Neural Canvas: Supporting Scenic Design Prototyping by Integrating 3D Sketching and Generative AI

Yulin Shen
yshen654@connect.hkust-gz.edu.cn
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China

Yifei Shen
yifei.shen@aya.yale.edu
Yale University
New Haven, Connecticut, USA

Jiawen Cheng
chiaki_meow@sjtu.edu.cn
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China

Chutian Jiang
cjiang893@connect.hkust-gz.edu.cn
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China

Mingming Fan
mingmingfan@ust.hk
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
The Hong Kong University of Science and Technology
Hong Kong SAR, China

Zeyu Wang*
zeyuwang@ust.hk
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
The Hong Kong University of Science and Technology
Hong Kong SAR, China

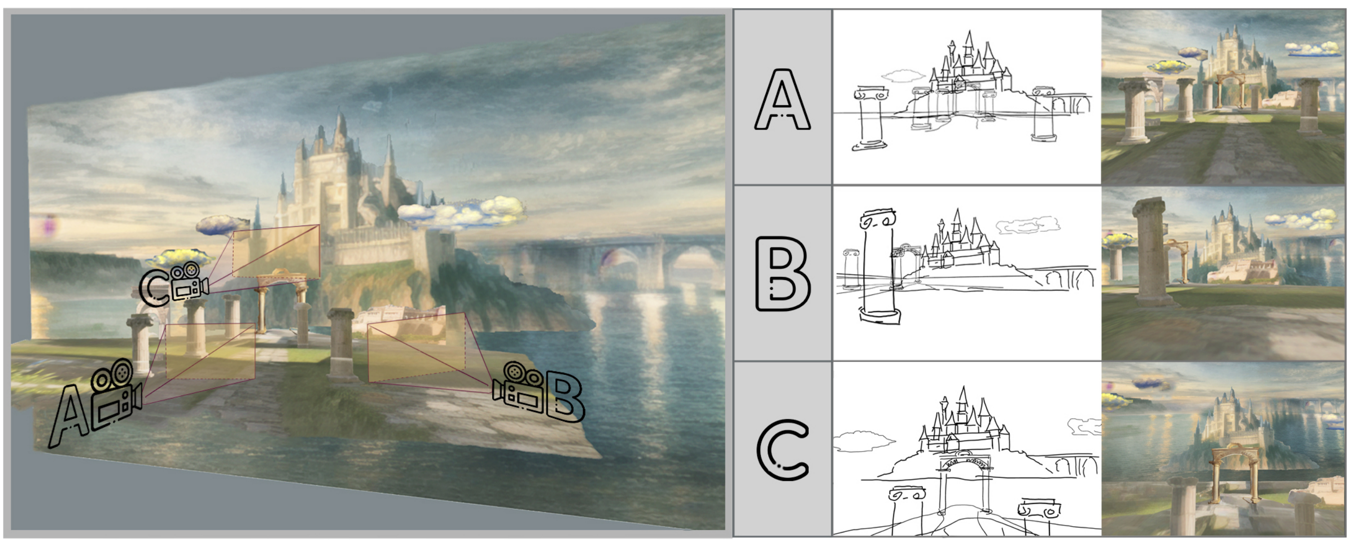

Figure 1: An example scenic design created using Neural Canvas with 3D sketching and generative AI models. Left: The 3D environment with three camera views. Right: 3D sketches and projected AI-generated appearance in the corresponding views.

ABSTRACT

We propose Neural Canvas, a lightweight 3D platform that integrates sketching and a collection of generative AI models to facilitate scenic design prototyping. Compared with traditional 3D tools, sketching in a 3D environment helps designers quickly express spatial ideas, but it does not facilitate the rapid prototyping of scene appearance or atmosphere. Neural Canvas integrates generative AI models into a 3D sketching interface and incorporates four types of projection operations to facilitate 2D-to-3D content creation. Our user study shows that Neural Canvas is an effective creativity support tool, enabling users to rapidly explore visual ideas and iterate 3D scenic designs. It also expedites the creative process for both novices and artists who wish to leverage generative AI technology, resulting in attractive and detailed 3D designs created more efficiently than using traditional modeling tools or individual generative AI platforms.

*Corresponding author.

CHI '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0330-0/24/05
https://doi.org/10.1145/3613904.3642096

CCS CONCEPTS

• **Computing methodologies** → **Graphics systems and interfaces**; • **Human-centered computing** → *Web-based interaction.*

KEYWORDS

generative AI, 3D sketching, scenic design, prototyping

ACM Reference Format:
Yulin Shen, Yifei Shen, Jiawen Cheng, Chutian Jiang, Mingming Fan, and Zeyu Wang. 2024. Neural Canvas: Supporting Scenic Design Prototyping by Integrating 3D Sketching and Generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3613904.3642096

1 INTRODUCTION

Scenic design refers to the creation of a 3D environment as theatrical scenery, which involves the design and configuration of sets, props, and other elements for dramas, games, films, and performance-based mediums. A good scenic design should offer the audience an immersive experience through aesthetic appeal and better storytelling, atmosphere, and spatial organization [4, 23]. Scenic design is challenging because of its conceptual and abstract nature. Designers often need to spend considerable time and budget, and must possess sufficient technical expertise in order to externalize their ideas. They also need to incorporate many divergent ideas into existing design drafts and then constantly modify and polish them.

Options for scenic design prototyping tools are currently rather limited, with sketching and 3D modeling being most commonly used in the design process. In a typical workflow, designers would collect many reference images to make a mood board, and then use traditional 3D modeling software to create polygon meshes, texture maps, and light sources so they could finally render a possible design. This workflow is time-consuming because tools such as 3D modeling software are better suited for final production rather than rapid prototyping. Designers often face a dilemma where they must spend considerable time finalizing designs before they can check whether their initial design ideas even work.

Sketching is widely regarded as an agile prototyping tool [19, 24, 89] and has often been used in scenic design, as it allows designers to quickly explore visual ideas and iterate design alternatives. However, it is not easy to use traditional sketching to present spatial and appearance information in 3D scenic design, in particular the relative depth relationships between objects and their material properties. Some recent 3D sketching interfaces [5, 44, 61] have proved their effectiveness in expressing spatial design ideas, but it is still difficult to create material appearances with these tools.

Recent advances in generative AI have made appearance creation much easier and offered new alternatives to sketching and traditional 3D modeling. In particular, recent diffusion models can produce photorealistic 2D images from text prompts and sketches [60, 68, 90, 95]. We conducted a formative study with professional scenic designers or related artists to understand how generative AI can facilitate scenic design prototyping. Our key finding is that an ideal scenic design prototyping tool needs to support 1) efficient creation of appearance, 2) intuitive externalization of spatial concepts, and 3) rapid idea exploration and design iteration. As shown in Table 1, existing tools cannot satisfy these requirements simultaneously, while the benefits of integrating 3D sketching and generative AI make this possible.

We propose Neural Canvas, a lightweight web-based 3D platform that integrates 3D sketching and generative AI to meet the above three requirements of a desired scenic design prototyping tool. With Neural Canvas, users begin by drawing sketches on canvases placed in a 3D environment and use projection functions to create an initial 3D design. Neural Canvas also integrates a collection of generative AI models, which enables users to easily convert 3D sketches into photorealistic assets with detailed appearance. Users can then edit them to iterate their designs until they achieve the desired visual effect.

We conducted a user study to evaluate the effectiveness of our system as a scenic design prototyping tool, and to explore new use cases enabled by integrating 3D sketching and generative AI. The study shows that Neural Canvas is an effective prototyping tool and supports the creation of diverse scenic designs. It also reveals new user behaviors enabled by Neural Canvas such as using 3D sketching to control the viewpoint and improve input quality for generative AI.

Our main contributions are as follows:

- We propose Neural Canvas, a web-based platform that supports scenic design prototyping. Users can prototype 3D shapes and their spatial relationships using operations like sketching, canvas manipulation, and projection.
- We integrate various generative AI models supporting different modalities into our platform, which allows users to avoid switching back and forth between individual AI applications, enabling them to discover new ideas with the convenient interaction between various AI models.
- We design two new projection methods working closely with generative AI functions, which easily project objects in a generated image back onto their corresponding sketch or desired planes. We also provide users with three ways to use 3D sketches as input to generative AI models.
- Our user study shows that our system can facilitate scenic design prototyping and help users explore new ideas that they would not have conceived otherwise. Neural Canvas makes it easier to express and prototype scenic design ideas resulting in more detailed and attractive 3D content.

2 RELATED WORK

2.1 Image-Based Modeling and Rendering

Designers often resort to geometric and appearance modeling to render a photorealistic scene. However, this approach requires a lot of expertise and can be time-consuming. How this process can be streamlined and expedited has, therefore, become the main focus of previous research. One approach focuses on using multiple 2D images to construct 3D scenes directly. For example, structure from motion (SfM) is an effective way of 3D reconstruction based on feature matching and camera pose estimation. Using photogrammetry, Snavely et al. [78] estimated camera pose information from crowdsourced images and reconstructed tourist attractions, enabling Photo Tourism [77], a 3D virtual tour of the reconstructed

Table 1: A summary of scenic design prototyping tools and whether they satisfy the three requirements. Note that sketching is limited in externalizing spatial concepts because 2D sketches may not represent full 3D information.

Tools	Efficient creation of appearance	Intuitive externalization of spatial concepts	Rapid idea exploration and design iteration
Mood Board	✓	✗	✗
Sketching	✗	limited	✓
3D Software	✓	✓	✗
Scale Model	✗	✓	✗
Neural Canvas (ours)	✓	✓	✓

architectures. Other studies use a single image to generate 3D models for viewing. Oh et al. [58] combined interactive tools with a decoupling filter for large- and small-scale textures to model a scene, and Hoiem et al. [29] introduced a model to automatically label different regions and "cut and fold" the image into a pop-up model for a simple scene. 360proto [57] can capture paper mockups of all components simply by taking a photo with a smartphone. It also allows organizing and editing captures, layering the captures, and making them interactive in a 360-degree AR/VR environment. Reframe [65] analyzes the security and privacy concerns of an AR environment by building scenarios with some preset image assets and interaction components. While these methods attempt to create scenes with preexisting images, we incorporate generative AI to produce images and textures on demand.

Another approach uses depth information embedded in planes to accomplish scene rendering. Layered Depth Images [71] extracts depth data from images and segments the corresponding components into multiple different planes to embed spatial positional relationships, and many studies have proven its effectiveness [39, 75]. Based on this method, more studies [21, 91] attempt to use deep learning for novel view synthesis. Zhou et al. [99] implemented a new method called multiplane images (MPI) for view synthesis. With depth information and transparency information learned from videos, it represents a scene with multiple layers of images. Based on the MPI method, AdaMPI [25] introduced two modules to adjust plane depth and color prediction and achieve better results in view synthesis. Neural Canvas is inspired by the idea of embedding depth information into multi-layered images, facilitating design with editable images on conveniently adjustable planes placed in a 3D environment.

Additionally, Neural Canvas decouples the process of generating multiplane images, namely depth estimation, object detection, image inpainting, etc. Users possess full control of the process, with the freedom to choose between manual design and generative AI at each step.

2.2 3D Sketching

Sketching is an effective tool for design prototyping. However, 2D sketching cannot represent depth information or the 3D spatial relationship between objects, which are essential in 3D scenic design. Therefore, many studies have explored 3D sketching in various design contexts.

One approach is to generate accurate 3D models by analyzing the structural information in 2D line drawings, known as sketch-based modeling [59]. Due to the lack of depth information and positional relationship of strokes, sketch-based modeling is an ill-posed problem. Researchers usually set some spatial constraints so that the 3D reconstruction problem can be solved procedurally. SmartCanvas [97] uses a background picture to build the context of 3D sketching and optimize results during the process of sketching. SweepCanvas [49] uses a pair of strokes to generate 3D surfaces and extracts spatial information from RGB-D images. DreamSketch [42] uses coarsely defined 3D sketches as the design context and uses generative algorithms to build a model with new strokes and background.

Another approach introduces planar canvases into 3D sketching for designing, using strokes to describe objects. This approach preserves descriptive details in 2D sketching while storing 3D spatial information in canvases. Mental Canvas [17] uses planes with adjustable positions to depict and generate corresponding scenes, and the scene of depth is achieved through the relationship between planes. Insitu [61] builds simple scenes with geometry positions and plane-based images. Also, some studies implement planar canvases in their systems to reconstruct cultural heritage sites. CHER-ish [70] introduces heritage images and corresponding camera positions for creators to create scenes with strokes, and Shen et al. [72] extract contours from photographs to facilitate reconstruction based on former work.

In addition, some studies focus on sketching directly in a 3D environment. With new hardware and systems, these methods map positions in the real world into a sketching coordinate space. Mobi3DSketch [46] uses mobile devices to set up planar or surface proxies in the real world and draw strokes onto them. Many other studies [5, 6, 43, 82] implement 3D sketching in VR and AR.

Neural Canvas implements a plane-based 3D sketching system for designers to create scene prototypes. Based on planar canvases, Neural Canvas combines sketch planes with image-based planes, encouraging users to make scene design prototypes with sketching and optimize the process with images. Furthermore, designers can iterate their designs with new ideas inspired by generated images.

2.3 Generative AI

Recently, generative AI has gained popularity thanks to efficient deep learning models and large-scale datasets. Several works have exhibited extraordinary performance in image synthesis.

Earlier research on AI generate models is mainly based on Generative Adversarial Networks (GANs) [96]. Isola et al. [37] introduced conditional GANs to achieve image-to-image translation tasks, such as synthesizing images from labels and colorizing images. StyleGAN [40] introduced a new data set and used style vectors to control high-level attributes in images and generate realistic pictures. GauGAN [62] can generate realistic pictures by filling the corresponding textures of users' sketched color blocks.

Recent image generation models use variants of the diffusion model [79], and they can produce more realistic pictures than previous methods [16, 28]. Stable Diffusion [68] is currently one of the most popular variants and is capable of generating realistic images given any text input. ControlNet [95] adds handles such as sketches, human body poses, and depth to prime image generation with Stable Diffusion.

Like generative AI, segmentation models also benefit from large-scale datasets. Previous works in segmentation mainly focus on tasks separately, such as interactive segmentation [11, 41], semantic segmentation [74, 80] and edge detection [3]. Segment Anything [45] built the largest segmentation dataset to date (by far) and proposed a new segmentation task model to accomplish challenging assignments. It provides users with interactive control and can solve a range of downstream segmentation problems effectively.

Many researchers release their pre-trained models to the public. To help users obtain desired models, Modelverse [50] provides a model search platform to retrieve a model or pre-trained network weights using an image, a sketch, or a text description. However, because different kinds of AI models were not well integrated, designers still had to switch between different websites and upload and download files multiple times to generate images. Neural Canvas solves this problem by integrating generative AI models into one single platform.

3 FORMATIVE STUDY

We conducted a formative study to gain insight into the current practice of scenic designers and to understand their desired features in a prototyping tool. We invited five professionals to participate in a one-hour semi-structured interview. Our participants include an off-Broadway stage designer (FP1), a TV show stage designer (FP2), a 3D content creator (FP3), and two new-media artists (FP4, FP5), all of whom have extensive experience and professional expertise in scenic design. We covered the following topics in this interview: 1) the typical practices and challenges in a scenic design workflow, and 2) the benefits and shortcomings of existing design tools.

3.1 Scenic Design Pipeline

From the responses of our interviewees, we were able to summarize a typical scenic design workflow. Fig. 2 shows this workflow. It consists of three phases: preparation phase, prototyping phase, and implementation phase. The preparation phase often involves the entire larger project in which the scene design is included. The preparation phase includes screenplay selection, story writing, team-up, etc. The prototyping phase consists of multiple rounds of communication, research, idea exploration, and visual representation making. The final plan will be delivered to the implementation phase to build a stage or build a virtual world according to the needs of the project.

During the prototyping phase, designers need to first communicate with other colleagues, such as directors or editors, on the theme, style, and atmosphere needed for the scene. Then, designers conduct research on the designated topic and explore initial ideas. They use visual representations like paintings or 3D software renderings to express their design ideas and then present them to others for further discussion. This process may be repeated many times until the final plan is determined so they can enter the next implementation stage.

3.2 Existing Creativity Support Tools

From the interviews, we gathered participants' feedback on existing design prototyping tools and identified issues in these tools.

Mood Board. FP1 mentioned that the mood board can serve as a representation of research results. She pastes the materials collected during her research on a whiteboard and wants to look for more inspiration from the materials she has already gathered. However, mood boards can only display past works and materials, lacking the ability to edit and present design ideas. Therefore, a mood board can only participate in the stages of research, idea exploration, and communication.

Sketching. All participants regarded traditional sketching as a very efficient prototyping tool before the emergence of digital tools, but they also agreed on some of its major drawbacks. All participants mentioned that sketches cannot easily present material appearance, which is vital for the theme and atmosphere of the scene. FP1 and FP2 also said that even sketches with good perspective relationships cannot accurately reflect spatial relationships. FP1 said *"I have to observe the relationship between the motion lines in space and the position of objects."* Because traditional sketching is unable to represent spatial and appearance information, its usefulness in the prototyping phase is limited. As more digital tools emerged, traditional sketching has been increasingly replaced by other tools.

3D Software. FP2 and FP3 used 3D software, such as Blender [22], C4D [51], SketchUP [36] extensively for their design. 3D software enables users to place and observe objects in a three-dimensional environment in a WYSIWYG manner. However, to achieve the expected appearance, users are required to do complex modeling operations and search for suitable materials and textures. Lighting is also another challenge. All the participants claimed that using existing 3D software costs a lot of time and it is difficult to find inspiration and make large-scale changes.

Scale Model. A scale model is a physical model that is geometrically similar to the prototype. Designers use real materials to build scale models, such as wood, metal, and fabric. FP1 referred to scale models as the final level of the prototyping phase. Scale models can provide designers with the most realistic material and lighting effects, but they are extremely inconvenient to produce and difficult to edit. Therefore, scale models are only used as a final check and a way to share the design with others.

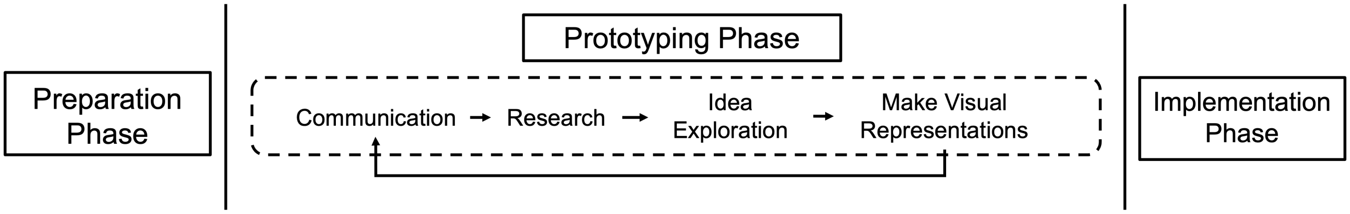

Figure 2: A typical workflow for scenic design. The prototyping phase usually requires multiple rounds of iteration, and each round may require communication with peers, where attractive and understandable visual expression is essential.

3.3 Design Considerations

Based on the interviews in our formative study, we summarized three key requirements that an ideal scenic design prototyping tool should meet simultaneously.

- **It should support efficient creation of appearance.** Appearance is very important to express the atmosphere and mood of a scenic design. Traditional sketching requires a lot of time to fill in color while existing 3D software requires skills to search and adjust material, texture, and lighting to create desired appearance.
- **It should support intuitive externalization of spatial concepts.** In scenic design, users need to consider the spatial relationships for object placement to provide the audience with a plausible and appealing environment. Traditional sketching fails to present 3D spatial information even with accurate perspectives.
- **It should support rapid idea exploration and design iteration.** As a prototyping tool, it should help users find inspirations and quickly modify and change the content whenever they would like to iterate their design ideas.

4 NEURAL CANVAS

Following the three principles we identified in the formative study, we designed Neural Canvas to leverage both 3D sketching and generative AI to support scenic design prototyping. 3D sketching allows users to embed spatial information in their strokes, whereas generative AI models can quickly generate detailed appearance. Both 3D sketching and AI content generation are lightweight operations from a user's perspective, making it easy to explore a variety of ideas and iteratively improve design.

Referring to the Cognitive Load Theory [83] and Hick–Hyman Law [27, 32], we think that compared with using each AI model on a separate platform, integrating these features into one platform can reduce the complexity and cognitive burden of user choices, improve user experience, and help users iterate design faster without switching between different platforms and applications. Fig. 3 depicts the overall system architecture of our platform, Fig. 4 showcases the user interface design, and Fig. 5 demonstrates the use-case scenarios of creators employing Neural Canvas.

4.1 System Architecture

Our platform is a web-based application with a client-server structure for easy access without requiring users to install or deploy the system. The front-end client includes an editor module, an AI module, a 3D renderer, a camera module, and basic image editing operations. The back-end services are divided into three parts: a web server, self-hosted services, and services provided by cloud or SaaS companies. We deploy some complex image editing algorithms and multiple open-source AI models on our server. We can also integrate services provided by open-source communities like models hosted on Huggingface, and companies like OpenAI.

Our front-end interface is implemented mainly using JavaScript, Vue.js [94], and Three.js [10]. The front end is responsible for visualization and basic image processing. AI functions and complex computations operate on a high-performance server designed to accommodate the huge computing power demands of deep learning models and equipped with a CPU of 64-core Intel Xeon Platinum 8370C 3.50GHz and 4 NVIDIA RTX 4090 GPUs. The back-end server is implemented with Nginx [67], Python, Flask [69], PyTorch [63], Gradio [1] and other scientific computing packages [9, 26]. Such architecture gives us the flexibility to plug in cutting-edge AI models and deploy them on different platforms including the cloud.

Neural Canvas integrates 3D sketching and more than ten generative AI models to support scenic design prototyping. To make the integration tighter and provide a smoother user experience, we propose three input methods for generative AI in Sec. 4.3.2 and two projection methods in Secs. 4.5.3 and 4.5.4.

4.2 Editor Module

The editor module implements an array of basic operations essential for 3D content creation. The Stroke Editor and Image Editor provide specialized tools for creating and modifying visual assets in our platform. The State Manager oversees the application's current state and provides undo/redo capabilities. The View Controller is used to manipulate the virtual camera. The Canvas Controller allows users to place and manipulate any geometric primitives in 3D that they can later draw on top of. The Layer Manager provides an interface for organizing all objects currently present in the scene. Finally, the user can use the file I/O capabilities of the File Manager to save/load their work and share their work with others. More implementation details for the Editor Module are provided in Appendix A.

4.3 AI Module

AI-generated content provides a way to enrich 3D sketching with appearance information and the ability to iterate quickly.

4.3.1 Integrated AI models. We integrated 13 AI functions to give users more choices to enhance the creative experience.

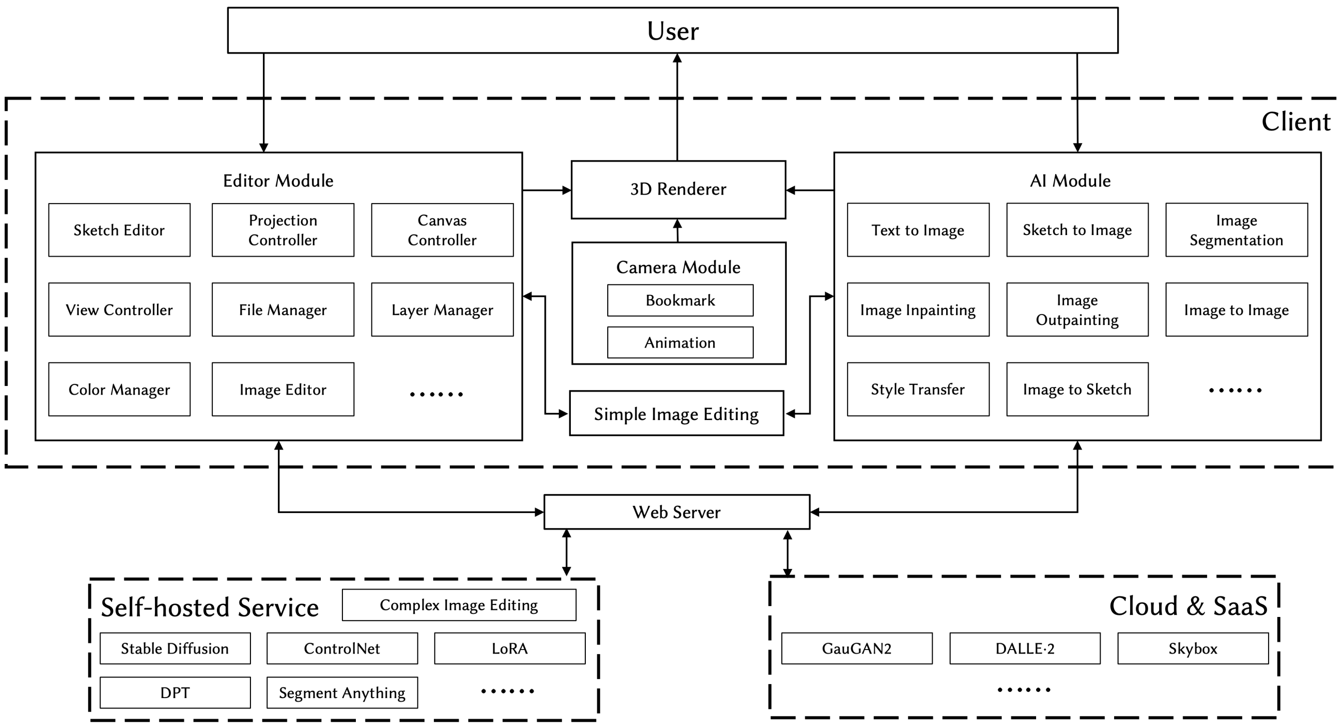

Figure 3: System architecture overview. The front-end and back-end separation design allows Neural Canvas to be used on different types of devices, and deploying AI functions on remote servers or clouds ensures the scalability of the AI Module to integrate cutting-edge generative AI models.

Text to Image. Text-to-image synthesis refers to the process of generating a visual representation, typically in the form of a photorealistic image, based on a textual description. Designers can use this function to generate desired art assets without externalizing visual ideas. In Neural Canvas, we adopt Stable Diffusion [68] for this purpose.

Sketch to Image. Converting sketches to images can greatly increase information capacity and is a very important functionality supported in our system. In order to make sketch-based synthesized images better reflect the user's creative intention, we need controllable image generation from sketches. To that end, we decided to deploy ControlNet [95] with Stable Diffusion because users can have more control over AI-generated results using this architecture.

Image Inpainting. Image inpainting refers to the process of filling in missing or corrupted parts of an image in a visually coherent and plausible manner. The goal is to restore the image such that the filled-in regions are indistinguishable from the original, uncorrupted parts. When we segment objects out from the image and project them elsewhere, they leave large blank holes behind in the original images. We use image inpainting algorithms like LaMa [81] and Stable Diffusion [68] to fill these holes. Neural Canvas sends two images to the backend server as inpainting inputs. The first image is the original image without any holes, and the second one is a binary image representing which areas are missing.

Image Outpainting. Image outpainting, or image extrapolation, is the process of extending the content of an image beyond its original boundaries while maintaining visual coherence. Outpainting techniques analyze the content of the input image, taking into account the patterns, textures, and structures, to extrapolate it to a larger image that appears as a natural extension of the original scene. This is particularly helpful when creators want to place their scenes within a larger background or panorama. Our system uses DALL·E 2 [60], a proprietary image synthesis service.

Image Segmentation. AI-generated content adds a lot of details, and sometimes designers only want to select a certain part of it to add to their work. In order to better help users quickly crop regions of interest, we have introduced an image segmentation algorithm using a deep learning model. Neural Canvas also provides traditional lasso and magnetic lasso tools to provide users with more flexible options.

Other AI Functions. Besides the above five frequently-used AI functions, eight additional AI functions are available to users. Their details can be found in Appendix B.

4.3.2 Input Methods. We provided users with three input methods to control the content generated by generative AI models.

Current Canvas. The user can input the drawn content on the currently selected canvas into an AI model. The AI model can generate an image based on content on the current canvas and a text prompt.

Full-Screen Content. In our system, users can create 3D sketches by drawing on multiple planar canvases. Generating images on one

Figure 4: The user interface of Neural Canvas. a1) Mode selection buttons. a2) Gerneral operation buttons. a3) Canvas operation buttons. a4) Brush operation buttons. a5) Image operation buttons. a6) Projection operation buttons. b1) The brush parameter panel. b2) The layer manager panel. b3) The camera manager panel. b4) The 3D primitive manager panel. b5) The environment control panel. b6) The text box panel. c) AI function panel: this drop-down panel consists of multiple tabs of different AI functions.

canvas at a time is not only cumbersome but also adverse to controlling the style of the overall generated content. We enabled providing what the user sees on the full screen as input to a generative model. Specifically, it uses the camera from the user's current perspective to render the entire scene and obtain a 2D picture as input to the generative model.

Current Canvas with Background Reference. Sometimes, users want to modify an existing 3D scene by drawing sketches on a new canvas. We offer a method for users to generate images on the new canvas and use the existing background as a reference. This input method is particularly useful for AI models that take multiple inputs like ControlNet [95]. This allows image generation with a similar style or color tone to the reference image and generated content is conditioned on user-drawn sketches.

4.4 Camera Module

The camera module records the parameters of bookmarked cameras. Inspired by previous work [77, 92], our camera module can generate virtual tours by interpolating parameters between different cameras. A simple animation system allows the camera view to navigate the scene according to the bookmarked cameras.

Using the camera module, users can quickly switch camera views to observe their design from bookmarked perspectives and then use AI functions to create from different perspectives. Collecting images from multiple points of view enables users to achieve an effect similar to modeling scenes from images [78].

4.5 Projection Operations

Projection maps any asset on a canvas to another canvas while keeping the scene from the current perspective unchanged, as shown in Figure 6. This process assigns spatial information to strokes and images through user-specified receiver canvases. Projections, along with many other operations, allow users to select objects from a 2D plane and place them in 3D space, offering more perspectives to observe their designs and build immersive environments.

To tightly integrate the 3D environment and AI-generated 2D images, we designed four projection operations to place images in 3D space, which are demonstrated in Fig. 6.

4.5.1 Stroke Projection. We designed the stroke projection function based on the plane-based 3D sketching methods [17, 87, 88]. Users can first sketch on a 2D plane and then project any strokes onto a specified plane. During this process, the strokes appear apparently unchanged from the user's point of view, but the 3D position of the strokes shifts from one plane to another.

We define a sketch as a list of 3D strokes, $S = \{s_1, s_2, \ldots, s_n\}$, where each stroke s_i consists of a list of vertices $V_i = \{v_1^i, v_2^i, \ldots, v_N^i\}$. We adopted a planar-based 3D sketch representation method. The user first selects the target plane P_t and the plane P_o where the current stroke is located, and then clicks on the stroke s_p to project. We project all vertices $v_1^p, v_2^p, \ldots, v_n^p$ of s_p onto the target plane P_t, and connect them to form a new 3D stroke.

The projection process is described as follows. Denote $\boldsymbol{x}_v$ as the 3D coordinates of a vertex, and $\boldsymbol{x}_c$ as the current camera position.

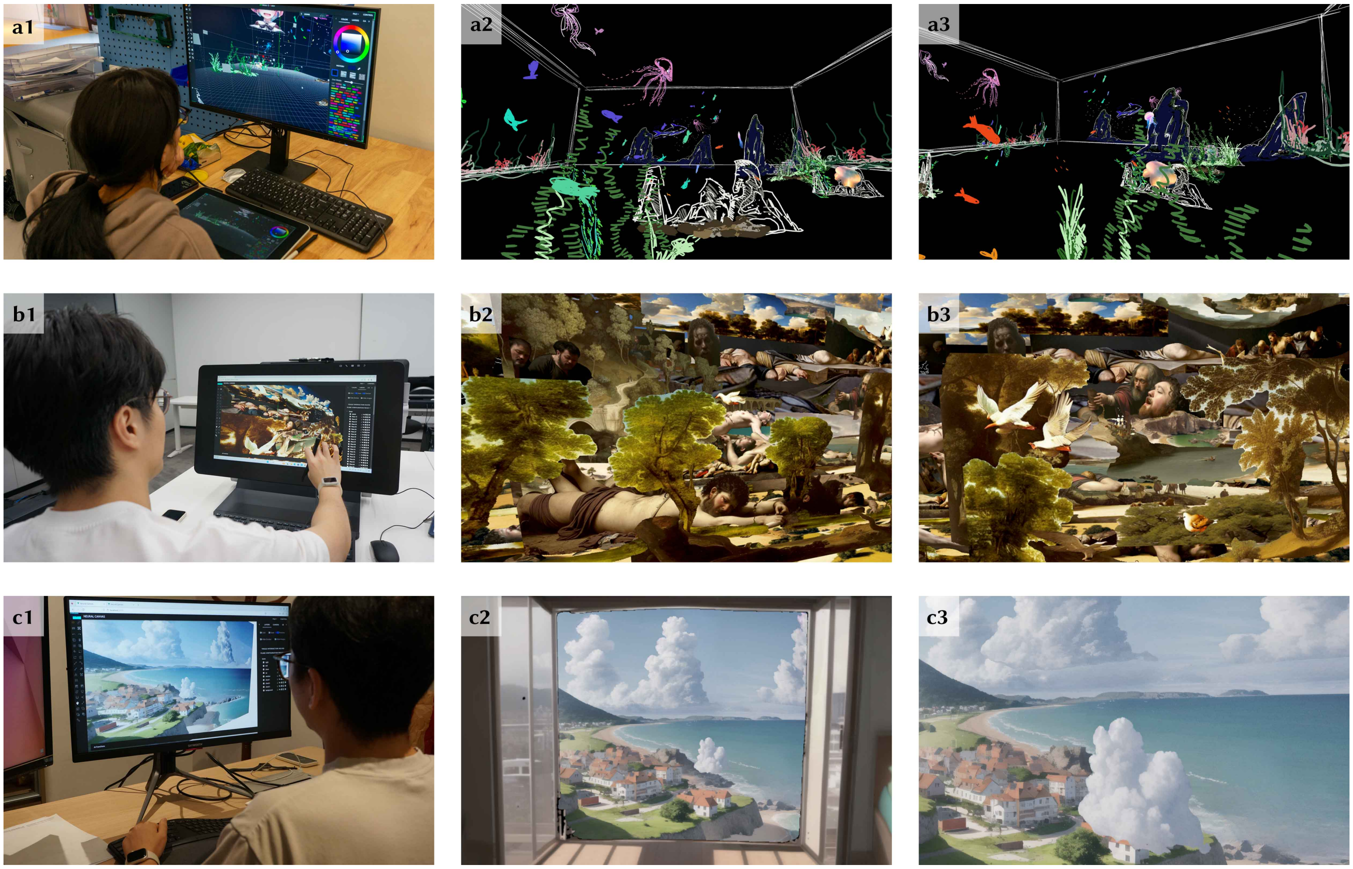

Figure 5: Neural Canvas in use. Left: Our system design supports various types of digital tablets and mouse input. Middle and Right: Two example views of their scenic design.

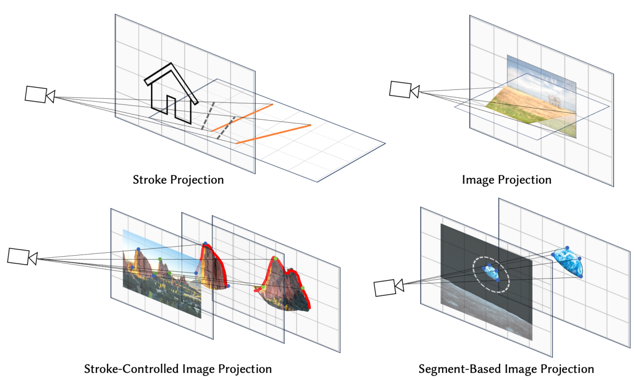

Figure 6: Illustration of four projection operations in Neural Canvas. These projection operations provide an intuitive way for users to create 3D content.

A straight ray can be obtained from these two points, on which each point can be expressed as $\boldsymbol{x}_c + \lambda\,(\boldsymbol{x}_v - \boldsymbol{x}_c)$. Given the plane P_t to project on, the position of the point after projection $\boldsymbol{x}'_v$ can be obtained by

$$\boldsymbol{x}'_v = \boldsymbol{x}_c - \frac{n \cdot \boldsymbol{x}_c + d}{n \cdot (\boldsymbol{x}_v - \boldsymbol{x}_c)} \cdot (\boldsymbol{x}_v - \boldsymbol{x}_c)\,,$$

where the equation of P_j is $\boldsymbol{n} \cdot \boldsymbol{x} + d = 0$, $\boldsymbol{n}$ is the normal vector of the plane, $\boldsymbol{x}$ is any point on the plane, and d is a scalar.

4.5.2 Image Projection. Image projection allows the image to be projected onto a plane as if by a slide projector. In this way, users can map an image to a specified plane in 3D space to construct a 3D scene. We first project the corners of the texture image to the target plane and then warp the image to fit the projected shape.

4.5.3 Stroke-Controlled Image Projection. When users create a scene via 3D sketching, they are likely to draw strokes on multiple planes. If the user subsequently chooses to generate an image from this 3D sketch using AI, it is not immediately obvious which parts of the image belong to which planes. It is tedious for users to segment objects of interest manually from the original image and then project them onto target planes.

We propose a projection method for AI-generated image parts controlled by their corresponding input strokes on multiple canvases, as shown in Fig. 7. Our method enables intuitive projection of an AI-generated 2D image to the 3D environment given a 3D sketch.

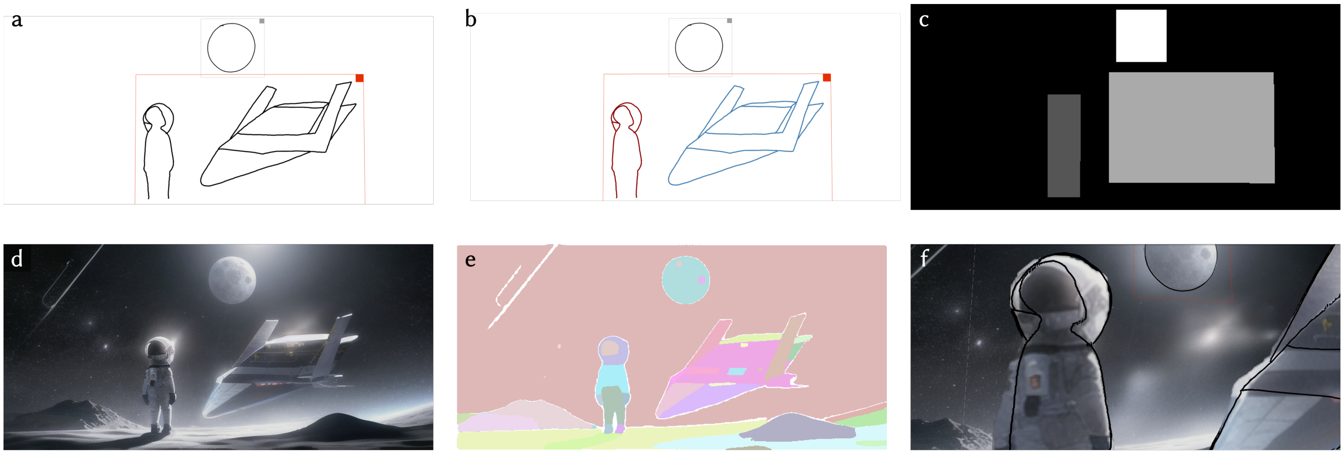

Figure 7: Illustration of stroke-controlled image projection. a) The original sketch. b) The connected strokes are divided into groups using the disjoint-set data structure [84]. c) Masks generated from the stroke groups. d) Generated image from the sketch. e) The segmentation map generated by SAM. f) Segmented objects are projected onto their corresponding sketch canvases, creating a parallax effect.

Technically speaking, we can denote a partial sketch drawn on a canvas as $S = \{s_1, s_2, ..., s_n\}$, where s_i represents the i^{th} stroke. We first find the bounding box B_i for each stroke s_i. Then we use the disjoint-set data structure to group the connected bounding boxes as $G = \{g_1, g_2, ..., g_m\}$. Group g_j represents a cluster of strokes that the user draws for object j. Next, we find the bounding box of group g_j on this canvas and project it back onto the AI-generated image plane to create a mask M. To better project pixels controlled by these strokes, we also applied the Segment Anything Model (SAM) [45] on this image to get its segments and project the segments onto the canvas if over half of the pixels of a segment fall in mask M.

4.5.4 Segment-Based Image Projection. AI-generated images often contain content beyond the initial input sketch. Such content will not be projected through stroke-controlled image projection. Neural Canvas also supports segment-based image projection, so users can create image segments using SAM [45] and project them onto any specified plane placed in the 3D environment. Segment-based image projection first maps the user-specified area from screen space to image space. Next, it uses SAM to segment the image. When the area covered by the designated area of the segmentation block reaches a certain threshold, we consider it to belong to the object that the user wants. Finally, we will merge the segmentation blocks that are covered by masks above a certain threshold and map them onto the projection plane.

5 USER STUDY

We conducted a user study to evaluate Neural Canvas's ability to support scenic design prototyping for both professionals and novices. The goal of the study is to understand how well Neural Canvas supported participants with their scenic design prototyping tasks. We recorded the screen and logged all user operations for further analysis.

5.1 Participants

We recruited 12 users to participate in our study. Eight were male, three were female, and one was non-binary. Their average age was 26.25. Six of them had experience with 3D software and a relevant design background, while the remaining participants with only 3D software experience were regarded as novice users. On a scale of 1–7, the average self-reported familiarity of the expert group with scenic design was 5.67, while that of the novice group was 3.33. Overall, users have used twelve kinds of 3D software (Blender [22], Solidworks [13], Inventor [34], Catia [85], Cinema4D [51], Unity [86], Unreal [20], Maya [35], SketchUp [36], 3DS MAX [33], ZBrush [52], Houdini [76]) and four kinds of AI tools (Midjourney [55], Stable Diffusion [68], DALL·E 2 [60], Meshy [54]) .

Table 2: Participants' groupings, assigned task 2 themes, and their occupations. Theme 1 is ' "aquarium," theme 2 is "the world outside the window," and theme 3 is "disaster."

#	Group	Theme	Occupation
P1	Novice	2	Researcher
P2	Novice	3	Researcher
P3	Novice	2	Researcher
P4	Expert	1	Designer
P5	Expert	3	Artist
P6	Novice	1	Researcher
P7	Novice	3	Researcher
P8	Expert	1	Artist
P9	Novice	1	Researcher
P10	Expert	2	Artist
P11	Expert	2	Artist
P12	Expert	3	Artist

5.2 Design Tasks

We designed two tasks for the participants. We asked them to complete design task 1 in 15 minutes and design task 2 in one hour.

Task 1 is designed as a trial task. Since our system is relatively large and usage is different from that of traditional 3D software, it takes some time for users to become familiar with the user interface and the projection functionality. We designed Task 1 to familiarize participants with one of our regular workflows. Participants are guided to author a simple still-painting scene. They are required to draw 2D sketches on a plane. Then, strokes are projected onto different 3D planes to construct a 3D scene. AI models generate content with detailed appearance information according to the 3D sketch. Finally, users project generated content to corresponding planes to complete the scene.

Task 2 is an open-ended design task to understand how participants would author a 3D scene using Neural Canvas. We designed three broad design themes: "aquarium," "the world outside the window," and "disaster." The aquarium theme focuses on appearance information. The theme of "the world outside the window" focuses on displaying the spatial information of the scene. The disaster theme encourages designers to use appearance information and spatial information to express the mood and atmosphere of the scene. Each theme has 4 participants involved in the design, including two experts and two novice users. All users randomly select a theme on a first-come, first-served basis; if the number of participants in a theme reaches the limit, the theme is then removed from the poll.

5.3 Procedure

The whole user study consists of four parts. First, we introduced the basic operations in our system. Then, participants were asked to complete Task 1 to gain familiarity with the system. After users were comfortable using the tool, they were introduced to Task 2 and the final deliverables. After task completion, the authoring experience of each participant was collected with a questionnaire, and a short semi-structured interview was conducted to collect users' opinions and feedback. Questions asked in the semi-structured interview can be found in the Appendix C. The study lasted about 120 minutes for each participant, and each participant was reimbursed $15.

5.4 Measurement

We recorded the screen and audio during the study for further analysis. The participants answered questions adapted from the Questionnaire for User Interface Satisfaction (QUIS) [12] and the System Usability Scale (SUS) [7, 38] questionnaire. In the *User Experience Ratings* section, we asked participants to rate how much they agreed with seven statements on a scale of 1–7, with 1 being "strongly disagree" and 7 "strongly agree." In the *System Features Ratings* section, we asked participants to rate how useful each feature was in helping them complete the design on a scale of 1–7, with 1 being "not useful at all" and 7 "extremely useful." We also collected the participants' comments on our system and the overall user experience during the interview.

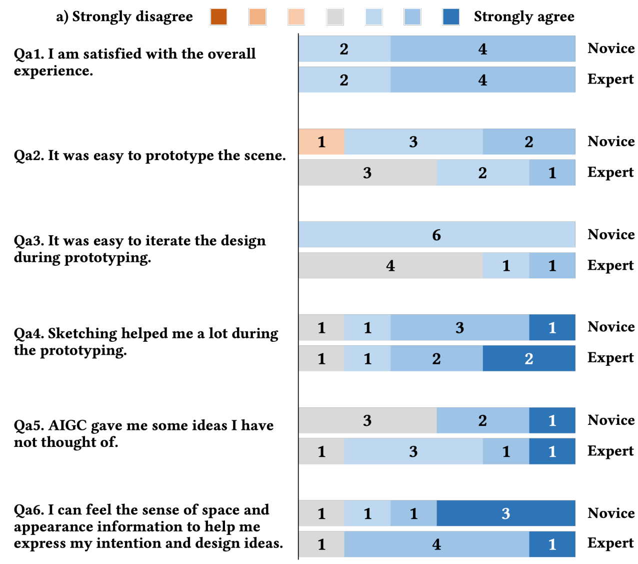

Figure 8: User experience ratings. The color of a bar represents how much participants agreed with a statement. The number in the bar represents the participants who submitted the same rating.

5.5 Results

We analyzed the differences in participants' questionnaires, interviews, logged operations, and recorded videos.

5.5.1 Overall Experience. Fig. 8 shows an overview of participants' ratings of their experience. Overall, the participants gave positive ratings on their experience using Neural Canvas (12/12 in Qa1). They found it very efficient to use the integrated AI functions with the 3D platform to create and iterate their scenic design. The ratings also showed that 3D sketching and generative AI are powerful tools for prototyping.

Neural Canvas gave the participants an intuitive sense of space and appearance. Qa6,*I can feel the sense of space and appearance information to help me express my intention and design ideas.*, received a high average score of 5.91 out of 7 (std=1.037). P2, P3, P8, P9 found it novel to sketch in an environment that they considered to have more liveliness than static 2D sketching. Additionally, the 3D platform reduced the time and effort of creation. P1, P3, P5, and P12 mentioned that they could arrange the assets once and change the location and angle of the camera to obtain a series of renderings with the same theme and asset composition but with different angles of view. Third, the 3D platform increased artistic freedom. P1, P2, P3, P4, P7, P8, P9, P10, and P12 agreed that the 3D platform allowed them to easily explore different relative locations and sizes between the assets, which helped them realize ideas they have never imagined.

As for the ease of prototyping and design iteration using Neural Canvas, most participants reported some frustration due to differences in shortcuts and mouse button functions between Neural Canvas and other 3D software that they were accustomed to. For Qa2 and Qa3, expert users gave lower scores than novice scores

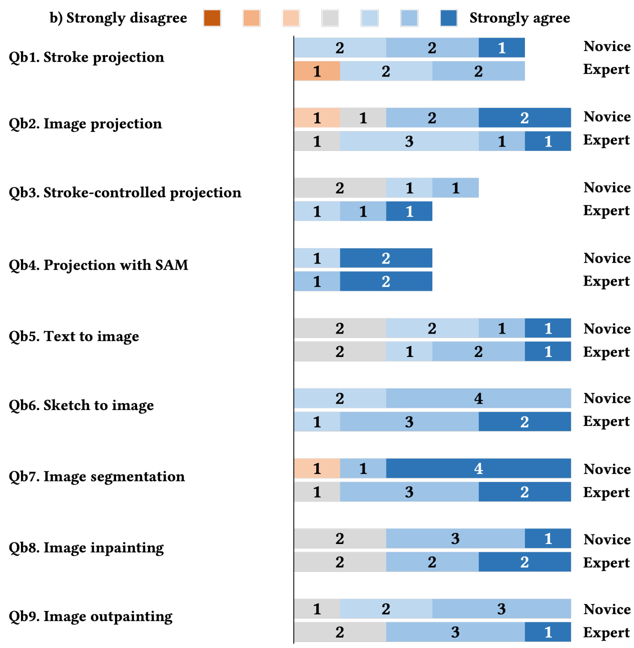

Figure 9: System features ratings. The color of a bar represents how useful participants found a feature. The number in the bar represents the participants who submitted the same rating. Not all users made use of the functions we introduced in the guide section, so some of the functions did not receive ratings from all 12 participants.

because they were more sensitive to this inconsistency with their familiar operations. P4, P5, P11, and P12 mentioned they would enjoy using Neural Canvas more if the shortcuts and key layout were more consistent with the software they are familiar with. Nevertheless, all users except P6 said it is easier to prototype and iterate scenes with Neural Canvas than with existing 3D software due to the cumbersome operations of modeling, finding materials, and adjusting lighting in 3D software. P10, a professional 2D animator, claimed *"though using Neural Canvas costs more time than sketching to prototype a scene, it gave me a sense of space and saved me from the complex work of manual coloring. And thus I could try a lot of different styles."*

5.5.2 Ratings of Main Functions. The ratings of separate features (shown in Fig. 9) prove the effectiveness of generative AI and 3D sketching in prototyping. In task 2, not all users made use of the functions we introduced in the guide section, so some of the functions did not receive ratings from all 12 participants.

According to Qb1, Qb2, Qb3, and Qb4, the four projection operations received average scores of 5.30, 5.41, 5.28, and 6.5 respectively. These four operations aim to integrate AI-generated content tightly with the 3D environment. P5, a digital artist drawing the most strokes among all participants, said *"with stroke projection, I can draw 2D sketches as usual and build 3D scenes with different layers very quickly."* P7 segmented a wall from a generated image and projected it onto a transverse plane, and commented *"image projection enables me to inversely render 2D images back to the 3D environment."* P4, P5, P8, and P12 thought stroke-controlled projection made a good combination of 3D sketching and *Sketch to Image* function, allowing them to quickly assign color and material information to their sketches.

For scenic design prototyping, AI functions were also well received. The five main AI functions introduced in the guide section received average scores of 5.25, 5.91, 6.08, 5.58, and 5.41.*Text to Image* and *Sketch to Image* inspired participants' creativity and improved efficiency. P1, P3, P8, and P9 compared these two functions to automatic coloring and emphasized their usefulness in scenic design prototyping, which provided different creation styles they had not tried before. *Image segmentation* received the highest average score for improving the user experience of obtaining desired assets. When users are interested in a part of the generated image, they can use these functions to isolate and segment that part very quickly. P1, P2, P3, P4, P5, P6, P8, P9, P10, and P12 claimed that it is especially useful to get assets to iterate design quickly.

5.5.3 Quality of Prototypes. We analyze the quality of created prototypes based on user evaluation and discuss how Neural Canvas facilitates prototyping. From Fig. 10, we can see the improvement that our system brings to the user's original hand-drawn sketches to the built 3D scenes.

Almost all participants said in Qa6 that they could feel the sense of space and appearance information brought by Neural Canvas. Compared with users in the expert group, users in the novice group are more impressed by the effects as they are not able to draw complete work. P1 said *"I didn't expect the AI to be able to generate such good pictures based on my sketches, and that I could convert them into 3D scenes."* P3 also expressed similar opinions. P9 drew a dog, which is hard for people to recognize, and said *"the generated dog is far better than what I drew. I could never draw like this."*

Expert users think the same quality of work requires them to spend several rounds of effort with other tools. For example, while Mental Canvas allows users to draw on different planes to bring depth information, creating 3D sketches with fine color and texture requires a significant amount of time and effort. Some content creators [15, 64] shared on video platforms that prototyping a scene on Mental Canvas requires dozens of hours.

During the interview, we noticed that participants in the expert group were more likely to notice artifacts and distortions in images than users in the novice group. P11 and P12 commented that segmentation by AI did not handle well at the edges of objects sometimes, and the projected images may be distorted if they are projected from the wrong points of view. Projection distortion can be avoided by undoing the operation, manually adjusting the camera position, and applying projection again. Segmentation artifacts can be alleviated by using SAM multiple times or manually editing.

6 DISCUSSION

In this section, we provide an in-depth discussion of the synergistic effects of integrating generative AI and 3D sketching, as well as new user behaviors that emerged from the process of using Neural Canvas. The discussion aims to provide insight into the future design of more user-friendly generative AI models and 3D content creation interfaces.

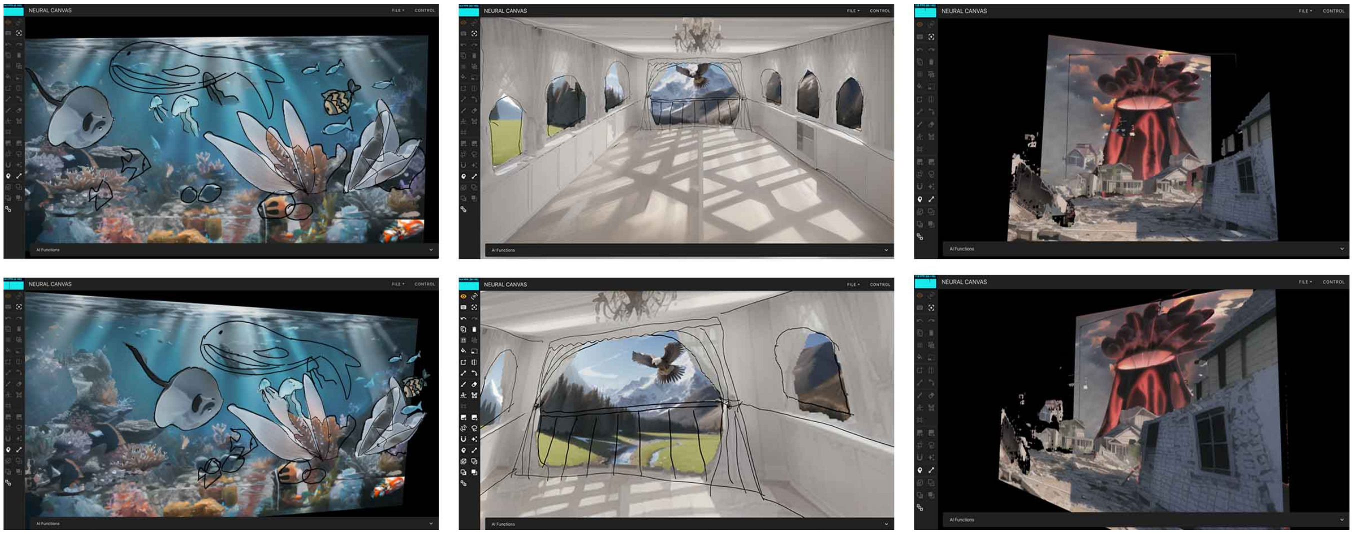

Figure 10: Two example views of scenic designs created by our study participants. Left: Theme 1 "aquarium." Middle: Theme 2 "the world outside the window." Right: Theme 3 "disaster."

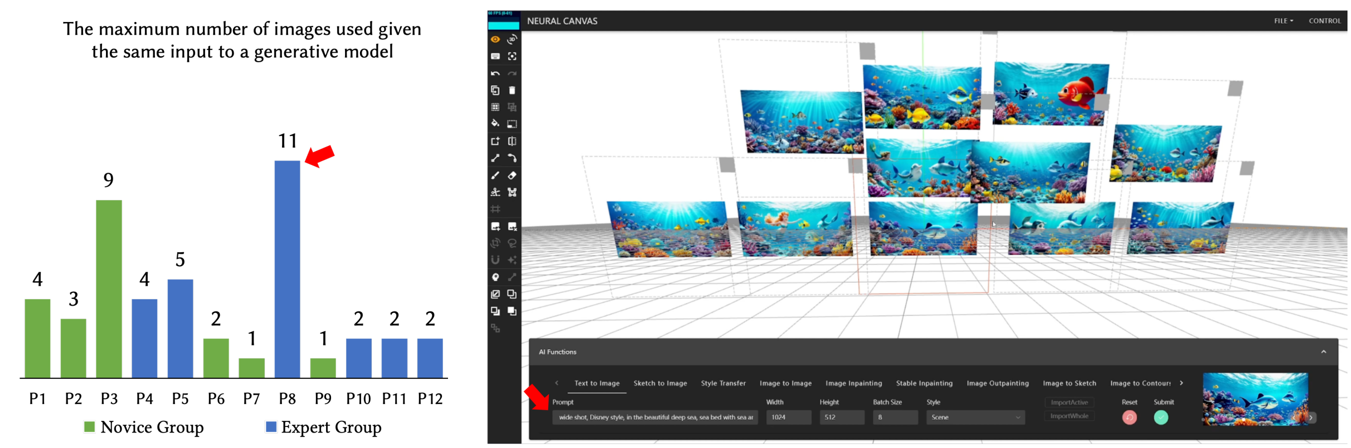

Figure 11: Visualization of how users develop multiple ideas from generated images. Left: The maximum number of images used given the same input to a generative model. Right: A screenshot of P8's design in progress. She used one prompt to generate multiple candidate images to develop the scenic design. The red arrows indicate that P8 used the same prompt to generate eleven candidate images.

6.1 How Integrating Multiple Generative AI Models Facilitates Prototyping

We observed that all participants were very willing to use AI functions during their scenic design prototyping. P1, P2, P3, P4, P7, P8, P9, and P12 mentioned that AI functions helped them a lot in designing an initial scene with mood and atmosphere.

6.1.1 Generative AI Offers New Ideas to Users. Some users emphasized that the *Text to Image* function offered them many insights by generating a lot of candidate images only from some keywords. They are willing to mix up their initial idea with synthesized images. More than two-thirds of the participants thought generative AI models gave them some ideas they had never thought of. P7 said *"It is as if an anonymous person is having a brainstorm with you."* Although *Sketch to Image* function generates images conformed to users' specified outline, the generated images still give creators more ideas in terms of color and atmosphere. P4 said *"Although I already had some rough ideas in my mind, AI showed my ideas in a more brilliant form."*

6.1.2 Generative AI Provides Appealing Content for Rapid Prototyping. Users would need to spend a lot of time drawing an image with an appealing appearance. Generative AI models could quickly externalize fleeting thoughts into images that could be refined by other AI functions or manually edited. Such a process can be seen in Fig. 10. P12 claimed that *"Neural Canvas allows users to freely explore* *which elements are needed in the initial stage, how to place them, and then make adjustments afterward. Designers have the opportunity to freely combine multiple elements and explore the possibilities of different combinations; AI can also accelerate the process of exploration."* Generative AI makes it much faster for designers to explore visual ideas and thus enables a rapid prototyping experience.

Principal benefits of effective prototyping tools include motivating users to explore a broader range of design possibilities through rapid iteration of a single concept and simultaneous exploration of various ideas [73]. Generative AI gives users a lot of candidates for design and can restart designing again from a desired time point as users wish. As shown in Fig. 11, most participants import two or more generated images for one idea into the 3D environment. They chose one image to build 3D scenes and kept some of the other images as references or tried to blend some of their elements. P6, P7, P11, and P12, four participants of the expert group, were not satisfied with their early-stage work and used previous 3D sketches to re-generate images to prototype a better 3D scene.

6.1.3 Multimodal Input Prompts to Generative AI Make Design Easier. Generative AI lowered the barrier of entry for novices. P1, P3, and P9 from the novice group mentioned that they could express their ideas clearly with sketches using AI-generated content. P9 said, *without the AI-generated content, I can only draw very simple and rough drafts, which were ineffective for sharing my ideas.* Furthermore, multiple modalities of the input to generative AI models bring users more control and creative possibilities than only using one modality. AI models with sketch prompts can offer good control of layout over generated images, while *Text to Image* models could quickly enrich the scene with very little manual effort. Almost all participants agreed that using multiple input modalities speeds up their creative process compared to using only one modality. P12 commented, *"First, I used sketching to set up objects I really need, and then I used Text to Image function to create surroundings with mood and atmosphere, which is not easy to express with strokes."*

6.1.4 Integration of Generative AI Models Improves Efficiency. The integration of generative AI models improved the creation efficiency. All participants spoke highly of integrating different generative AI models so they do not need to use them on different websites. P1, P2, P3, P4, P8, and P12 mentioned that the AI models can rapidly generate content, saving time and effort compared to searching on the internet or creating by themselves. P8 said, *It took time to consider the assets' textures, shades, and light effects in 3D software. However, generative AI can save some time so they can focus on arranging the assets and choosing angles.* P12, a professional artist, mentioned the completion time of using our system (*Text to Image* followed by *Image Segmentation*) was within one minute compared to his accustomed workflow (Midjourney [55] + Photoshop [2]) that took more than ten minutes.

6.2 How 3D Sketching Facilitates AI Content Generation

Though the input to AI models is mostly text and 2D images, our user study showed that 3D sketching plays an important role in facilitating AI content generation.

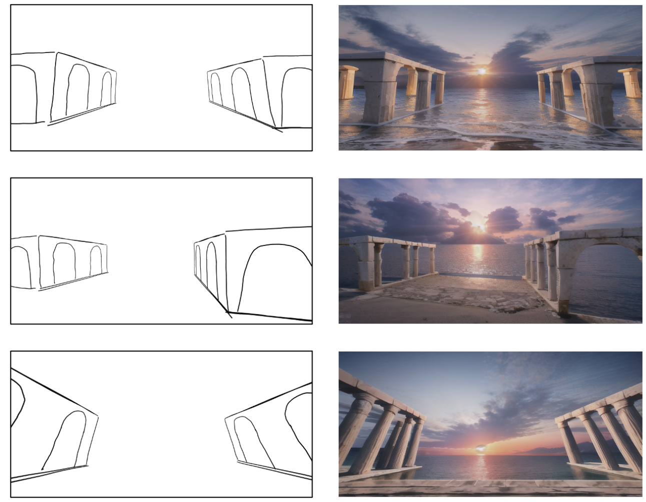

Figure 12: AI-generated images from the same 3D sketch from different views.

6.2.1 3D Sketching Offers Control over Viewpoint for Generative AI. Viewpoint control of AI-generated content is widely considered a challenging task. However, for 3D sketching, users can change the angle of view in the 3D environment, and AI models can generate images from corresponding viewpoints (shown in Fig. 12). P2, P3, and P10 used sketching to build a 3D scene and then feed AI functions with renderings from different perspectives. In this way, users directly control the perspective of generated content. P2 said *"I drew a sketch of the corridor, then viewed it from different angles, and finally adopted an angle that I was satisfied with and input it into the AI model for generation."* Designers may not have thought about the perspective from which they initially want to portray their ideas. But with 3D sketching, they can look at their designs from different angles and subsequently use the AI function to generate better renderings that meet their expectations.

6.2.2 3D Sketching Improves the Quality of Input to Generative AI. Users can check whether their sketch conforms to the perspective relationship in the 3D environment, which is very useful for those not very skilled in sketching. Most current AI models are trained on photographs and paintings by professional artists, which typically have an accurate perspective relationship. AI models will generate inferior output if the sketch prompt has poor perspectives or proportions. P1 mentioned *"I do not have a lot of drawing experience, so the sketches I draw look a bit weird; but when I map them from the 2D plane to different positions in the 3D space, I quickly found the perspective and scale relationships of my drawings are not correct, and then I can edit them very quickly."* 3D sketching could help users, especially novices, avoid errors in their sketches so that they can obtain better AI-generated output.

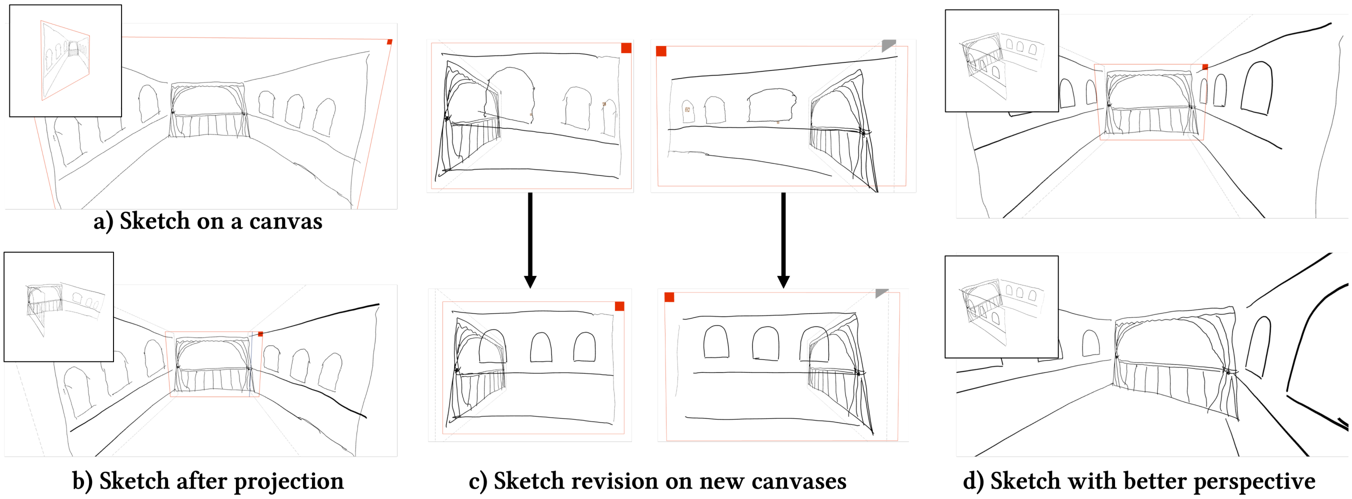

Figure 13: Correcting perspective errors in 3D sketching. Users can easily observe perspective errors with sketches after they are projected onto multiple planes, such as the distorted windows and walls at the top of (c. Users can then edit sketches quickly to obtain high-quality input to generative AI models.

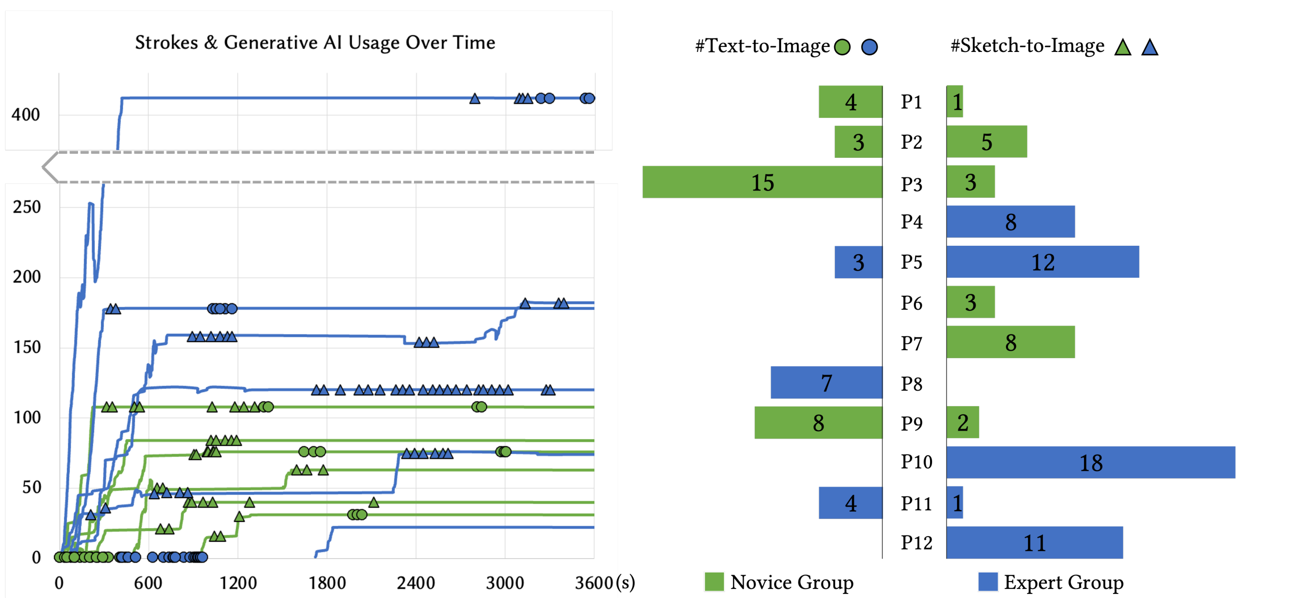

Figure 14: Visualization of user behavior patterns between novice and expert groups. Left: The number of strokes and usage of AI functions over time. Overall, experts drew more strokes than novices and tended to use *Sketch to Image* throughout the entire design process. The dots indicate *Text to Image*, and the triangles indicate *Sketch to Image*. Right: The number of *Text to Image* and *Sketch to Image* calls. Novices preferred using *Text to Image* while experts preferred using *Sketch to Image*.

6.3 Different User Behavior Between Novices and Experts

We analyzed the workflows of both novices and experts using our collected data, and visualized them in Fig. 14. We noticed some usage patterns that differ across novices and experts.

Between the *Text to Image* and *Sketch to Image* functions, novices seemed to prefer *Text to Image* while experts seemed to prefer *Sketch to Image*. According to our data, novices used image generation functions 8.6 times on average, with 5.0 out of 8.6 times being *Text to Image* and 3.6 out of 8.6 times being *Sketch to Image*. On the other hand, experts used image generation functions 10.6 times on average, with 2.3 out of 10.6 times being *Text to Image* and 8.3 out of 10.6 times being *Sketch to Image*. In addition, experts drew an average of 164.6 strokes during their design process while novices only drew 67 strokes on average, further demonstrating the experts' strong preference for sketching.

Experts also seemed to be more selective than novices about which generated images they decided to keep in their designs. Novices on average adopted 6.6 out of 8.6 generated images, whereas experts on average adopted 6.8 out of 10.6 generated images. Experts, despite generating 2.0 more images than novices, only adopted 0.2 more images in their final designs than novices.

We logged the timestamps of *Sketch to Image* and *Text to Image* used in participants' design sessions, as shown in Fig. 14. We noticed that expert users tended to use *Sketch to Image* throughout their entire design process while novice users mainly used *Text to Image* at the beginning of their design.

7 LIMITATIONS AND FUTURE WORK

Despite participants' favorable reviews of Neural Canvas, our system still has some limitations in terms of the overall user experience. This section discusses possible improvements in the future.

Layer and Perspective Selection for Projection. While projection operations have demonstrated their efficiency, some users initially had trouble understanding how they work. Additionally, a few users voiced their frustration with the need to switch viewpoints frequently during the projection process. P10 explained that "the thought process involved in projection significantly differs from the intuitive approach used in painting. Painting relies on an instinctive understanding, while projection demands a more rational consideration of perspective and spatial relationships. I plan to employ projection selectively for critical tasks, resorting to simpler operations to complete the scene." A more intuitive and streamlined projection interaction method is needed to alleviate user frustration and facilitate a smoother transition between sketching and projection. One potential solution is using image depth estimation algorithms [56, 66] to automatically provide users with candidate layers. Introducing an auxiliary view may also help users understand the 3D environment from multiple views at the same time, which is adopted by Mental Canvas [53].

Layer Management. As scenic design progresses into its later stages, the number of layers present in the scene increases substantially. The default naming scheme for layers (layer type suffixed by an index, e.g., Plane 0) makes it difficult to mentally keep track of and select layers when the number of layers is large. To address this issue and enhance the overall user experience, it is worth exploring the dynamic interactions between 3D layers, possibly with techniques such as physical gesture control [18, 47, 93]. Additionally, incorporating layer thumbnails could provide a visual aid, allowing users to swiftly identify and locate specific layers amidst the growing complexity of the scene. Furthermore, there is an exciting avenue to explore the integration of AI methods to automatically assign meaningful names to layers, thereby reducing the burden on users and streamlining their workflow. These potential improvements could significantly enhance the efficiency and usability of any 3D software, making it more accessible and user-friendly for all users.

Arsenal of Generative AI Models. In our user study, participants expressed a strong desire to employ various styles of models in their scene designs. However, it is important to note that our AI functions are limited to a pre-selected collection of generative AI models. This was an intentional design choice so that users would not be overwhelmed by too many AI models. Nonetheless, expanding the system to accommodate a broader spectrum of generative AI models is a worthwhile direction for future research. The main objective is to ensure scalability and enhance style diversity within the system. One promising approach to achieve this is to interface with model sharing and search platforms like Modelverse [50]. This would enable designers to access and incorporate a rich array of generative models, thereby fostering creativity and versatility in their 3D design.

8 CONCLUSION

In this paper, we have proposed Neural Canvas, a lightweight platform that leverages 3D sketching and the integration of multiple generative AI models to facilitate scenic design prototyping. With Neural Canvas, designers can easily draw sketches and place them in the 3D environment to embed spatial information. The integration and interaction of multiple cutting-edge deep generative models also make it easier for designers to create attractive and detailed content from sketches for scenic design prototyping.

Our user study showed that Neural Canvas helped users easily express their early-stage visuospatial ideas and concepts. By viewing scenes from different perspectives in 3D space and integrating multiple AI generative models, users can rapidly express and iterate their design ideas and freely explore new possibilities.

We also observed from our user study that 3D sketching benefits AI content generation. Such findings may lead to a new and more controllable modality of prompt in the future generative AI research. In addition, we also proved that integrating multiple generative AI models facilitates scenic design prototyping in many ways.

We think Neural Canvas can serve as a creativity support tool in scenic design prototyping, and we plan to release the platform and support user-customized generative AI models so that creative communities can maximize the synergy between 3D sketching and the latest AI technologies.

ACKNOWLEDGMENTS

We would like to thank Jiayang Huang and Yiran Chen for their creative work using Neural Canvas and thank the reviewers for their valuable comments and helpful suggestions. This work is supported by Guangzhou-HKUST(GZ) Joint Funding #2023A03J0670 and Guangzhou Basic Research Scheme #2024A04J4229.

REFERENCES

[1] Abubakar Abid, Ali Abdalla, Ali Abid, Dawood Khan, Abdulrahman Alfozan, and James Zou. 2019. Gradio: Hassle-Free Sharing and Testing of ML Models in the Wild. *arXiv preprint 1906.02569* (2019).

[2] Adobe. 2023. Photoshop. https://www.adobe.com/products/photoshop.html. Accessed Apr 5, 2023.

[3] Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik. 2010. Contour Detection and Hierarchical Image Segmentation. *IEEE transactions on pattern analysis and machine intelligence* 33, 5 (2010), 898–916.

[4] Arnold Aronson. 2005. *Looking into the Abyss: Essays on Scenography.* University of Michigan Press.

[5] Rahul Arora, Rubaiat Habib Kazi, Tovi Grossman, George Fitzmaurice, and Karan Singh. 2018. SymbiosisSketch: Combining 2D & 3D Sketching for Designing Detailed 3D Objects in Situ. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association

for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3173574.3173759

[6] Rahul Arora, Rubaiat Habib Kazi, Fraser Anderson, Tovi Grossman, Karan Singh, and George W Fitzmaurice. 2017. Experimental Evaluation of Sketching on Surfaces in VR.. In *CHI*, Vol. 17. 5643–5654.

[7] Aaron Bangor, Philip T Kortum, and James T Miller. 2008. An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction* 24, 6 (2008), 574–594.

[8] Blockade Labs. 2023. Skybox - AI-Generated 3D Worlds. https://skybox.blockadelabs.com/. Accessed Sep 12, 2023.

[9] Gary Bradski. 2000. The OpenCV Library. *Dr. Dobb's Journal of Software Tools* (2000).

[10] Ricardo Cabello. 2023. three.js. https://threejs.org/. Accessed Apr 5, 2023.

[11] Xi Chen, Zhiyan Zhao, Feiwu Yu, Yilei Zhang, and Manni Duan. 2021. Conditional Diffusion for Interactive Segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 7345–7354.

[12] John P Chin, Virginia A Diehl, and Kent L Norman. 1988. Development of an Instrument Measuring User Satisfaction of the Human-Computer Interface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 213–218.

[13] SolidWorks Corporation. 2023. SOLIDWORKS | 3D CAD Design Software & PDM Systems. https://www.solidworks.com/. Accessed Sep 12, 2023.

[14] Thomas Cover and Peter Hart. 1967. Nearest Neighbor Pattern Classification. *IEEE Transactions on Information Theory* 13, 1 (1967), 21–27.

[15] Fan DanDan. 2023. Interstellar: A One-Shot Scene Created Using Procreate and Mental Canvas With One Month's Hardworking. https://www.bilibili.com/video/BV1k24y16761. Accessed Dec 13, 2023.

[16] Prafulla Dhariwal and Alexander Nichol. 2021. Diffusion Models Beat Gans on Image Synthesis. *Advances in neural information processing systems* 34 (2021), 8780–8794.

[17] Julie Dorsey, Songhua Xu, Gabe Smedresman, Holly Rushmeier, and Leonard McMillan. 2007. The Mental Canvas: A Tool for Conceptual Architectural Design and Analysis. In *Proceedings of the Pacific Conference on Computer Graphics and Applications (PG)*. IEEE, 201–210.

[18] Tomas Dorta, Gokce Kinayoglu, and Michael Hoffmann. 2014. Hyve-3D: A New Embodied Interface for Immersive Collaborative 3D Sketching. In *ACM SIGGRAPH 2014 Studio*. 1–1.

[19] Catherine Elsen, Anders Häggman, Tomonori Honda, and Maria C Yang. 2012. Representation in Early Stage Design: An Analysis of the Influence of Sketching and Prototyping in Design Projects. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Vol. 45066. American Society of Mechanical Engineers, 737–747.

[20] Inc. Epic Games. 2023. The most powerful real-time 3D creation tool - Unreal Engine. https://www.unrealengine.com/. Accessed Sep 12, 2023.

[21] John Flynn, Michael Broxton, Paul Debevec, Matthew DuVall, Graham Fyffe, Ryan Overbeck, Noah Snavely, and Richard Tucker. 2019. Deepview: View Synthesis With Learned Gradient Descent. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2367–2376.

[22] Blender Foundation. 2023. Blender: A 3D Modeling and Rendering Package. http://www.blender.org. Accessed Apr 5, 2023.

[23] J Michael Gillette. 2012. *Theatrical Design and Production: An Introduction to Scene Design and Construction, Lighting, Sound, Costume, and Makeup.* McGraw-Hill Higher Education.

[24] Michael Graves. 2005. The Necessity for Drawing. In *Michael Graves: Images of a Grand Tour*. Princeton Architectural Press, New York, NY, 235–245. https://doi.org/10.1007/1-56898-657-2_9

[25] Yuxuan Han, Ruicheng Wang, and Jiaolong Yang. 2022. Single-View View Synthesis in the Wild with Learned Adaptive Multiplane Images. In *ACM SIGGRAPH 2022 Conference Proceedings*. 1–8.

[26] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. 2020. Array Programming with NumPy. *Nature* 585, 7825 (Sept. 2020), 357–362. https://doi.org/10.1038/s41586-020-2649-2

[27] William E Hick. 1952. On the Rate of Gain of Information. *Quarterly Journal of experimental psychology* 4, 1 (1952), 11–26.

[28] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising Diffusion Probabilistic Models. *Advances in neural information processing systems* 33 (2020), 6840–6851.

[29] Derek Hoiem, Alexei A. Efros, and Martial Hebert. 2005. Automatic Photo Pop-Up. In *ACM SIGGRAPH 2005 Papers* (Los Angeles, California) *(SIGGRAPH '05)*. Association for Computing Machinery, New York, NY, USA, 577–584. https://doi.org/10.1145/1186822.1073232

[30] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. LoRA: Low-Rank Adaptation of Large Language Models. In *International Conference on Learning Representations*.

[31] Xun Huang, Arun Mallya, Ting-Chun Wang, and Ming-Yu Liu. 2022. Multimodal Conditional Image Synthesis with Product-of-Experts GANs. In *Proceedings of the European Conference on Computer Vision*. Springer, 91–109.

[32] Ray Hyman. 1953. Stimulus Information as a Determinant of Reaction Time. *Journal of experimental psychology* 45, 3 (1953), 188.

[33] Autodesk Inc. 2023. Autodesk 3ds Max: Create massive worlds and high-quality designs. https://www.autodesk.com/products/3ds-max. Accessed Sep 12, 2023.

[34] Autodesk Inc. 2023. Autodesk Inventor: Mechanical design software for ambitious ideas. https://www.autodesk.com/products/inventor. Accessed Sep 12, 2023.

[35] Autodesk Inc. 2023. Autodesk Maya: Create expansive worlds, complex characters, and dazzling effect. https://www.autodesk.com/products/maya. Accessed Sep 12, 2023.

[36] Trimble Inc. 2023. 3D Design Software | 3D Modeling on the Web | SketchUp. https://www.sketchup.com/. Accessed Sep 10, 2023.

[37] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-to-Image Translation with Conditional Adversarial Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1125–1134.

[38] Patrick W Jordan, Bruce Thomas, Ian Lyall McClelland, and Bernard Weerdmeester. 1996. *Usability Evaluation in Industry*. CRC Press.

[39] Sing Bing Kang. 1998. Survey of Image-Based Rendering Techniques. In *Videometrics VI*, Vol. 3641. SPIE, 2–16.

[40] Tero Karras, Samuli Laine, and Timo Aila. 2019. A Style-Based Generator Architecture for Generative Adversarial Networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4401–4410.

[41] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. 1988. Snakes: Active Contour Models. *International journal of computer vision* 1, 4 (1988), 321–331.

[42] Rubaiat Habib Kazi, Tovi Grossman, Hyunmin Cheong, Ali Hashemi, and George W Fitzmaurice. 2017. DreamSketch: Early Stage 3D Design Explorations with Sketching and Generative Design. In *Proceedings of the ACM Symposium on User Interface Software and Technology*. 401–414.

[43] Florian Kern, Peter Kullmann, Elisabeth Ganal, Kristof Korwisi, René Stingl, Florian Niebling, and Marc Erich Latoschik. 2021. Off-The-Shelf Stylus: Using XR Devices for Handwriting and Sketching on Physically Aligned Virtual Surfaces. *Frontiers in Virtual Reality* 2 (2021), 684498.

[44] Yongkwan Kim, Sang-Gyun An, Joon Hyub Lee, and Seok-Hyung Bae. 2018. Agile 3D Sketching with Air Scaffolding. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173812

[45] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. 2023. Segment Anything. *arXiv preprint arXiv:2304.02643* (2023).

[46] Kin Chung Kwan and Hongbo Fu. 2019. Mobi3DSketch: 3D Sketching in Mobile AR. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–11.

[47] Paul Lapides, Ehud Sharlin, Mario Costa Sousa, and Lisa Streit. 2006. The 3D Tractus: A Three-Dimensional Drawing Board. In *First IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP'06)*. IEEE, 8–pp.

[48] Mengtian Li, Zhe Lin, Radomír M˘ ech, Ersin Yumer, and Deva Ramanan. 2019. Photo-Sketching: Inferring Contour Drawings from Images. In *Proceedings of the Winter Conference on Applications of Computer Vision*.

[49] Yuwei Li, Xi Luo, Youyi Zheng, Pengfei Xu, and Hongbo Fu. 2017. SweepCanvas: Sketch-based 3D Prototyping on an RGB-D Image. In *Proceedings of the ACM Symposium on User Interface Software and Technology*. 387–399.

[50] Daohan Lu, Sheng-Yu Wang, Nupur Kumari, Rohan Agarwal, David Bau, and Jun-Yan Zhu. 2022. Content-Based Search for Deep Generative Models. *arXiv preprint 2210.03116* (2022).

[51] Maxon. 2023. Cinema 4D: 3D Computer Animation, Modeling, Simulation, and Rendering Software. https://www.maxon.net/en/cinema-4d. Accessed Apr 5, 2023.

[52] Maxon. 2023. ZBrush. https://www.maxon.net/en/zbrush. Accessed Sep 12, 2023.

[53] Mental Canvas, Inc. 2024. Mental Canvas. https://www.mentalcanvas.com/. Accessed Feb 20, 2024.

[54] Meshy. 2023. Meshy | Create 3D Game Assets with Generative AI. https://www.meshy.ai/. Accessed Sep 12, 2023.

[55] Midjourney. 2023. Midjourney. https://www.midjourney.com/. Accessed Sep 12, 2023.

[56] Yue Ming, Xuyang Meng, Chunxiao Fan, and Hui Yu. 2021. Deep Learning for Monocular Depth Estimation: A Review. *Neurocomputing* 438 (2021), 14–33.

[57] Michael Nebeling and Katy Madier. 2019. 360proto: Making Interactive Virtual Reality & Augmented Reality Prototypes From Paper. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.

[58] Byong Mok Oh, Max Chen, Julie Dorsey, and Frédo Durand. 2001. Image-Based Modeling and Photo Editing. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '01)*. Association for Computing Machinery, New York, NY, USA, 433–442. https://doi.org/10.1145/

383259.383310

[59] Luke Olsen, Faramarz F Samavati, Mario Costa Sousa, and Joaquim A Jorge. 2009. Sketch-Based Modeling: A Survey. *Computers & Graphics* 33, 1 (2009), 85–103.

[60] OpenAI. 2023. DALL·E 2: Creating Realistic Images and Art from a Description in Natural Language. https://openai.com/product/dall-e-2. Accessed Apr 5, 2023.

[61] Patrick Paczkowski, Min H. Kim, Yann Morvan, Julie Dorsey, Holly Rushmeier, and Carol O'Sullivan. 2011. Insitu: Sketching Architectural Designs in Context. *ACM Trans. Graph.* 30, 6 (Dec. 2011), 1–10. https://doi.org/10.1145/2070781.2024216

[62] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. 2019. Semantic Image Synthesis with Spatially-Adaptive Normalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 2337–2346.

[63] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 8024–8035.

[64] Dong Men Ban Qiao. 2021. 50 Hours to Build the Magnificent Ancient Mythical World: Your First Experience With Mental Canvas. https://www.bilibili.com/video/BV12b4y1871E. Accessed Dec 13, 2023.

[65] Shwetha Rajaram, Franziska Roesner, and Michael Nebeling. 2023. Reframe: An Augmented Reality Storyboarding Tool for Character-Driven Analysis of Security & Privacy Concerns. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–15.

[66] René Ranftl, Alexey Bochkovskiy, and Vladlen Koltun. 2021. Vision Transformers for Dense Prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 12179–12188.

[67] Will Reese. 2008. Nginx: the High-Performance Web Server and Reverse Proxy. *Linux Journal* 2008, 173 (2008), 2.

[68] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis with Latent Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10684–10695.

[69] Armin Ronacher. 2023. Flask. https://flask.palletsprojects.com/en/2.2.x/. Accessed Apr 5, 2023.

[70] Victoria Rudakova, Nathan Lin, Natallia Trayan, Tevfik Metin Sezgin, Julie Dorsey, and Holly E Rushmeier. 2017. CHER-ish: A Sketch-and Image-based System for 3D Representation and Documentation of Cultural Heritage Sites. In *GCH*. 195–199.

[71] Jonathan Shade, Steven Gortler, Li-wei He, and Richard Szeliski. 1998. Layered Depth Images. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques*. 231–242.

[72] Yifei Shen, Zeyu Wang, Qinying Sun, Anne Chen, and Holly Rushmeier. 2021. Reconstructing Dura-Europos From Sparse Photo Collections Using Deep Contour Extraction. In *Eurographics Workshop on Graphics and Cultural Heritage*, Alan Chalmers and Vedad Hulusic (Eds.). The Eurographics Association.

[73] Ben Shneiderman. 2007. Creativity Support Tools: Accelerating Discovery and Innovation. *Commun. ACM* 50, 12 (2007), 20–32.

[74] Jamie Shotton, John Winn, Carsten Rother, and Antonio Criminisi. 2006. Textonboost: Joint Appearance, Shape and Context Modeling for Multi-Class Object Recognition and Segmentation. In *Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I 9*. Springer, 1–15.

[75] Harry Shum and Sing Bing Kang. 2000. Review of Image-Based Rendering Techniques. In *Visual Communications and Image Processing 2000*, Vol. 4067. SPIE, 2–13.

[76] SideFX. 2023. Houdini | 3D Procedural Software for Film, TV & Gamedev | SideFX. https://www.sidefx.com/products/houdini/. Accessed Sep 12, 2023.

[77] Noah Snavely, Steven M. Seitz, and Richard Szeliski. 2006. Photo Tourism: Exploring Photo Collections in 3D. *ACM Transactions on Graphics* 25, 3 (July 2006), 835—-846.

[78] Noah Snavely, Steven M Seitz, and Richard Szeliski. 2008. Modeling the World from Internet Photo Collections. *International Journal of Computer Vision* 80 (2008), 189–210.

[79] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. 2015. Deep Unsupervised Learning Using Nonequilibrium Thermodynamics. In *International Conference on Machine Learning*. PMLR, 2256–2265.

[80] Robin Strudel, Ricardo Garcia, Ivan Laptev, and Cordelia Schmid. 2021. Segmenter: Transformer for Semantic Segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*. 7262–7272.

[81] Roman Suvorov, Elizaveta Logacheva, Anton Mashikhin, Anastasia Remizova, Arsenii Ashukha, Aleksei Silvestrov, Naejin Kong, Harshith Goka, Kiwoong Park, and Victor Lempitsky. 2021. Resolution-robust Large Mask Inpainting with Fourier Convolutions. *arXiv preprint arXiv:2109.07161* (2021).

[82] Ryo Suzuki, Rubaiat Habib Kazi, Li-Yi Wei, Stephen DiVerdi, Wilmot Li, and Daniel Leithinger. 2020. RealitySketch: Embedding Responsive Graphics and Visualizations in AR with Dynamic Sketching. In *Proceedings of the ACM Symposium on User Interface Software and Technology*. 166–181.

[83] John Sweller. 1988. Cognitive Load During Problem Solving: Effects on Learning. *Cognitive science* 12, 2 (1988), 257–285.

[84] Robert Endre Tarjan. 1975. Efficiency of a Good but Not Linear Set Union Algorithm. *Journal of the ACM (JACM)* 22, 2 (1975), 215–225.

[85] TECHNIA. 2023. CATIA | TECHNIA (US). https://www.technia.us/software/catia/. Accessed Sep 12, 2023.

[86] Unity Technologies. 2023. Unity Real-Time Development Platform | 3D, 2D, VR & AR Engine. https://unity.com/. Accessed Sep 12, 2023.

[87] Osama Tolba, Julie Dorsey, and Leonard McMillan. 1999. Sketching With Projective 2D Strokes. In *Proceedings of the 12th annual ACM symposium on User interface software and technology*. 149–157.

[88] Osama Tolba, Julie Dorsey, and Leonard McMillan. 2001. A Projective Drawing System. In *Proceedings of the 2001 symposium on Interactive 3D graphics*. 25–34.

[89] Barbara Tversky. 2002. What do Sketches say about Thinking. In *2002 AAAI Spring Symposium, Sketch Understanding Workshop, Stanford University, AAAI Technical Report SS-02-08*, Vol. 148. 151.

[90] Andrey Voynov, Kfir Aberman, and Daniel Cohen-Or. 2023. Sketch-Guided Text-To-Image Diffusion Models. In *ACM SIGGRAPH 2023 Conference Proceedings*. 1–11.

[91] Qianqian Wang, Zhicheng Wang, Kyle Genova, Pratul P Srinivasan, Howard Zhou, Jonathan T Barron, Ricardo Martin-Brualla, Noah Snavely, and Thomas Funkhouser. 2021. Ibrnet: Learning Multi-View Image-Based Rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4690–4699.

[92] Zeyu Wang, Weiqi Shi, Kiraz Akoglu, Eleni Kotoula, Ying Yang, and Holly Rushmeier. 2018. CHER-Ob: A Tool for Shared Analysis and Video Dissemination. *Journal on Computing and Cultural Heritage (JOCCH)* 11, 4 (2018), 1–22.

[93] Min Xin, Ehud Sharlin, and Mario Costa Sousa. 2008. Napkin Sketch: Handheld Mixed Reality 3D Sketching. In *Proceedings of the 2008 ACM symposium on Virtual reality software and technology*. 223–226.

[94] Evan You. 2023. Vue.js. https://vuejs.org/. Accessed Apr 5, 2023.

[95] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2023. Adding Conditional Control to Text-to-Image Diffusion Models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 3836–3847.

[96] Xiaoming Zhao, Fangchang Ma, David Güera, Zhile Ren, Alexander G Schwing, and Alex Colburn. 2022. Generative Multiplane Images: Making a 2D GAN 3D-Aware. In *Proceedings of the European Conference on Computer Vision (ECCV)*. Springer, 18–35.

[97] Youyi Zheng, Han Liu, Julie Dorsey, and Niloy J Mitra. 2016. SmartCanvas: Context-inferred Interpretation of Sketches for Preparatory Design Studies. In *Proceedings of the 37th Annual Conference of the European Association for Computer Graphics*. 37–48.

[98] Bolei Zhou, Hang Zhao, Xavier Puig, Tete Xiao, Sanja Fidler, Adela Barriuso, and Antonio Torralba. 2019. Semantic Understanding of Scenes Through the ADE20K Dataset. *International Journal of Computer Vision* 127, 3 (2019), 302–321.

[99] Tinghui Zhou, Richard Tucker, John Flynn, Graham Fyffe, and Noah Snavely. 2018. Stereo Magnification: Learning View Synthesis Using Multiplane Images. *ACM Transactions on Graphics* 37, 4 (2018), 1–12.

A SUBMODULES OF EDITOR MODULE

Other submodules of the editor module are detailed in this section.

Sketch Editor. Neural Canvas uses 3D vector lines for sketching, enhancing the user's ability to create and visualize drawings with a greater sense of depth and spatial awareness while also providing an immersive and intuitive drawing experience. The sketch editor consists of functions for drawing, deletion, splitting, polygon drawing, color selection, and projection.

State manager. The state manager holds a series of states in the current project, including objects that need to be rendered, the current operating mode of the platform, redo and undo queues, and other common basic functions.

View controller. We provide separate view modes for creating and examining a scene respectively. The first enables mouse control of the rotation, pan, and zoom of the field of view. The second enables navigation through the scene in a first-person perspective using the arrow keys.

Canvas controller. We use canvases, i.e., planes with unlimited size placed in 3D space, as the basic medium of creation. Users can draw strokes and edit images on a canvas. Users can transform and rotate these planar canvases in 3D space to embed spatial information. To better embed the spatial relationship in each plane, the canvas controller allows hinge rotation, which is the rotation of a plane around the intersection line of two planes.

File manager. The file manager of our system is designed to facilitate efficient input and output of 3D scene data. All relevant scene information, such as geometry, strokes, texture, and material, can be serialized and stored in a JSON file. This functionality enables designers to save and load working projects for further complex design, making it easier to collaborate on projects with others.

Layer manager. Layer management is an essential feature in 3D platforms that allows users to efficiently organize and manipulate various elements or objects within a scene. Layers in a 3D platform serve as containers that hold different objects or components, making it easier to work with complex scenes by isolating or grouping elements. The layer manager in Neural Canvas provides functions including adding, deleting, renaming, grouping, and regrouping layers, as well as visibility control.

Image editor. The image editor enables image uploading, deletion, scaling, lasso, magnet lasso, etc. These functions provide designers with more freedom and more fine-grained operations.

B OTHER LOADED AI FUNCTIONS

The AI module currently contains thirteen functions. Some of them are deployed on our back-end server, and others are offered by companies or public applications deployed on the cloud.

Image to Image. The image-to-image method converts an image to another with a similar color distribution. We decide to implement ControlNet [95] with Stable Diffusion. Therefore, users can generate stylistically congruent components.

Style Transfer. Style transfer can transform the style expression of images based on descriptive prompts. We choose Stable Diffusion for style transfer. Although other prior studies [37, 40] have had great success in style transfer, the huge volume of training data of Stable Diffusion makes it suitable for more styles. Additionally, the open-source community provides a large number of model weights tailored toward different styles.

Image Inpainting with Prompts. We introduce another image inpainting method based on Stable Diffusion [68]. Users can use texts to describe objects to fill holes under the image context. Designers may gain more ideas from this method, as the AI model may generate unexpected elements.

Image to Sketch. We introduce *Image to Sketch* to enhance the fusion of drawn strokes and generated images in sketch-based expression. To convert existing images to detailed sketches, we built a pipeline leveraging Stable Diffusion [68], ControlNet [95] and LoRA [30]. We use the model weights of LoRA trained on a sketch dataset to force Stable Diffusion to generate sketches, while ControlNet forces Stable Diffusion to generate content based on input images. We also load PhotoSketching [48] as an alternative, which is a legacy deep model to extract the contours from images.

Image to Contours. When users use *Image to Sketch* for sketch generation, results might be too complicated or filled with cluttered lines. Therefore, we introduce *Image to Contour* based on PhotoSketching [48] for transformation from images to sketches. Users can make scene design prototypes with simplified sketches.

Figure 15: Example output from the image-to-sketch function.

Segmentation to Image. Designers can use color blocks to draw a semantic map to represent the layout of a scene. With built-in semantic brushes, users can draw semantic maps that can be used to generate pictures as input for *Segmentation to Image*. Existing deep generative models [31, 62] can synthesize images from semantic maps. In Neural Canvas, we use ADE20K [98] protocol to define color-semantic pairs. Two models are integrated into our platform: GauGAN [62] provided by the original authors on the cloud and ControlNet [95] deployed on our back-end server.

Layer Split by Depth. We provide an alternative workflow allowing users to split layers besides various lasso tools and SAM, which leverages the depth information of images to split different layers automatically. First, we leverage a single view depth estimation model called DPT [66] to predict the depth of each pixel. Then, we use the KNN [14] clustering algorithm to decide the layer each pixel belongs to, where the number of clusters is the number of split layers specified by users.

Skybox Generation. We set a sphere geometry as the sky dome in our system for better immersion. We introduce *Skybox Generation* provided by Blockade Labs [8] to generate stylistic textures of the sky dome with text prompts.

C QUESTIONS OF THE SEMI-STRUCTURED INTERVIEW

- What features are easy to use? What features gave you a hard time? What can we do to further improve the user experience with Neural Canvas?
- What AI functions did you use in your design, except those introduced in the guide section? And why did you use them?
- Compared with traditional sketching and existing 3D software respectively, which features of Neural Canvas give you deep impressions?
- Please list some AI functions you used in the tasks and illustrate why you used them together. How can you realize the same effect if there is no Neural Canvas? How much more time and effort would it take?
- How do you think 3D sketching helps generative AI? How do you like the integration of 3D sketching and generative AI in Neural Canvas?