



# ImmerseSketch:

## Transforming Creative Prompts into Vivid 3D Environments in VR

SIGGRAPH 2024  
DENVER+ 28 JUL – 1 AUG

ALFRED LAN<sup>1</sup>, TAI-CHEN TSAI<sup>1</sup>, CHIH-CHUAN HUAN<sup>1</sup>, PU CHING<sup>1</sup>, TSE-YU PAN<sup>2</sup>, MIN-CHUN HU<sup>1</sup>

<sup>1</sup>NATIONAL TSING HUA UNIVERSITY, <sup>2</sup>NATIONAL TAIWAN UNIVERSITY OF SCIENCE AND TECHNOLOGY

### PROBLEM

- Artists utilize reference in their artistic process, yet most available reference are confined to flat images, which do not provide a sense of depth and is limited to single angle information. Physical reference provides more information for artists to glean from, but are harder to obtain or do not exist.
- Drawing in thematically relevant environments provides inspiration and assists in ideation, but attending diverse physical environments require travel, and many fantasy themed or stylized locations do not exist in real life.

### RELATED WORK / MOTIVATION

- Past works like 360proto [Nebeling and Madier 2019] and 360theater [Speicher et al. 2021] utilize VR to preview and prototype spaces made from panoramic sketches and drawings on physical paper.
- Recent work sPellorama [Chen et al. 2023] harnesses generative AI to create 2D panoramas via users' voice prompts and are shown to the user as a skybox in VR.
- These frameworks fall short on providing three-dimensional virtual spaces or furnishing tools for active creation within VR, their main purpose is to preview ideas proposed by users. We wish to extend from that concept and develop a system that generates a variety of 3D environments and objects that also provides tools for users' to sketch and ideate with.

### APPROACH

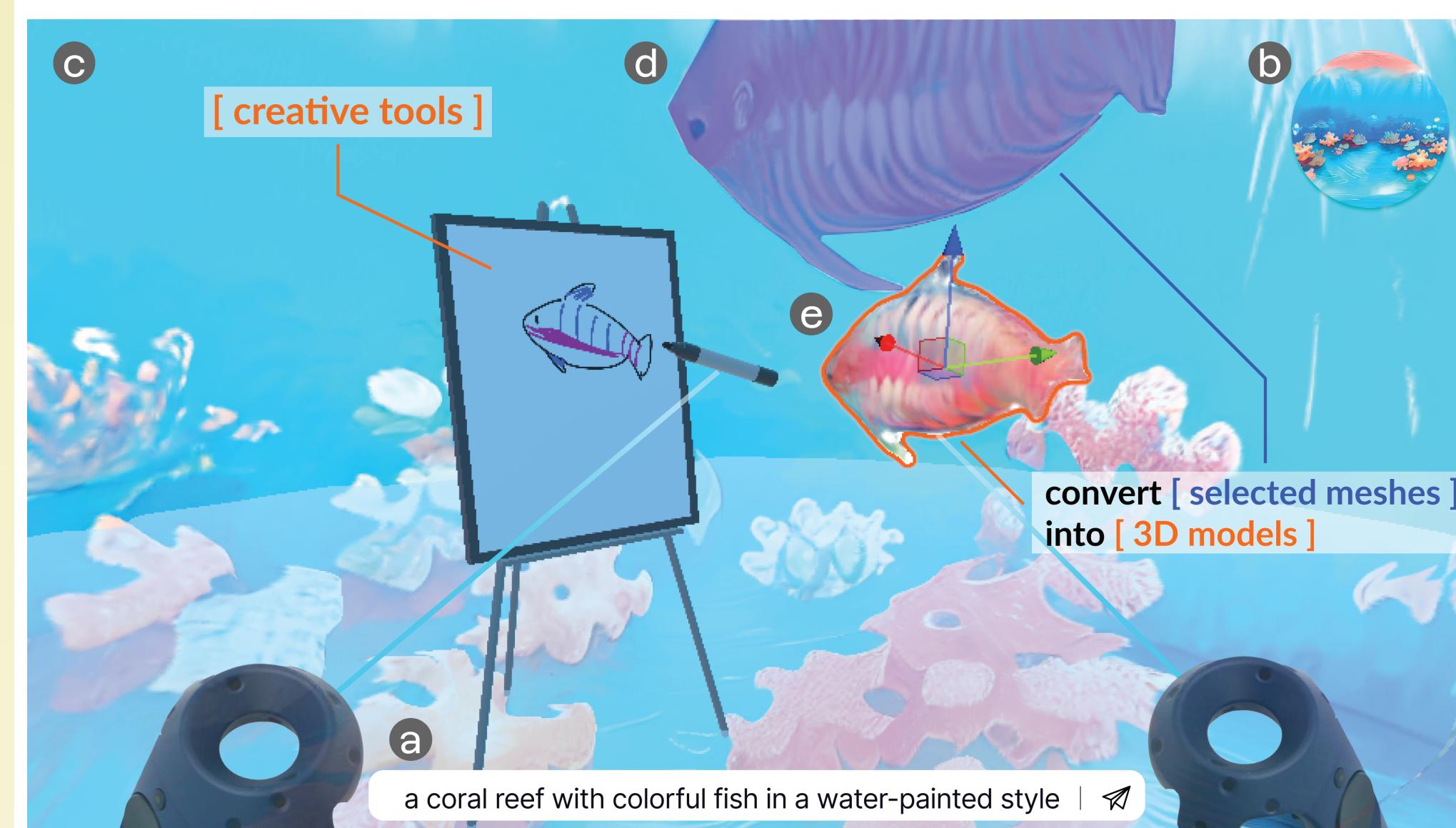
We utilize VR to tap into its ability to give users' a sense of depth, physicality, and presence. With generative AI, we can provide users with virtual environments that are stylized or ones that do not exist in real life. We can also provide users with objects that they can observe and interact with similar to physical reference.

### REFERENCES

- [1] Timothy Chen, Miguel Ying Jie Then, Jing-Yuan Huang, Yang-Sheng Chen, Ping-Hsuan Han, and Yi-Ping Hung. 2023. sPellorama: An Immersive Prototyping Tool using Generative Panorama and Voice-to-Prompts. In ACM SIGGRAPH 2023 Posters (Los Angeles, CA, USA) (SIGGRAPH '23). Association for Computing Machinery, New York, NY, USA, Article 42, 2 pages. <https://doi.org/10.1145/3588028.3603667>
- [2] Mengyang Feng, Jinlin Liu, Miaoqiao Cui, and Xuansong Xie. 2023. Diffusion360: Seamless 360 Degree Panoramic Image Generation based on Diffusion Models. arXiv preprint arXiv:2311.13141 (2023).
- [3] Minghua Liu, Chao Xu, Haian Jin, Linghao Chen, Mukund Varma T, Zexiang Xu, and Hao Su. 2024. One-2-3-45: Any single image to 3d mesh in 45 seconds without per-shape optimization. Advances in Neural Information Processing Systems 36 (2024).
- [4] 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.
- [5] Maximilian Speicher, Katy Lewis, and Michael Nebeling. 2021. Designers, the stage is yours! medium-fidelity prototyping of augmented & virtual reality interfaces with 360theater. Proceedings of the ACM on human-computer interaction 5, EICS (2021), 1–25.
- [6] Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. 2024. Depth anything: Unleashing the power of large-scale unlabeled data. arXiv preprint arXiv:2401.10891 (2024).

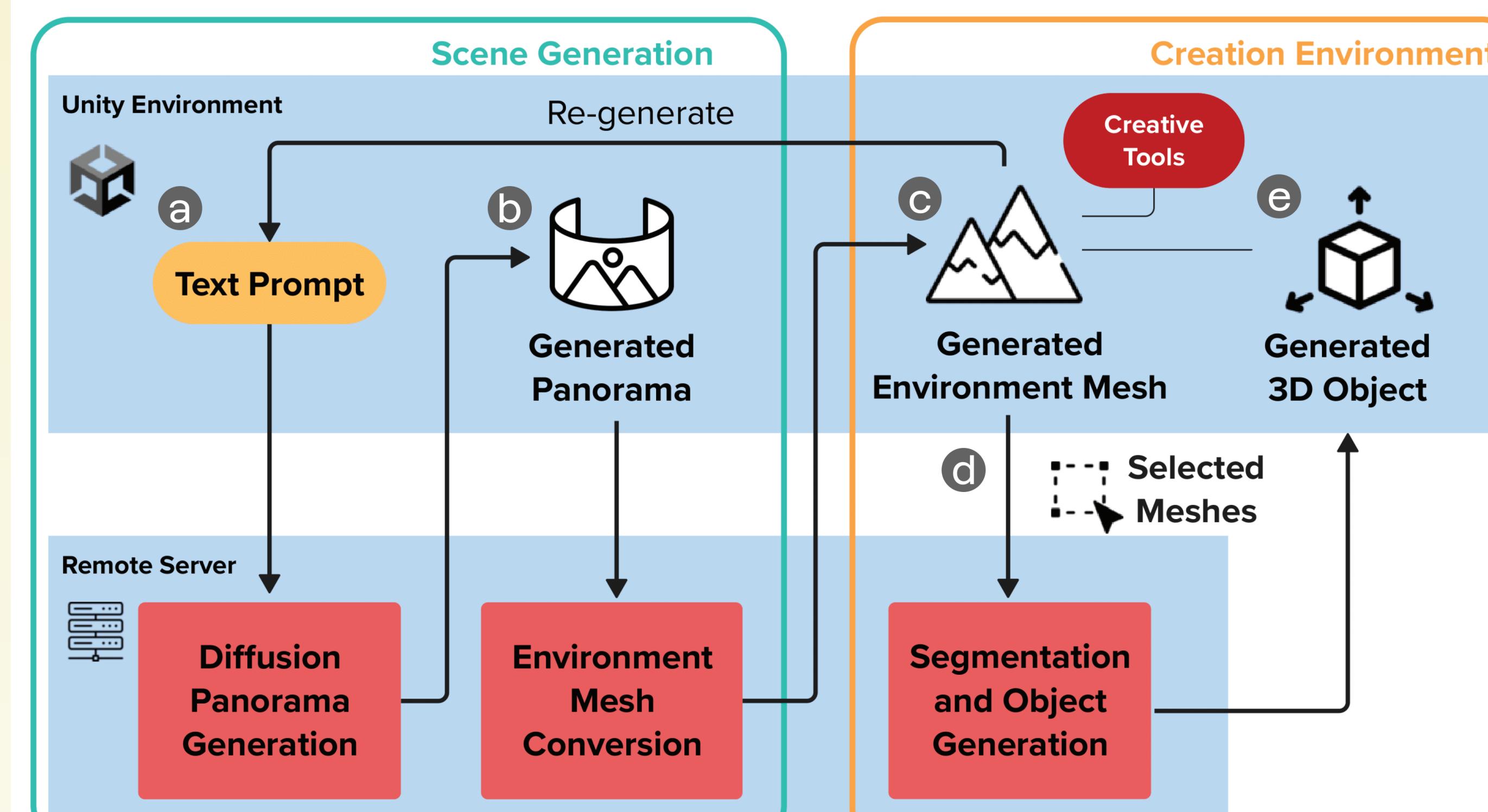
### METHOD

Our proposed VR system provides artists a method to generate 3D environments that they can create in. This process involves 3 main modules.



#### Application Scenario

- (a) Scene description given by the user
- (b.) Generated panorama
- (c.) Generated mesh of the environment converted from panorama.
- (d.) Selected section of the environment mesh
- (e.) Generated standalone 3D object based on the selected section



(e.) Generated 3D Object



To generate a standalone object, we take a screenshot from the user's viewpoint and (d.) their selection location as input to obtain a segmentation mask, we then use this segmentation to generate a 3D object model with One-2-3-45 [Liu et al. 2024].

(a.) Text Prompt

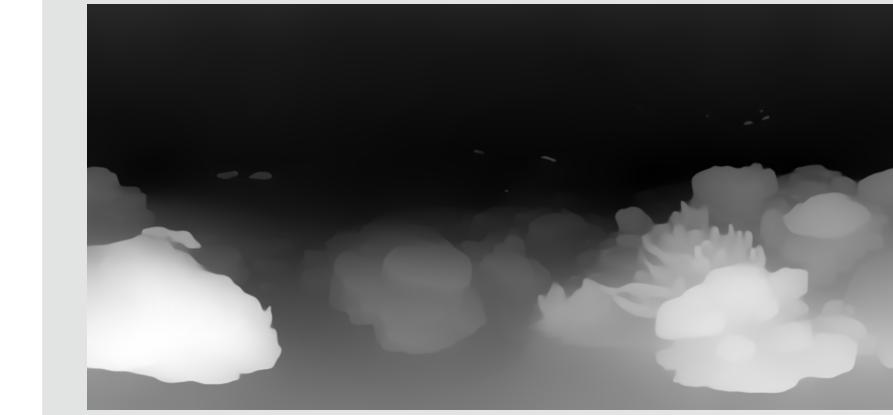
"under the ocean, water is azure with some light rays piercing the water above, shining some light into the deep blue waters, a few fishes can be seen"

(b.) Generated Panorama



Uses Diffusion360 [Feng et al. 2023], takes user given text prompt as input, outputs panoramas.

(c.) Generated Environment Mesh



Takes selected panorama as input and outputs a 3D environment mesh by combining the original image with a depth map generated using DepthAnything [Yang et al. 2024].

### RESULTS



Top: A participant sketching with our system  
Right: Several sketches made by different participants during the study



- The pilot study was conducted with 14 participants who have a passion for painting. The experimental scenes were designed to cover three styles: realistic, natural, and futuristic fantasy. Each participant was instructed to create three sketches, one for each scene. Afterwards, each participant completed a System Usability Scale Questionnaire (SUS), a Slatner-Usoh-Steed Presence Questionnaire (SUS-PQ), and participated in an interview to share their experience.
- The resulting SUS score was **77.1±12 (B Grade)**, and the SUS-PQ score **4.3±1.6 (out of 7)**. During the interviews, participants unanimously agreed that the generated environments aided in ideating concepts and provided an immersive, relaxing setting for painting.
- Our system is currently in its experimental phase, offering opportunities for refinement. Our primary objective for future work is to enhance immersion and the quality of generated 3D environments. Based on participant feedback, we would like to focus on incorporating generated soundscapes to enhance immersion. Additionally, recent works such as DreamGaussian demonstrate improvements in quality and speed using 3D Gaussian Splatting. We plan to incorporate this technique to improve the consistency and quality of generated environments while maintaining efficient rendering.