Accuracy Evaluation of Tree Images Created Using Generative Artificial Intelligence

Jaeheon Kim[1], Yumi Lee[2]

[1]Seoul National University, Seoul/Republic of Korea
[2]Seoul National University, Seoul/Republic of Korea · yumil@snu.ac.kr

Abstract: This study evaluated the precision of artificial intelligence (AI) in generating tree images and aimed to integrate AI advancements with landscape design needs, fostering more innovative and efficient design approaches. The AI image-generation tool, Midjourney, was used to assess the accuracy of images for three tree species, to determine its applicability to landscape design. The results demonstrate that although the AI emphasized the color characteristics and seasonal changes of the trees with a certain level of accuracy, it exhibited limitations in accurately replicating the overall form and texture of the trees, particularly the complex patterns and textures of the bark. Consequently, AI image-generation technology still faces challenges in accurately reproducing the detailed characteristics of natural elements, which underscores the enhancements needed for practical applications in landscape design. These findings suggest AI's potential as a rapid visualization tool in the initial stages of landscape design while highlighting the need for more sophisticated technological development and improvements.

Keywords: Text-to-image AI, image generation, design tool, tree representation, image accuracy

1 Introduction

The importance of visualization in modern landscape design is being increasingly recognized, as it is crucial for conveying the aesthetic and functional characteristics of designed spaces (SINENKO et al. 2019). In this context, artificial intelligence- (AI) powered image generation offers an innovative approach to landscape design (RADHAKRISHNAN 2023). There is a growing trend toward AI-driven research and practice in landscape architecture, similar to allied disciplines (FERNBERG & CHAMBERLAIN 2023).

Recent studies, such as that by Ragab (2022), provide deep insights into the technical processes and outcomes of AI art, as well as its relationship with human creative responses. In landscape design, the role of visualization in the planning and design phases is crucial, and designers are required to predict and visualize how specific trees will appear in a given space in different seasons. The shift towards more modern, technology-based approaches is necessary for improved efficiency and accuracy, given that traditional visualization methods in landscape design are often time-consuming and labor-intensive (LIU & NIJHUIS 2020).

The advent of AI image-generation technology offers a solution to these challenges. However, systematic research on the effective use of AI image-generation technology in landscape design remains lacking. Research on the effective application of AI in various design processes is ongoing (THOLANDER & JONSSON 2023), but in landscape design – particularly involving real plants – the need is to verify AI's accuracy. If landscape designers can accurately generate images of specific trees, considering their shapes and seasonal changes, the design process can become significantly more efficient and precise. Therefore, this study explored the potential and limitations of the state-of-the-art AI technology in landscape design to propose methods for professionals in the field to use AI more effectively.

Journal of Digital Landscape Architecture, 9-2024, pp. 1029-1037. © Wichmann Verlag, VDE VERLAG GMBH · Berlin · Offenbach. ISBN 978-3-87907-752-6, ISSN 2367-4253, e-ISSN 2511-624X, doi:10.14627/537752098.

This study evaluated the accuracy of tree images generated using image creation AI and the images were compared with actual trees to assess how well the AI represents their characteristics in real environments. The ultimate goal of this research is to bridge the gap between the technological advancements in image-generation AI and the requirements of landscape design.

2 Methodology

2.1 AI Tool Selection

After evaluating several AI tools for image generation based on the quality, diversity, and suitability of the images produced for landscape design, we selected Midjourney. Midjourney AI both showcases its creative capabilities and reveals its limitations, particularly in generating diverse outputs that foster creativity in conceptual design (RADHAKRISHNAN 2023). For Midjourney to generate images involves inputting a prompt, which produces four different images simultaneously. The prompts can range from simple words to full sentences, thereby allowing users to input exact details of their choosing, including the option of uploading image files. In addition, Midjourney offers settings that users can use to adjust, among other features, the version and image aspect ratio. The capability of Midjourney to generate images in various styles and forms based on the user-input prompts makes it a well-suited choice for the objectives of this study.

2.2 Species and Prompt Selection

Native Korean trees, such as *Ginkgo biloba*, *Prunus serrata f. spontanea*, *Zelkova serrata, Pinus koraiensis, Acer palmatum, Juniperus chinensis*, and *Chionanthus retusus*, were considered as the research subjects. These species represent a range of characteristics, including coniferous, deciduous, evergreen, and broadleaf trees, and were selected to ensure diversity and comprehensiveness. The typical forms and textures of these species, as well as how they are spatially used, were considered to evaluate how AI reproduces the unique characteristics of each species. In selecting trees for this study, popular names with higher search data volumes were used instead of scientific names based on the results produced for both. Trees with colors or other specific nouns in their names were excluded because of the tendency for these names to lead to abnormal image distortions or images that focus on the specific noun rather than the tree itself. Notable examples included the red pine, red palmate maple, and cherry blossom trees. Three species (*Ginkgo biloba, Juniperus chinensis,* and *Chionanthus retusus*) were ultimately selected, and images were generated using their most common names and not their scientific names. In selecting these three species, consideration was given to trees that stand out with spring blossoms among evergreens, as well as those with colorful autumn foliage.

The prompt critically influences the output of image-generation AI tools, because the word choices determine the results (WANG & JIAQI et al. 2023). To assess the accuracy of the tree images produced by Midjourney, we used images corresponding to four different groups for each tree. These were: Overall Shape of the Tree Group (Tree Group), Leaf Group, Bark Group, and Implementation in Space Group (Space Group). When deciding on the prompts for each group, the images generated from concise prompts and those with additional adjectives were compared. The rate of inaccurate images was lower with concise prompts, which

led to the decision to use these for the final image generation. The prompts that were used to evaluate the final images for accuracy were:

1) Tree Group prompt: "specific (common name of tree)" tree --v 5.0 --s 1000.

2) Leaf Group prompt: "specific" tree's leaves --v 5.0 --s 1000.

3) Bark Group prompt: bark of "specific" tree –no leaves --v 5.1 --style raw --s 1000.

4) Space Group prompt: city park with "specific" tree --v 5.0 --s 1000.

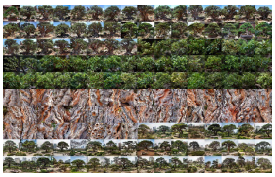
The prompts used in the study include using parameters to set the version of Midjourney, with "--v 5.0" or "--v 5.1" denoting the version settings. During the experiment, versions up to 5.2 were available and the optimal version for each group was selected based on the version that resulted in the fewest excluded images. The "--s" parameter is used to adjust the level of "artistic" styling and was included to obtain the most detailed and realistic results. Distinct parameters, namely "–no leaves --v 5.1 --style raw --s 1000," were used for the Bark Group.

This decision was made due to the high frequency of images that failed to depict accurately the texture and appearance of tree bark. There was a high failure tendency for the generation that included 'leaves' abnormally, to produce typical and accurate images of bark. To address this, the "–no leaves" parameter was added for the Bark Group. A city park, which represents a common landscape space in urban environments, was selected as the setting for the Space Group.

2.3 Image Generation and Image Selection

Images for various tree species were generated using Midjourney version 5.0 and 5.1. The prompts that were input into Midjourney were finalized after a pilot test based on related preliminary research. Overly detailed prompts were avoided and a conversational style was adopted. Images were created for the four groups using the finalized prompts. Forty images were generated for each group for three selected species, resulting in the creation of 480 images (Tab. 1). Midjourney produced four selectable images in each generation cycle. After 10 iterations, this process produced 40 generated images. Those deemed inappropriate based on a qualitative assessment were excluded.

Table 1: Complete image set for each tree species

Category	Ginkgo biloba	Juniperus chinensis	Chionanthus retusus
Image set			

The excluded images were considered as being difficult to assess accurately. The reasons for exclusion included black and white representations, with filter applied, bark and leaves combined in unrealistic and abnormal forms, and trees represented in atypically small sizes, e. g., bonsai (Tab. 2).

Table 2: Examples of Excluded Images

Category	Example 1	Example 2	Example 3	Example 4
Excluded images				
Reason	Bonsai	Unusual shape of bark	Black and white representations	Watercolor filters

These selections were performed internally. The percentage of images excluded per group is as follows. On average, approximately 4.38% of the images were excluded per group. Certain groups had no images excluded, while the maximum exclusion rate was 7.5%. Exclusions were not intended to lower or raise the accuracy measurements of the generated images artificially, but rather to measure the accuracy when using appropriate prompts that would probabilistically produce results above a certain quality threshold. Therefore, after excluding inaccurate images that were unnecessary for the accuracy measurement from the entire image set, five images per group were randomly selected from the complete image set (Tab. 3) for further analysis.

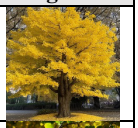
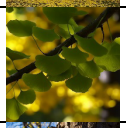
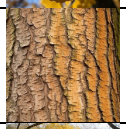
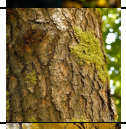
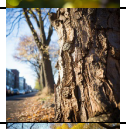
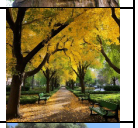
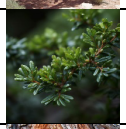
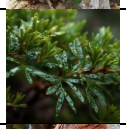
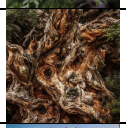
3 Image Evaluation

3.1 Method

The accuracies of the generated images were assessed using two methods. First, a survey using a five-point scale was conducted among landscape design experts to evaluate how the AI-generated images reflected the actual form and texture of the trees. The scale started from Perfectly Consistent (5 points), to Consistent, Partially Consistent, Mostly Inconsistent, and finally Completely Inconsistent (1 point). The images used in the survey can be found in Table 3.

Second, a plant recognition application was used to assess the accuracy of each image in representing the respective tree species. The selected application, PictureThis, was developed in China and contains a database of over 10,000 plant species that are predominantly Asian. PictureThis was selected from among the top six most downloaded plant identification applications because of its high accuracy (SCHMIDT et al. 2022). Although plant recognition applications cannot yet replace the expertise of skilled professionals, they offer high accuracy relative to time efficiency. This dual approach was intended to enhance the study's objectivity and reliability.

Table 3: Selected images

Tree	Group	Image a	Image b	Image c	Image d	Image e
Ginkgo biloba	Tree					
	Leaf					
	Bark					
	Space					
Juniperus chinensis	Tree					
	Leaf					
	Bark					
	Space					
Chionanthus retusus	Tree					
	Leaf					
	Bark					
	Space					

3.2 Participants and Measures

The evaluators were graduates with a bachelor's or higher degree in forestry or landscape architecture and 14 experts including botanical garden researchers and landscape design practitioners. The evaluation method involved a simple questionnaire for gathering accuracy ratings and comments. The questionnaire was designed to rate each item on a five-point scale. To enhance accuracy, in the sequence for evaluation alongside the AI-generated images, we randomly included two actual photographs, provided by the laboratory, of each of the three tree species. The evaluation scores from the actual tree photographs were used to exclude one evaluator's scores, which were deemed unreliable. Thus, 13 evaluations were analyzed.

3.3 Data Analysis and Results

The accuracy scores for the images evaluated in the survey are presented in Figure 1. The scores for each image rated on the five-point scale, were averaged. Notably, the Tree and Space Groups exhibited greater variance in accuracy ratings than the Leaf and Bark Groups.

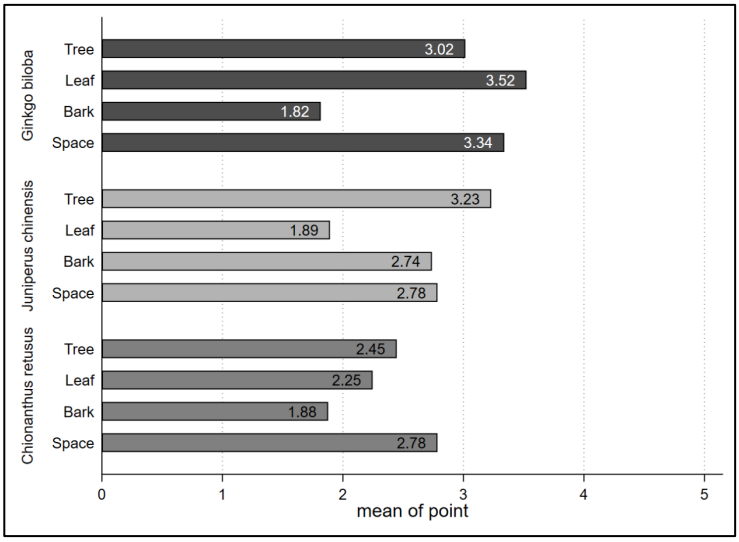


Fig. 1: Mean point of selected images

The accuracy evaluation of the tree-specific images revealed distinct patterns. Regarding the leaf representation, the AI replicated the shape and edges fairly accurately but struggled to capture the unique form, leaf arrangement, and intricate texture and patterns of the bark. This highlights the tendency towards a biased representation of certain features.

A noticeable decrease in accuracy was observed for the juniper tree images. The elaborate representation of the AI often failed to convey the typical features of the juniper tree and produced unnatural textures in the canopy and shape. Furthermore, the leaf images did not effectively capture the coniferous traits of the juniper tree, with issues such as an overly wide leaf span and disproportionately high representation of the needles.

The AI-generated images did not successfully reproduce the characteristic umbrella-shaped form and leaf structure of the fringe tree. The arrangement and texture of the leaves lacked clarity, leading to significant disparities with the actual leaves. Regarding the bark representation, although some accuracy was exhibited in depicting the bark of the fringe tree, the AI failed to capture adequately the overall characteristics of the tree as a streetscape feature.

Feedback from evaluators highlighted specific issues with AI-generated images of various tree species. *Ginkgo biloba's* form was criticized for its unnatural appearance, especially in the Space Group. While the leaf shape was somewhat accurate, the arrangement and bark depiction lacked clarity and specificity. The *Juniperus chinensis* images were misleading, with an overly wide leaf span and a pine-like appearance for both leaves and bark. *Chionanthus retusus,* however, showed better accuracy in flower representation, though it still struggled with form, leaves, and bark accuracy.

The results according to category are described as follows. Images that were generated through prompts for the overall form of the tree species tended to focus on the color characteristics of the tree types. Features such as the flowering season, foliage color, and general color traits of the tree species were prioritized. In depicting the form the accuracy was found to be relatively lower. For the Leaf Group, while the overall shape of the leaves was captured, the implementation failed to represent accurately the specific features unique to each species. In the Bark Group, despite adjusting the prompts to generate images of normal bark, the expert evaluations indicated a general decrease in accuracy across all tree types.

The accuracy of AI-generated images in the Space Group varied with the typical planting locations of the species. This underscores how AI's dataset influences image creation, emphasizing the need to match plants with their natural or common environments to improve accuracy. Notably, even without specific descriptors like colors or fruits in a plant's name, mentioning a region (e. g., “Chinese”) can lead to images that reflect the characteristics of that region more distinctly. Seasonal color changes and natural habitats play a significant role in this variability, highlighting the nuanced relationship between AI image accuracy and the specific traits of each plant species.

The results from the PictureThis application were as follows. The images of *Ginkgo biloba* exhibited a higher recognition rate than those of the other two trees, probably because of the distinct shape of *Ginkgo biloba* leaves and the color of their fall foliage. For *Juniperus chinensis*, the application tended to recognize images in the Tree group, similar to the other types of conifers. This result matches that of the evaluation by experts. The images of the leaves and bark of the *Chionanthus retusus* generally exhibited an inability to reproduce identifiable characteristics. However, the images for the *Chionanthus retusus* in the Tree Group and Space Group had relatively higher recognition rates, which was a contrasting result to that of the expert evaluation. The Tree Group and Space Group had higher recognition rates than those of the Leaf Group for deciduous trees. Given the nature of the Midjourney images, which often failed to create accurate representations of the bark alone, trusting the recognition rates for the Bark Group was difficult. Images with an unnatural mix of bark and leaves could have led PictureThis to recognize leaves rather than bark. A characteristic of the tree images of Midjourney is their tendency to represent highly seasonal features. This could be a reason for the difference in recognition rates for the Tree Group and Space Group compared with the expert evaluations.

4 Discussion

In this study, we conducted an in-depth analysis of the limitations of tree image generation using AI. In particular, images that were created using Midjourney exhibited a certain level of accuracy for various tree species such as *Ginkgo biloba, Juniperus chinensis*, and *Chionanthus retusus.* However, this accuracy was biased towards specific qualities. Although the tendency of AI to emphasize the color characteristics and seasonal properties of trees is noteworthy, it indicates a failure to capture critical features such as the overall form and complex textures of the bark. Thus, AI image-generation technology still faces challenges in accurately reproducing the detailed characteristics of natural elements, which suggests that improvement is needed for its practical application in landscape design.

These findings can guide the use of AI in landscape design. AI-generated images can be useful rapid visualization tools in the initial design phase. However, achieving precise design implementation requires further technological advancements and a more strategic approach to generating desired plant images. As the landscape of image-generating AI continues to evolve with new technologies and updates, developing adaptable workflows and understanding each AI's unique capabilities becomes more important than compiling a list of common prompts. This may involve refining prompt methodologies and diversifying training datasets. Additionally, the subjectivity of expert evaluations and data analysis limitations highlight the need for a more objective and comprehensive evaluation system. These findings lay the groundwork for future research into seamlessly integrating AI technology into landscape design, providing essential data for its improved application in the field.

5 Conclusions

The results of this study revealed both the potential and practical limitations of AI image-generation technology in landscape design. The AI applications demonstrated their ability to reproduce the color and *ambience* of trees effectively under certain conditions; however, limitations were encountered in accurately replicating the complex forms and textures of the selected trees. These findings suggest that AI technology could be utilized as an initial visualization tool in landscape design, but further technological advancements are required for precise design implementation and execution.

To overcome these limitations and improve accuracy, targeted research and development are essential. This could include more adept AI tools, capable of capturing the form, color, and texture of plants, being specifically constructed for landscape design. Moreover, expanding AI training datasets to encompass a broader range of plant species is crucial for enabling AI to learn and replicate diverse plant characteristics more precisely. Building a high-quality plant image database to ensure botanical accuracy is also vital. Also, this database should correlate scientific and common names of plant species and encompass data on the varied appearances of plants throughout the seasons.

This study provides essential foundational data for exploring the potential of AI technology in landscape design. The ongoing advancements in AI technology offer the potential for creative changes in the future of landscape design, necessitating in-depth research on techno-

logical progress and practical application strategies. Integrating AI with actual landscape design projects could prove highly beneficial, enabling an analysis of how AI tools can be seamlessly incorporated into the design workflow. This is pivotal for evaluating the real-world applicability and value of AI in landscape design, fostering the effective use of AI at various stages of the process, and encouraging professionals to explore more efficient ways to leverage this technology.

Acknowledgments

This work was supported by SNU Environmental Planning Institute.

References

FERNBERG, P. & CHAMBERLAIN, B. (2023), Artificial Intelligence in Landscape Architecture: A Literature Review. Landscape Journal, 42 (1), 13-35.

LIU, M. & NIJHUIS, S. (2020), Mapping landscape spaces: Methods for understanding spatial-visual characteristics in landscape design. Environmental Impact Assessment Review, 82, 106376.

MIDJOURNEY (2023), https://www.midjourney.com.

PICTURETHIS (3.68) [Mobile application software]. Hangzhou (Zhejiang, People's Republic of China): Glority Global Group Ltd. 20 September 2023. https://apps.apple.com/kr/app/picturethis/id1252497129.

RADHAKRISHNAN, A. M. (2023), Is Midjourney-AI the New Anti-Hero of Architectural Imagery & Creativity? GSJ, 11 (1).

RAGAB, H. (2022), A (Mid) Journey To The Virtual World Of Hassan Ragab. ParametricArchitecture.

SCHMIDT, R. J., CASARIO, B. M., ZIPSE, P.C. & GRABOSKY, J. C. (2022), An Analysis of the Accuracy of Photo-Based Plant Identification Applications on 55 Tree Species. Arboric. Urban For., 48 (1), 27-43.

SINENKO, S., POZNAKHIRKO, T. & OBODNIKOV, V. (2019) Automation of visualization process for organizational and technological design solutions. In MATEC Web of Conferences (Vol. 270, p. 05008). EDP Sciences.

THOLANDER, J. & JONSSON, M. (2023), Design Ideation with AI – Sketching, Thinking and Talking with Generative Machine Learning Models. In Proceedings of the 2023 ACM Designing Interactive Systems Conference (pp. 1930-1940).

WANG, J., LIU, Z., ZHAO, L., WU, Z., MA, C. & YU, S. et al. (2023), Review of Large Vision Models and Visual Prompt Engineering Meta-Radiology, 100047.