Dreamcrafter: A Virtual Reality 3D Content Generation and Editing System Assisted by Generative AI

CYRUS VACHHA, ZACH DIVE, ARIA KANG, ASHWAT CHIDAMBARAM, and ANIK GUPTA

Fig. 1. System Overview: (Left) View of edits and spatial annotations made in scene. (Right) Pipeline overview of system including three distinct modules to edit scenes: edit NeRF models, create new NeRF models, and create 2D stable diffusion renders of scenes. Input files generated from Unity are in green, generative model components are in yellow, and saved results are in red.

1 INTRODUCTION

The intersection of Virtual Reality (VR) and Generative AI technologies, notably Stable Diffusion and novel 3D representations like Neural Radiance Fields (NeRF), presents an exciting frontier in the realm of 3D environment generation and design. Recent work in 3D generative diffusion models and novel 3D representations show potential as the future of 3D content creation and interaction. Traditional methods of creating and editing 3D content often entail complex interfaces and require specialized skills, limiting accessibility and stifling creativity. We propose Dreamcrafter, a Virtual Reality 3D content generation and editing system Assisted by generative AI. Our system addresses this gap by proposing a system that harnesses the immersive experience and spatial interactions of VR, coupled with the advanced capabilities of generative AI, to enhance the process of 3D environment creation and editing.

While tools like Google's Tilt Brush [4] have demonstrated the potential of VR in 3D content creation, they fall short in fully realizing the creation capabilities and details possible with generative AI. While these tools allow for creative work in 3D, they have not yet integrated generative AI capabilities to make a smoother and more efficient experience. NeRF and diffusion models offer unparalleled realism and detail in rendering, however, their integration into user-friendly platforms for 3D environment creation is still in its infancy. Dreamcrafter aims to bridge these gaps by creating a seamless and intuitive interface for users to design, modify, and generate complex and photorealistic 3D environments. Editing of radiance fields and generative 3D objects is currently limited to text prompts or limited 2D interfaces. Current research in NeRFs and diffusion models is primarily on enhancing image/reconstruction quality, and we aim to address the noticeable lack of exploration in the application of user interfaces designed for editing and controllability of these models and novel 3D representations.

VR is becoming increasingly ubiquitous with the rise of consumer displays and commercial VR platforms, requiring low latency and high-quality rendering of synthetic imagery with reduced compute overheads [7]. NeRF and 3D Gaussian Splatting (3DGS) has demonstrated that photo-realistic quality and continuous view changes of 3D scenes can be achieved without loss of view-dependent effects [7]. Dreamcrafter leverages the photorealistic rendering of NeRF and 3DGS to enable users to create realistic scenes.

At the core of our approach is a VR-based system that allows users to interact with and manipulate 3D objects and environments in real-time. Dreamcrafter involves two subsystems systems which leverage novel 3D representations and stable diffusion. The stable diffusion powered system assigns semantically mapped spatial tags to 3D primitive objects to generate stable diffusion previews of scenes. Our second subsystem leverages NeRFs and 3D Gaussian Splatting for rendering and editing of 3D photo realistic scenes. Dreamcrafter is designed to be simple to use, lowering the barrier to entry for users without extensive experience in 3D modeling, while still providing realistic output results.

In our overview figure accompanying this introduction (Fig. 1), we show the different components of our modular design. We have built a Unity app which presents two different scenes: a NeRF scene and a 3D primitive stable diffusion scene. We have separate modules that can run independently from each other, including NeRF editing with

Authors' address: Cyrus Vachha; Zach Dive; Aria Kang; Ashwat Chidambaram; Anik Gupta.

text, 3D model generation from text, and stable diffusion renderings of scenes from text. These components are modular, containing a JSON interface to allow for switching out our generative models for better ones in the future.

The anticipated outcomes of this project include a significant reduction in the complexity and time required to create detailed 3D environments. Through user studies, we aim to validate that Dreamcrafter enables both novice and experienced users to produce high-quality designs efficiently. These results would represent a substantial advancement in democratizing 3D design technology. The contributions of this paper are twofold: (1) We introduce an innovative VR-based system that integrates generative AI technologies for simplified and enhanced 3D environment design. (2) We demonstrate the system's efficacy in making complex 3D modeling processes more accessible and efficient.

2 RELATED WORK

Over the last year, promising developments in novel view synthesis and generative diffusion models (2D and 3D) have shown promising results in creating realistic results and enabling controllability of generated 3D content. There is a line of work that explores 3D generation from NeRFs [15] and 3D Gaussian Splatting (3DGS) [12] as well as 3D foundation models and creative tools to generate 3D models from text or semantic descriptions. User friendly interfaces offered by Nerfstudio [21] and Luma Labs AI [1] have increased the adoption of these 3D representations in content creation and visual effects.

2.1 Generation/editing of novel 3D representations - Neural Radiance Fields (NeRF)

In recent years, the evolution from traditional 3D graphics using meshes and geometries to more photorealistic methods such as Neural Radiance Fields (NeRFs) [15] and Gaussian Splatting [12] rendering techniques has shown to be promising in changing the way the world is being captured for VR experiences. These techniques have been shown to be highly effective at modeling details with realistic lighting, shadowing, and surfaces for real-world captures. And, with the increase in applications requiring 3D content, these models can be effectively used to quickly capture and create assets for graphic designers.

The popularity of these techniques have caused a boom in neural rendering techniques, allowing for the development of many different variants enabling generation or semantic querying including PERF (radiance fields from single panoramas) [22], LERFs (language embedded radiance fields) [13], and Instruct-NeRF2NeRF [11]. Models like PERF, which only require a single panorama to build radiance fields for scene reconstruction, further indicate the trend towards increasing ease-of-use and speeding up the process of generating 3D content. NeRFs originally required a well-crafted set of images from multiple views along with camera poses in order to build a proper representation, and these newer models are being used to relax those constraints with 3D generative foundation models.

There is a line of work to use natural language for understanding, semantically labeling, editing, and generating such models. LERF use CLIP embeddings [19] to allow users to query a NeRF using natural language to determine regions of interest. Users can search for any object in the scene and view a heatmap of the prompted item. ConceptGraphs [9] also uses a similar technique with CLIP embeddings but processes a more traditional 3D representation of point clouds rather than NeRFs. These developments have indicated the importance of object-centric labeling and editing in the systems that are built and have guided our design of editing components of 3D scenes in VR. In addition to labeling objects in captured scenes, language models are increasingly being used to generate new content and edit existing content. Instruct-NeRF2NeRF allows users to input a text prompt and an existing NeRF and edits it to reflect that change mentioned in the prompt, outputting a new NeRF. It utilizes 2D text guided image diffusion models such as InstructPix2Pix [3]. We interface with Instruct-NeRF2NeRF to allow users to build and edit their 3D scenes and objects. DreamFusion [18] also uses natural language but allows users to build a NeRF or mesh from scratch using a text prompt and can be used for designing quick mockups of objects a graphic artist may not have in their library.

Overall, newer rendering and modeling techniques are allowing for quicker prototyping for designing scenes using realistic 3D assets generated from the real world. Natural language has also come to the forefront in the development of these tools when it comes to labeling objects, editing them, or generating new models. However, no VR interface exists that makes these tools easy to use together and allows users to modify and experience their creation in a headset.

2.2 Stable Diffusion (SD) Models and Further Work

Playing a pivotal role in the field of generative AI, and in particular image synthesis, stable diffusion has become an essential foundation for these applications. While initial approaches were developed by OpenAI and Stability AI, since then there have been a plethora of applications of this technology with modifications and improvements to make it even better or specialized for certain tasks. Beginning with efficiency improvements, High-Resolution Image Synthesis with Latent Diffusion Models [20] demonstrates the concept of latent diffusion models (LDMs), which intend to improve the resolution and quality of image generation while achieving these results at significantly lower computational costs. On a tangential note of expanding the rich features and capabilities of stable diffusion, ZestGuide [6] demonstrates the ability to constrain stable diffusion outputs to be localized to specific spatial regions of the images, providing powerful and precise spatial control over generation/modification.

Taking these research papers from theory into practice, Blockade Labs [14] has integrated stable diffusion's generative capabilities, and extended them to fully generating entire environments that can be traversed through. Similarly, WOMBO [2] has created the ability to further modify and control the output of diffusion models by integrating traditional prompt input with additional fixed stylization controls, ensuring that various different prompts can be generated in a similar fashion and thus in theory can appear to fit a common theme.

The generative AI and image synthesis space has completely taken off since the introduction of diffusion models into the research landscape, and has quickly trickled into real-world startups using this technology to create direct applications for consumers to create and explore new horizons. However, there is yet to be an application that puts together all these ideas and concepts into a VR application, allowing users to not only take advantage of all that stable diffusion has to offer, but to further visualize and interact with the outputs in a virtual environment.

2.3 Creation Systems in VR

The transition from 2D CAD to 3D modeling in Virtual Reality (VR) represents a significant evolution, addressing the spatial limitations of traditional design methods [23]. 3DM [5] laid the groundwork with a 3D modeling system operated via a 6-DoF mouse, offering a novel way to interact with digital objects in three-dimensional space. Building on this, ISAAC [16] introduced scene editing within Virtual Environments, which allowed for a more intuitive and immersive design process.

Further innovations were introduced by CaveCAD [17] a system for freeform virtual sculpting of organic shapes, enabling artists and designers to conceptualize and iterate on their creations in an intuitive manner that closely mirrors the physical sculpting process. As the field has progressed, commercial applications have also begun to explore the potential of VR in creative expression. Google's Tilt-Brush [4] revolutionized the artistic landscape by allowing creators to paint with virtual light and textures, extending the canvas beyond the limits of traditional media. Similarly, VR games like Dreams [8], Figmin XR [25], and Horizon Worlds [24] have provided valuable insights into user interaction models, offering a glimpse into how VR can facilitate complex design tasks while maintaining user-friendly interfaces. Han et al., [10] demonstrates the next steps in HCI design and interaction with virtual environments by increased accuracy and range of physical gesture recognition, lending itself to more natural and user-friendly interaction with the surrounding virtual environment.

VR CAD systems are not without challenges: limitations in precision and expressiveness have been noted [16], which could potentially be mitigated by emerging technologies like generative AI. This integration promises enhanced detail and creative control by interpreting user input and generating refined designs, suggesting a future where AI augments human creativity in VR, enabling more accurate and expressive 3D models. As VR and AI technologies converge, they are set to redefine the capabilities and efficiency of 3D modeling, making it an indispensable tool in various creative and design disciplines.

In conclusion, The evolution of VR CAD has transformed design, offering immersive, intuitive modeling. Despite its progress, challenges in precision and expressiveness remain, with generative AI poised to revolutionize this space by enhancing detail and streamlining creation, signaling a new era for design and creative arts.

3 SYSTEM DESCRIPTION AND IMPLEMENTATION

3.1 Unity interactions in the Stable Diffusion system

In the Stable Diffusion system in Unity, there are several interactions available:

1. Automatic Image Saving: The camera is programmed to save images automatically every 10 seconds. These images are stored in a designated folder.

2. 3D Model Creation: Users have the ability to create 3D models within the scene. They can select objects from a menu and place them accordingly.

3. Load rendered stable diffusion model: The textManager component is responsible for loading the rendered image. This image, generated using a stable diffusion model, is then applied as a texture to a panel. This allows for showcasing the result of the image-to-image generation process.

3.1.1 Automatic Image Saving. 1. Initialization: The ‘Start()‘ method triggers the image capture process, using ‘InvokeRepeating‘ to call ‘CaptureAndSaveImage‘ repeatedly every ‘captureInterval‘ seconds, starting immediately.

2. Capture and Save Process: First, checks and creates ‘folderPath‘ if it doesn't exist. Second, captures the scene with ‘SDCamera‘ using a ‘RenderTexture‘ and transfers it to a ‘Texture2D‘ object. Third, encodes the ‘Texture2D‘ image to PNG and saves it in ‘folderPath‘ as "unity_scene.png". Finally, automatic Repeating: Repeats this process at the set interval.

3. Cleanup: Destroys the temporary ‘RenderTexture‘ and ‘Texture2D‘ after each capture to free memory.

3.1.2 3D Model Creation. Users can easily select the desired model by clicking on the button in the menu, as shown in figure.2. The system will accurately detect the position of the headset and generate objects that are positioned 1 meter away from the headset.

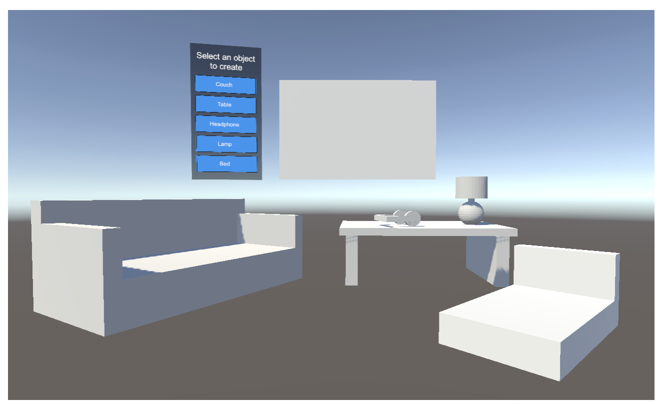

Fig. 2. Users can select object to create from the menu

3.1.3 Load rendered stable diffusion model. The TextureUpdateManager script in Unity is designed for continuously updating a GameObject's texture. It uses a coroutine, UpdateTextureRoutine, which runs indefinitely. Every updateInterval seconds, this coroutine attempts to load a new texture from a specified imagePath. If successful, the texture is applied to the GameObject's renderer. The texture is loaded using the LoadTexture method, which reads image data from the file and creates a Texture2D object. This script effectively allows for dynamic texture changes on a GameObject, using images periodically loaded from a given file path.

3.2 Spatial Annotation System

We propose a spatial annotation system for tracking edits across 3D objects in the scene, and facilitating real-time interactions and feedback despite non-real-time processes through generating proxy image or model previews of NeRF edits or generations. The spatial annotations or tags encode relevant information to allow user-specified edits to be applied offline. We implement this by recording the instructions that the user performs (editing NeRF objects or generating new 3D objects) along with a given text prompt which the user provides as speech input (e.g. "make the sofa blue" or "place a table"), as well as recording object specific information (location, name, etc.). Users can receive real-time feedback from their instructions as a text prompt hovering over the object as well as a 2D preview/proxy visualization of their proposed edit. For instance, this would be showing a 2D image plane of table created with text2image when the user asks to generate a 3D table) or an image of the edited NeRF using Instruct-Pix2Pix, the image conditioned diffusion model to perform 3D edits. We record this information as a JSON file which is parsed offline after all instructions have been noted. Our system then runs these instructions in a modular method by making API calls to other systems such as Nerfstudio to input the given object to edit and receive an output mesh or ply file. After these instructions are performed, our system places these objects into the Unity scene to create the final edited scene.

3.3 NeRF Editing System

One of the systems in Dreamcrafter is the ability to edit scenes with novel 3D representations such as NeRFs and 3DGS using our spatial annotation system. Leveraging generative methods like Instruct-NeRF2NeRF or Instruct-GS2GS (edit NeRFs/GS with text) and DreamFusion or DreamGaussian (text to 3D model/GS), our system performs semantic edits on NeRFs and generates 3D objects offline and reinserts them into the scene from the given scene spatial annotations. Users can import their captured NeRFs or Gaussian Splats into our system and place these objects within the NeRF environment (indoor or outdoor scenes) using the VR controllers to position and rotate them in the scene. To perform object edits, the user can simply point and press the primary button to initiate a spatial annotation, and then speak the instruction text-prompt (e.g. "make the chair retro style"). The user can also generate new 3D objects by pointing at the ground and speaking the text-prompt ("place a desk"). Dreamcrafter's spatial annotation system records these instructions and places the edited objects back into the scene in their corresponding locations. Figure 4 shows the spatial annotations placed over the NeRF objects in the scene and the final scene with edited NeRF objects placed in.

3.4 Stable Diffusion Creation System

The goal of stable diffusion and ControlNet part of the system is to be able to use barebone primitive 3D objects like cubes and cylinders to layout a scene and build a "blueprint" for a stylistic 2D render. For example, cubes and rectangular prisms can be stacked and arranged to form objects like tables, chairs, couches, and beds. Once the user has set up their scene using these primitives, they can place a virtual camera and take snapshots of the scene. With a text prompt that can be entered or dictated, the user has the ability to provide details on how they want to style the scene beyond the barebones primitives. They will then receive a 2D rendering of the scene on their headset to review, and it will be stylistically rendered according to their prompt. Figure 3 shows the virtual camera and a ControlNet render visible from the VR user's perspective.

In order to implement this, we followed the architecture detailed in the third component row of Figure 1. This component relies on the Unity interaction system, similar to the NeRF components. Users interact with 3D primitive objects and the virtual camera by moving them around. Next, every ten seconds, the virtual Unity camera takes a picture of the scene and saves it to the local file directory along with instructions and prompts in a JSON file. This is visualized as the green box in the figure. A Python script runs continuously in the background scanning for updates to the saved image in the directory and instructions. This online process is shown as the yellow box in the system diagram. It runs a MiDaS depth estimator and then runs ControlNet with the text prompt. ControlNet uses the depth estimation of the scene to ground its stable diffusion 2D image generation to be consistent with the user-defined 3D shape and configuration of the scene. This way, the rendered image stays grounded with the user's intent. The rendered image is then written to a predefined location on disk (indicated in the red box), and a texture in the Unity scene is updated to visualize the results in the headset. These renders are also saved for offline viewing.

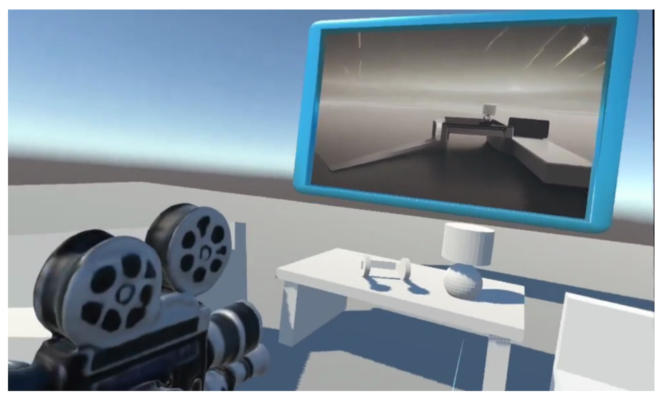

Fig. 3. Stable Diffusion System showing the controllable virtual camera and ControlNet output result.

4 EVALUATION

4.0.1 Methodology. We plan to conduct an evaluation to assess if Dreamcrafter met its design objectives of enhancing efficiency, intuitiveness, and immersion in 3D content creation, compared to conventional software like Blender, Maya, and 3ds Max. Experienced 3D designers and novices would be tasked to create similar 3D environments using Dreamcrafter and these traditional tools, aiming to test the hypothesis that Dreamcrafter's VR and AI integration would streamline the design process and does it provide an effective interface for using generative AI models.

4.0.2 Expected Findings. We hope that Dreamcrafter users could complete their tasks more quickly, highlighting its efficiency. This would be attributed to the intuitive VR interface and AI-driven real-time generation features which simplified design tasks. Novices, would be able to create more detailed and complex environments with Dreamcrafter, showing the AI's role in aiding those with limited experience. We hope that user study feedback would indicate a strong preference for Dreamcrafter's immersive experience and the ease of idea translation into designs, a task found challenging with traditional software. We also designed our spatial annotation system to be modular allowing new and better generative models for NeRF editing or 3D object generation to be easily replaced over time.

4.0.3 NeRF System Results. Our demonstrations using Dreamcrafter's NeRF system included designing living room interiors and outdoor scenes. The NeRF system allowed for detailed manipulation of spatial elements and object modifications through text prompts. The results, especially in the living room design, showcased high photo-realism and complexity, demonstrating the system's robustness in various environments.

Fig. 4. (Top) Spatial annotation tags are placed over the NeRF objects with given instructions and preview edits. (Bottom) The edited NeRF objects are placed in the scene after running the AI models offline.

4.0.4 Stable Diffusion System Results. Using Dreamcrafter's Stable Diffusion system, we transformed basic arrangements of primitive objects into detailed scenes, guided by user-provided text prompts. This demonstrated the system's capability to produce stylistically diverse environments from the same layout, highlighting the flexibility and power of generative AI in creative design.

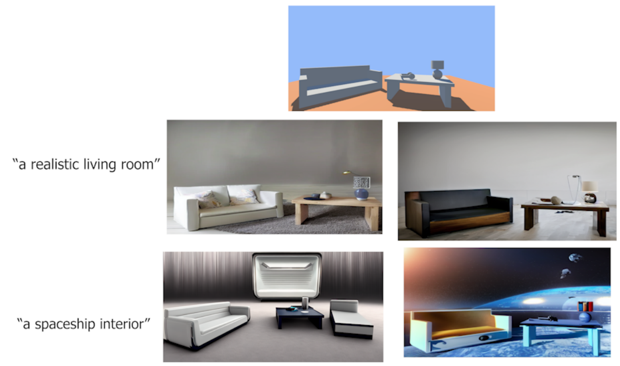

Fig. 5. Results from the Stable Diffusion system

4.0.5 Impact on Modular Design and HCI. We hope that a study will underscore the effectiveness of Dreamcrafter's modular design in building 3D content and scenes. Adaptable to future AI and VR innovations, it provides a flexible and scalable framework. The integration of these technologies not only enriches user experience but also signifies a notable progession in HCI systems, particularly in 3D design. The evaluation and results confirm that Dreamcrafter successfully achieves its goals of offering a more efficient, intuitive, and immersive approach to 3D environment design compared to traditional methods. Its innovative use of VR as an interface, combined with seamless generative AI, opens new avenues in 3D content generation and editing, posing relevant questions to the field of human-computer interaction.

5 DISCUSSION / LIMITATIONS

Although DreamCrafter offers a initial solution to creating generative environments in VR, we recognize some of its limitations. From the aspect of NeRF scene editing, one assumption is that our environment currently does not have any other existing editable objects within it. Essentially, this means that if a user were to create a NeRF of their furnished living room, we are currently unable to identify and auto-segment existing objects such as the furniture (couch, bed, etc.). Thus, this means that those real-world objects cannot be modified as well, and subsequently are unable to be edited through our existing pipeline.

In regards to the aspect of our Stable Diffusion system, a current limitation of DreamCrafter is that it only outputs 2D image renders. As a result, this means that we are only able to stylize and generate new environments from specific fixed perspectives and viewpoints of a scene, and subsequently can only be viewed from that angle while the scene inherently remains unchanged. This limitation means that we are not able to directly modify the underlying 3D environment itself, which limits the user to not being able to view the stylized environment from any perspective they wish.

Beyond the core aspects of NeRF and Stable Diffusion, there exist a few other limitations in more fine-grained aspects as well. For example, there is currently a limited assortment of primitive objects for the user to choose from to place into the scene. The ideal goal would be for the user to manually "create" these objects such as a dining table or a bedframe within the VR environment, but these interactions are a bit too difficult and time-consuming to accomplish given the current system. In addition, all our queries and requests to Instruct-NeRF2NeRF run offline, meaning that the user unfortunately cannot see these changes happen in live time. Furthermore, while our ControlNet renders of the scene for Stable Diffusion are indeed running online, there is inherently a slight latency within the render generation which may make the user lose a bit of the "real-time editing" feeling that we tried our best to accomplish.

Overall, while our current DreamCrafter system accomplishes many of the core problems we explored and provides diverse functionality for the user to generate and edit 3D environments in real-time, our discussion above aims to highlight some of the potential limitations of our system as well. Nevertheless, given more time we would be sure to address these concerns in future work.

6 CONCLUSION AND FUTURE WORK

In this paper, we introduced Dreamcrafter, an innovative VR system that leverages generative AI for 3D content generation and editing. Through our system, users can interact with and manipulate 3D objects and environments in real-time, with a focus on enhancing the 3D modeling process through intuitive VR interfaces and advanced AI technologies. Dreamcrafter's integration of Neural Radiance Fields (NeRF) and Stable Diffusion (SD) models marks a significant advancement in the field of virtual design and content creation.

However, as with any pioneering technology, Dreamcrafter is not without areas for further development and expansion. One key area of future work is the enhancement of our Stable Diffusion system to output 3D models directly. Currently, Dreamcrafter's SD system produces 2D images, which limits the user's ability to explore and interact with generated environments from multiple perspectives. Developing a method to convert these 2D images into 3D models would greatly enhance the versatility and immersive experience of our system.

Another vital area for development is the automatic segmentation of objects in NeRF-generated environments. This would allow users to individually edit and modify objects within a pre-existing 3D scene, significantly increasing the flexibility and utility of our NeRF editing system. By automating the segmentation process, we can streamline the editing workflow, making it more accessible and efficient for users. Moreover, an intriguing direction for future work is the integration of our Stable Diffusion and NeRF systems. One potential approach could be to utilize the SD-generated images at various camera angles to create 3D outputs in the form of NeRFs. This integration would not only enhance the realism and detail of the generated environments but also expand the creative possibilities available to users.

To further enhance user interaction, a key area of focus will be the development of more intuitive ways to place and interact with NeRF objects. For instance, incorporating voice commands like "place the table next to the blue chair" would make the system more user-friendly and accessible, particularly for those without extensive experience in 3D modeling. It's also worth noting that the modularity of our content generation and spatial annotation system is a cornerstone of Dreamcrafter's design. This approach not only facilitates future enhancements and integrations of different AI technologies but also serves as a robust framework for the development of AI-assisted interfaces in various applications.

In conclusion, Dreamcrafter represents a significant step forward in the realm of AI-assisted 3D content creation and editing. By continuing to develop and integrate advanced AI technologies, we aim to further democratize the process of 3D design, making it more accessible, intuitive, and powerful for a wide range of users. The future of Dreamcrafter holds the promise of even more seamless integration of AI and VR technologies, pushing the boundaries of what's possible in immersive design and content creation.

REFERENCES

[1] Luma Labs AI. 2023. lumalabs.ai.
[2] WOMBO AI. 2023. https://www.w.ai/.
[3] Tim Brooks, Aleksander Holynski, and Alexei A. Efros. 2023. InstructPix2Pix: Learning to Follow Image Editing Instructions. In *CVPR*.
[4] Tilt Brush. 2016. https://www.tiltbrush.com/.
[5] Jeffrey A. Butterworth, Andrew Davidson, Stephen Hench, and Marc Olano. 1992. 3DM: a three dimensional modeler using a head-mounted display. In *ACM Symposium on Interactive 3D Graphics and Games*. https://api.semanticscholar.org/CorpusID:9197179
[6] Guillaume Couairon, Marlène Careil, Matthieu Cord, Stéphane Lathuilière, and Jakob Verbeek. 2023. Zero-shot spatial layout conditioning for text-to-image diffusion models. arXiv:2306.13754 [cs.CV]
[7] Nianchen Deng, Zhenyi He, Jiannan Ye, Budmonde Duinkharjav, Praneeth Chakravarthula, Xubo Yang, and Qi Sun. 2022. FoV-NeRF: Foveated Neural Radiance Fields for Virtual Reality. arXiv:2103.16365 [cs.GR]
[8] Dreams. 2020. https://www.playstation.com/en-us/games/dreams/.
[9] Qiao Gu, Alihusein Kuwajerwala, Sacha Morin, Krishna Murthy Jatavallabhula, Bipasha Sen, Aditya Agarwal, Corban Rivera, William Paul, Kirsty Ellis, Rama Chellappa, Chuang Gan, Celso Miguel de Melo, Joshua B. Tenenbaum, Antonio Torralba, Florian Shkurti, and Liam Paull. 2023. ConceptGraphs: Open-Vocabulary 3D Scene Graphs for Perception and Planning. *arXiv* (2023).
[10] Sujuan Han, Shuo Liu, and Lili Ren. 2023. Application of human-computer interaction virtual reality technology in urban cultural creative design. *Sci. Rep.* 13, 1 (Sept. 2023), 14352.
[11] Ayaan Haque, Matthew Tancik, Alexei Efros, Aleksander Holynski, and Angjoo Kanazawa. 2023. Instruct-NeRF2NeRF: Editing 3D Scenes with Instructions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*.
[12] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics* 42, 4 (July 2023). https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/
[13] Justin Kerr, Chung Min Kim, Ken Goldberg, Angjoo Kanazawa, and Matthew Tancik. 2023. LERF: Language Embedded Radiance Fields. In *International Conference on Computer Vision (ICCV)*.
[14] Blockade Labs. 2023. https://www.blockadelabs.com/.
[15] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *ECCV*.
[16] Mark Mine. 1995. ISAAC : A Virtual Environment Tool for the Interactive Construction of Virtual Worlds. (06 1995).
[17] Kevin Ponto, Ross Tredinnick, Aaron Bartholomew, Carrie Roy, Dan Szafir, Daniel Greenheck, and Joe Kohlmann. 2013. SculptUp: A rapid, immersive 3D modeling environment. In *2013 IEEE Symposium on 3D User Interfaces (3DUI)*. 199–200. https://doi.org/10.1109/3DUI.2013.6550247
[18] Ben Poole, Ajay Jain, Jonathan T. Barron, and Ben Mildenhall. 2022. DreamFusion: Text-to-3D using 2D Diffusion. *arXiv* (2022).
[19] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning Transferable Visual Models From Natural Language Supervision. arXiv:2103.00020 [cs.CV]
[20] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis with Latent Diffusion Models. arXiv:2112.10752 [cs.CV]

[21] Matthew Tancik, Ethan Weber, Evonne Ng, Ruilong Li, Brent Yi, Terrance Wang, Alexander Kristoffersen, Jake Austin, Kamyar Salahi, Abhik Ahuja, David Mcallister, Justin Kerr, and Angjoo Kanazawa. 2023. Nerfstudio: A Modular Framework for Neural Radiance Field Development. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Proceedings*. ACM. https://doi.org/10.1145/3588432.3591516

[22] Guangcong Wang, Peng Wang, Zhaoxi Chen, Wenping Wang, Chen Change Loy, and Ziwei Liu. 2023. PERF: Panoramic Neural Radiance Field from a Single Panorama. arXiv:2310.16831 [cs.CV]

[23] Josef Wolfartsberger. 2019. Analyzing the potential of Virtual Reality for engineering design review. *Automation in Construction* 104 (2019), 27–37. https://doi.org/10.1016/j.autcon.2019.03.018

[24] Meta Horizon Worlds. 2020. http://www.oculus.com/facebookhorizon.

[25] Figmin XR. 2022. https://overlaymr.com/.