these software advance in learning they may move from improving access to assets for renderings to fundamentally disrupting the creative clockworks of landscape design. The oscillations of technology adaptation and abandonment often parallel the development of skills needed to use technology. AI image generators represent a major evolution of technology that could revolutionize the role of drawing and graphics skills. Ubiquitous use could begin to suggest that graphical skills may no longer be as critical as they once were, but it would require replacing with a skill to translate a graphical vision into a semantic or ontological description that can be understood by an AI system (FERNBERG et al. 2021). Here human and machine become more intimate and symbiotic than when a designer simply commands graphical software. The ability to learn from AI and to teach AI could fundamentally shift the creative process.

In each of the instances presented in this paper, industry will certainly reflect on the costs and benefits, whether as economic terms, or perhaps, existential terms. The software tested, represent early movers which are readily useful, turn-key solutions. However, the field of rapid image generation is advancing rapidly and only represent one of many potential disruptions to the design process. Future research and discussion are needed to reflect on the topic of AI in design representation. Additional topics to explore include artistic blockbusting with features like inpainting, linguistics models as design tools, AI generators in participatory processes and community visioning, or ethical considerations such as ownership of images, in-

tellectual property, and AI copyright. We hope this case study helps open and contribute to further conversations about the role of creative AI in digital landscape architecture.

References

BARBARASH, D., RASHEED, M. & GUPTA, A. (2022), Automated Recording of Human Movement Using an Artificial Intelligence Identification and Mapping System. Journal of Digital Landscape Architecture. https://doi.org/10.14627/537724007.

BRÜTTING, J. F. G. (2020), Optimum design of low environmental impact structures through component reuse [EPFL]. https://doi.org/10.5075/epfl-thesis-8448.

FERNBERG, P., STURLA, P. & CHAMBERLAIN, B. (2021), Pursuing an AI Ontology for Landscape Architecture. Journal of Digital Landscape Architecture, 6-2021, 452-460. https://doi.org/10.14627/537705040.

JOHNS, R. L., WERMELINGER, M., MASCARO, R., JUD, D., GRAMAZIO, F., KOHLER, M., CHLI, M. & HUTTER, M. (2020), Autonomous dry stone. Construction Robotics, 4 (3), 127-140. https://doi.org/10.1007/s41693-020-00037-6.

LABBAN, A. & BIZZI, L. (2022), Are social media good or bad for employees? It depends on when they use them. Behaviour & Information Technology, 41 (4), 678-693. https://doi.org/10.1080/0144929X.2020.1830174.

LIU, X. & TIAN, R. (2022), RiverGAN: Fluvial Landform Generation Based on Physical Simulations and Generative Adversarial Network. Journal of Digital Landscape Architecture, 7-2022. https://doi.org/10.14627/537724011.

RAMAN, T. A., KOLLAR, J. & PENMAN, S. (2022), Chapter 17 – SASAKI: Filling the design gap – Urban impressions with AI. In: AS, I. BASU, P. & TALWAR, P. (Eds.), Artificial Intelligence in Urban Planning and Design (pp. 339-362). Elsevier. https://doi.org/10.1016/B978-0-12-823941-4.00002-0.

ROBINSON, B. (2019, August 23), How Artificial Intelligence Is Preventing Cognitive Overload, Compassion Fatigue and Job Burnout. Forbes. https://www.forbes.com/sites/bryanrobinson/2019/08/23/how-artificial-intelligence-is-preventing-cognitive-overload-compassion-fatigue-and-job-burnout/.

SCOLERE, L. & HUMPHREYS, L. (2016), Pinning Design: The Curatorial Labor of Creative Professionals. Social Media + Society, 2 (1), 2056305116633481. https://doi.org/10.1177/2056305116633481.

SMITS, J. & BORGHUIS, T. (2022), Generative AI and Intellectual Property Rights. In: CUSTERS, B. & FOSCH-VILLARONGA, E. (Eds.), Law and Artificial Intelligence: Regulating AI and Applying AI in Legal Practice (pp. 323-344). T. M. C. Asser Press. https://doi.org/10.1007/978-94-6265-523-2_17.

ZEILINGER, M. (2021), Tactical Entanglements: AI Art, Creative Agency, and the Limits of Intellectual Property. meson press. https://doi.org/10.14619/1839.