Compos3D: Interactive Part-Based Composition for Creative Control in Generative 3D Models

Faraz Faruqi
faraz.faruqi@autodesk.com
Autodesk Research
Boston, MA, USA

Sean J. Liu
sean.liu@autodesk.com
Autodesk Research
San Francisco, California, USA

George Fitzmaurice
George.Fitzmaurice@autodesk.com
Autodesk Research
Toronto, Ontario, Canada

Justin Matejka
Justin.Matejka@autodesk.com
Autodesk Research
Toronto, Ontario, Canada

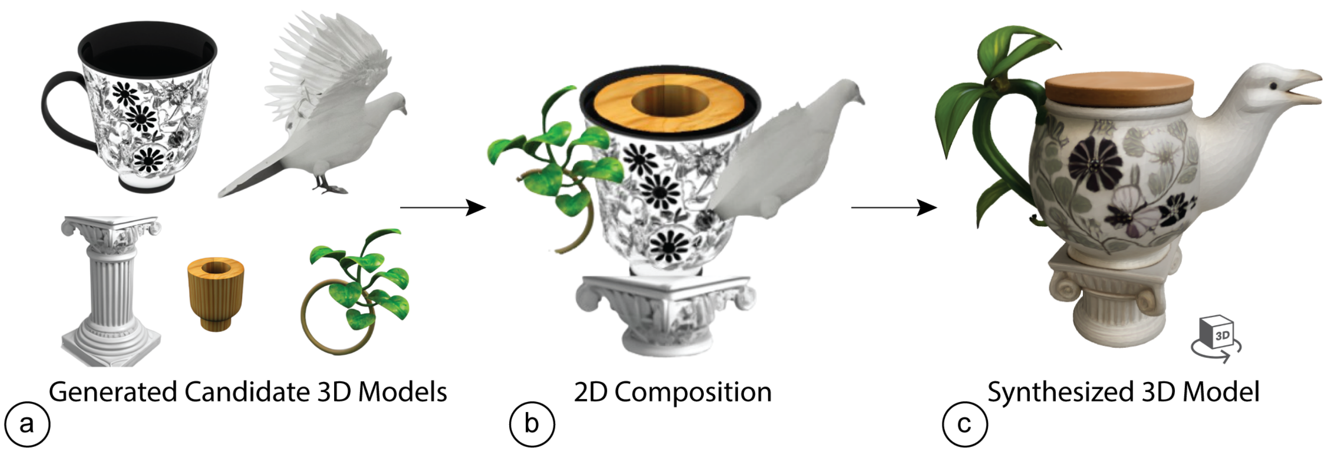

Figure 1: Compos3D allows users to create new 3D designs by remixing elements drawn from different generated outputs. (a) Users begin with candidate models generated (or uploaded). (b) Users select desired components (e.g., floral motif mug body, dove bird spout, vine handle, roman pedestal base, and bamboo lid) and arrange them into a rough composition. (c) The system uses 3D Generative AI to synthesize a coherent final model of a teapot that integrates the chosen parts into a unified design.

Abstract

While generative AI has unlocked new opportunities for 3D content creation, current workflows often rely on multiple regenerations, which provides limited control and unpredictable outcomes. We present Compos3D, a system that introduces a compositional workflow for generative 3D modeling through remixing. Instead of repeatedly regenerating models, users generate multiple candidates from text or image prompts, select parts of interest via 2D image regions or 3D mesh segments, and assemble them into a coherent design. The system synthesizes these compositions into a refined 3D model, preserving high-level intent while resolving low-level geometry. To evaluate this approach, we conducted a controlled user study comparing remixing and regeneration workflows across both 2D and 3D modalities. Results show that the remixing workflow provides participants with greater creative control, stronger alignment with their intent, and higher satisfaction. We conclude with design recommendations for future AI-assisted 3D modeling workflows.

Keywords

generative AI; 3d modeling; creativity-support tools

1 Introduction

The rapid growth in 3D Generative AI tools has enabled users to create increasingly complex 3D models from simple text or image prompts. This has opened up new opportunities for creative design, where novice creators can rapidly realize their design ideas with descriptive prompts without learning complex 3D modeling techniques. However, most 3D generation systems currently operate as one-shot black boxes: users describe an object, receive a single or multiple 3D models, but have little ability to iteratively manipulate its structure. When users need specific changes, they face a frustrating "*dice roll*" scenario—repeatedly regenerating a portion of or the entire model with modified prompts, hoping the system will preserve desired features while incorporating new ones that align with their intent. Each regeneration risks losing elements the user wanted to keep, turning the design process into a prompt-refinement task rather than an intentional creative task.

In contrast, human creative practice across domains, such as generative art [2], creative coding [43], or music design [16], relies heavily on recombining and recontextualizing parts. Designers often express and explore ideas through moodboards, especially when ideas are hard to express verbally [8]. Prior research has shown that compositional approaches allow users to engage users more deeply in creative exploration than prompt-based generation alone [6, 31, 34]. For example, in image generation, interactive collage and composition systems allow users to iteratively build and refine designs with immediate visual feedback, supporting richer ideation and reflection [25, 33, 34]. Similarly, in music generation, part-based remix interfaces have been shown to foster creative engagement by enabling users to directly manipulate, arrange, and recombine components [16, 38]. In current generative tools, studies investigating users' interaction [44, 56] underscore the limitations of text and image prompting alone. Moreover, research has emphasized the value of precise, fine-grained control, such as localized edits [11, 48] and direct manipulation interfaces [31], in generative tools. In this work, we bring these two dimensions together in 3D.

We introduce *Compos3D*, a system that supports *interactive remixing* of generative 3D models through both image-based and 3D segment-based interactions. We define remixing as the practice of taking elements from disparate designs and combining them into a refined model that preserves and integrates distinct features from multiple sources. Unlike existing workflows that rely on regeneration through text prompts and highlights, Compos3D allows users to select desired parts from multiple generated outputs and composite them into their final result. One advantage of this workflow is that it aligns more closely with human creative processes, as evidenced by numerous studies in other domains showing that remixing and composing outputs foster creativity [25, 33, 34]. Another advantage is that it gives users fine-grained control over the final 3D model, bringing elements of WYSIWYG[1]-style interaction into 3D generative modeling.

With Compos3D, users first generate multiple 3D candidates with text or image prompts and then select specific parts to compose with - such as the wings of a 3D dragon model and the body of another — using selected regions either from images or 3D mesh segments. Our system then synthesizes new models that fuse the selected components into a coherent whole, while retaining the form and semantic intent of the original parts. In this way, the system provides the user with local control of individual segments, while the AI model fills in the topological and textural details of the refined 3D model.

To evaluate our approach, we conducted a user study in which novice 3D modelers were asked to construct a 3D model following a multi-part design brief and describe their experience during the process. We compared our remixing tool against the regeneration approach, where users regenerate models using highlights and textual edits. This allowed us to investigate whether the remixing workflow provides any benefits over current regeneration approaches for creating 3D models. For selecting individual parts, the study compared both 2D (image-based) and 3D (mesh-segment-based) interactions, allowing us to investigate whether users prefer a specific modality for 3D model design. Results show that our remixing workflow gave users greater creative control and alignment with design intent, and demonstrated trade-offs between 2D and 3D input modalities. Finally, we utilize the results of our study to derive a set of design recommendations for future 3D modeling interfaces with Generative AI, discussing how compositional and multi-modal interfaces can better support precision, compositional reasoning, and creative control than regeneration alone.

In summary, our work contributes:

(1) A novel system for multi-modal remixing of generative 3D models using image and mesh-based part selection;
(2) A user study comparing the Remixing and Regeneration workflows
(3) A comparison of 2D and 3D remixing modalities in terms of user performance, control, and satisfaction;
(4) Design recommendations for future work on remixing workflows for 3D generation.

2 Related Work

We situate our work at the intersection of human-computer interaction, generative AI, and computational design. Specifically, we draw on prior research in three key areas: (1) advances in 3D generative AI, (2) current support tools for 3D modeling, and (3) remixing as a creativity-support paradigm.

2.1 3D Generative AI Workflows

Reconstructing 3D models from a single monocular image or a text description is a longstanding challenge in computer vision. Recent machine learning methods address this problem by learning priors over large collections of 3D shapes to infer the full geometry of an object from a single image. Several approaches have been proposed for this problem, including but not limited to GANs [7, 17], flow networks [24, 54], VAEs [32, 50], and diffusion models [10, 21]. The recent Large-Reconstruction Model (LRM) architecture [20]-based methods [5, 46, 52] have demonstrated generalizable and high-quality 3D reconstruction.

These generative models have enabled the development of support tools for creative exploration and rapid prototyping [41]. 3DALL-E [30] integrates text-to-image AI into CAD software to provide designers with image-based inspiration. VRCopilot [57] proposed a mixed-initiative VR authoring system that integrates generative AI with multimodal interactions and intermediate wireframe representations. Style2Fab [14] introduces a method to selectively personalize 3D models while preserving their affordances for physical use, while TactStyle [15] enabled the stylization of 3D models using image prompts while controlling their visual and tactile properties.

While these advances have lowered the barrier for users to create unique and high-quality 3D models, they remain largely prompt-driven. Users typically regenerate entire models to incorporate changes, with little ability to preserve desired features or combine elements across outputs. In contrast, image-generation research has shown that compositional interfaces such as Generative Photomontage [29], which use collages as a canvas and enable localized editing, give users more concrete control for creative exploration. Inspired by these approaches, we explore how similar compositional workflows can be brought into 3D generative modeling, enabling

[1]What You See Is What You Get

novices to remix parts across multiple outputs rather than rely solely on trial-and-error prompting.

2.2 3D Modeling Support Tools for Novices

3D modeling requires both design expertise and proficiency with complex CAD software, making it difficult for novices to translate intent into design.

HCI researchers have addressed these barriers with tools that scaffold the modeling process and lower the threshold for participation. For instance, Meshmixer [39] lets users combine parts extracted from a library of meshes, while PARTS [19] provides a method for designers to label their geometry with custom programs, making the models more reusable by novice makers. Alternatively, AutoConnect [26] allows users to connect two meshes without modifying them by creating a connector as a new mesh. Grafter [36] creates new mechanisms from existing parts by scaling the meshes to fit the mechanisms without affecting their functionality. Attribit [9] enables replacement using semantic attributes, and parametric design systems [40, 42, 47] provide interactive customization while preserving validity and manufacturability.

While these 3D modeling support tools enable novice users to personalize existing meshes or models, they are either limited to a static library of models, or only allow limited customization. Generative AI methods offer a way beyond static libraries and limited parametric customization; however, current 3D tools still rely on one-shot generation or regeneration, limiting their usability. Our work addresses this gap by integrating remixing into 3D generative workflows.

2.3 Human-in-the-Loop Customization with Generative Models

Across domains, HCI researchers have designed systems that situate generative models in human-in-the-loop workflows, giving users ways to steer, remix, and reuse outputs rather than rely on one-shot prompts. Previous studies on non-experts with AI-based models [44, 56] highlight why such interfaces are needed: novices struggle with underspecified prompts, opaque model reasoning, and lack of feedback. Interactive systems that afford multi-modal interaction and provide intuitive and controllable feedback mechanisms mitigate these challenges by scaffolding exploration, making intermediate steps visible, and enabling correction or recombination.

In creative coding, tools such as Spellburst [2], and XCreation [53] provide node- or graph-based interfaces to compose multimodal generative pipelines, enabling branching, recombination, and semantic programming. WorldSmith [13] and AngleKindling [35] let writers fork, merge, and refine prompts for narrative and journalistic content. In visual art and design, systems such as PromptPaint [11], Promptify [6], and PromptCharm [48] allow localized steering of image generation through brush strokes, prompt suggestions, attention-map reweighting, or direct manipulation. DirectGPT [31] generalizes this idea to text, code, and graphics, translating GUI actions into prompt edits that embody direct manipulation principles.

Other work has emphasized organizing and reusing outputs. HistoryPalette [4] and DreamSheets [1] capture and cluster generations, supporting revision and exploration of alternatives. Studies of example galleries [27, 55] highlight how curating and reusing design examples is integral to creative practice. Research on remixing communities [43] shows that branching, forking, and reuse are central to creative culture, aligning with systems that scaffold iterative exploration. In the LLM domain, systems like EvalLLM [22], AIChains [49], and Sensecape [45] allow breaking complex tasks into smaller steps, generate multiple results for different prompts, and evaluate them on user-defined criteria.

Together, this body of work shows that compositional interactive workflows are useful across nearly all domains of generative AI: from writing to art, to programming. They suggest that giving users explicit ways to preserve, recombine, and organize generative outputs fosters creativity, control, and ownership. We extend this line of work into 3D generative modeling — allowing users to combine parts across multiple outputs while retaining precise control over how they are integrated.

3 System Design and User Interface

Existing generative 3D tools often treat creation as a one-shot process: a prompt yields a complete model, which users can only adjust through global style edits or repeated regeneration. This limits fine-grained control and prevents designers from reusing desirable elements across different outputs. Compos3D addresses this gap by enabling part-based remixing: users can extract meaningful components from multiple generated or uploaded models and assemble them into new hybrid designs. Inspired by creative practices like collaging and kitbashing, our system treats intermediate generations as raw material for composition rather than discarded attempts.

To support this workflow, Compos3D provides a unified pipeline with three phases—Part Selection, Composition, and Synthesis—available in both 2D (image-based) and 3D (mesh-based) modalities, for creating 3D models. In Part Selection, users segment and select desired parts from candidate models; in Composition, they arrange these parts into rough collages on a 2D canvas or within a 3D scene; and in Synthesis, the system refines these collages into coherent, high-quality 3D models that preserve the user's design intent. By supporting both 2D and 3D modalities, Compos3D allows us to explore the tradeoff between the speed and flexibility in 2D, and the precision and spatial control in 3D.

In the next sections, we describe each step in more detail.

3.1 Initial Model Generation and Remix Context

Users begin with generating initial candidate 3D models using either text or image prompts (Fig.2a). Given a prompt such as "office chair", Compos3D produces a diverse set of images with a text-to-image model. Users curate this set by selecting images of interest into a carousel, which our system then converts into candidate 3D models with an image-to-3D model [51] (Fig. 2b). The resulting 3D models are rendered in the interface for the user's inspection. The user can open them in a larger viewer and view the geometry from multiple angles. Then, users can select models to remix by adding

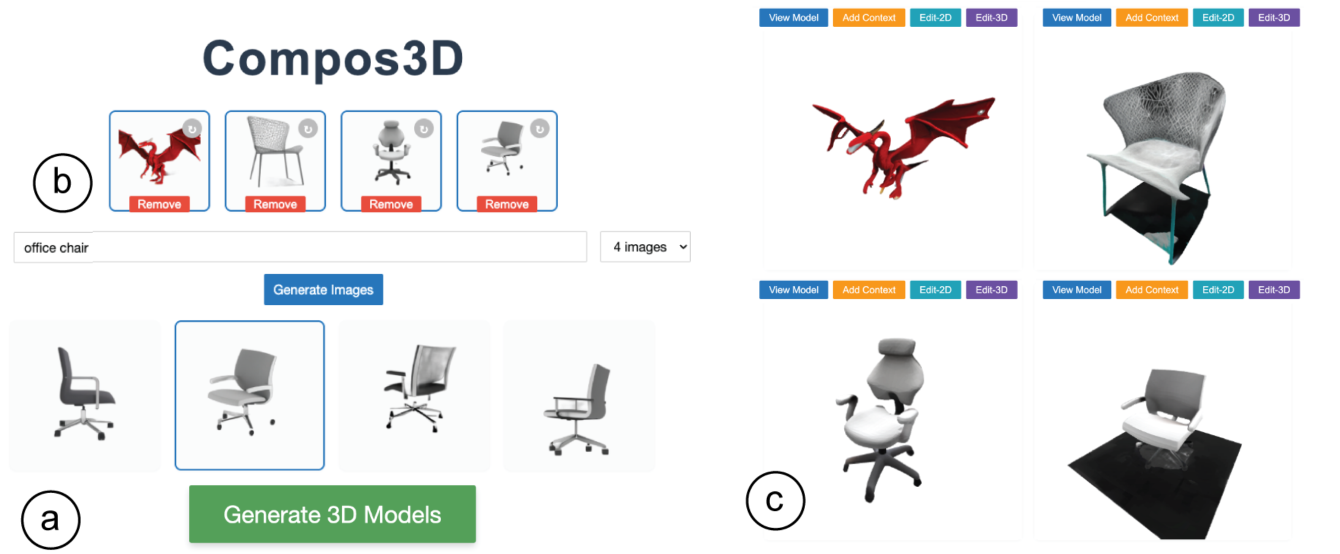

Figure 2: Compos3D 3D Model Generation Interface. (a) Users begin by uploading candidate images or generating them using a text-to-image model. (b) The user selects images of interest to add to the collage. (c) Once satisfied with the set of images. the user clicks 'Generate 3D models', and the system generates 3D models based on the images using an image-to-3D generative AI model. Upon inspecting the results, users select models to remix by adding them to a working canvas with the 'Add to Context' Button.

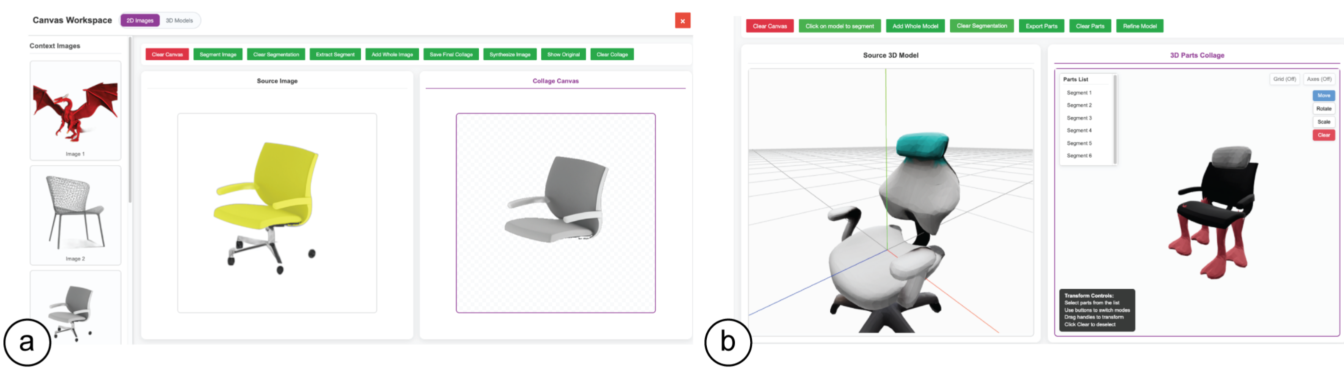

Figure 3: Compose3D Remixing User interface. The user's selected models are added to the 'Remix Context.' Here, users can source parts for composition. (a) In the 2D workspace, users select and extract regions from source images using click-based interactive segmentation and arrange them on a collage canvas. (b) In the 3D workspace, users select segments from the 3D model using click-based interactive segmentation and add them to a 3D parts collage. After the user is satisfied with the composite, the collage is synthesized into a 3D model.

promising designs into a Remix Context. This curated pool of 3D models serves as the working set from which parts can later be extracted and recombined.

In addition to generated outputs, users may also import their own 2D or 3D assets, allowing remixing with elements drawn from existing libraries or real-world artifacts.

3.2 Part Selection

Once the user is satisfied with the models added to the 'Remix Context', they move to the Remixing page. Here, users segment and extract 3D parts for composition. Compos3D supports interactive segmentation via 2D images or 3D meshes, enabling fine-grained selection of regions to be carried forward into composition.

In both modalities, users isolate meaningful regions for remixing (e.g., handles, surfaces, decorative elements) through an interactive prompt-based segmentation workflow. Selection operates via positive (`left-click`) and negative (`right-click`) clicks, allowing users to iteratively refine highlighted regions with real-time feedback until satisfied.

Figure 4: Interactive Part Selection in Compos3D. Users can segment regions either from (a) 2D images or from (b) 3D meshes. Both modalities support positive (green) and negative (red) prompts to iteratively refine selections until only the desired part is highlighted. Once finalized, selected regions are added to the remixing context for composition.

In the 2D interface, users operate on individual images. We employ the Segment Anything Model (SAM) [23], which supports fine-grained interactive segmentation. Users load an image, apply positive and negative clicks to adjust the highlighted region, view feedback in real time (Fig. 4a), and finalize the segment before adding it to the collage (Fig. 3a). This process can be repeated across multiple images to collect parts for remixing.

In the 3D interface, users interact directly with meshes in a 3D viewer. We employ Point-SAM [58], which extends the positive/negative click paradigm to 3D models. Highlighted regions are rendered directly on the mesh surface in real time (Fig. 4b). Once refined, the selected region is extracted as a mesh and added to the collage for later composition.

3.3 Composition

Once parts are selected and added, the system provides a canvas for composition, where users can explore different design possibilities by combining elements into new forms. This stage functions similarly to a canvas in creative practice, where designers assemble segments to explore ideas, compare alternatives, and evaluate new directions. Using the freeform canvas, users can visually experiment with the structure and style of their composition before finalizing the model in the synthesis stage.

In the 2D canvas, extracted image segments are placed as independent movable elements. Users can scale, rotate, and overlap them to explore alternative collages that represent their design intent. If they find a certain element that is missing, they can add additional parts from other images.

In the 3D canvas, the previously added segments appear as independent 3D meshes with the original material properties. Users can rotate, translate, and scale these parts individually in the 3D space to semantically align them according to their design intent. Unlike the flat compositional space in 2D, the 3D canvas affords precise control over spatial arrangement, allowing users to resolve perspective and proportion as desired.

3.4 Synthesis

In this final stage, users transform their rough compositions into coherent designs. The freeform compositing in the 'Composition' phase allows users to experiment and clarify their design intent, producing a rough approximation of the intended structure and style. But since these segments are eclectically sourced from different pre-generated models and positioned manually by the user, these rough compositions may lack the coherent topological details required in a final mesh. The synthesis stage resolves this gap, connecting the 3D parts together to form watertight geometry and preserve surface detail.

For 2D synthesis, Compos3D processes the rough collage created by the user into a refined composite image, using an image-generation model. Optionally, the user can add a text prompt to provide additional detail (Fig. 6a). The system treats the collage as a structural guide, producing coherent images that retain the composition but add detail and consistency. If the user wishes to further update the image, they can optionally regenerate alternatives until the result reflects their design intent. The finalized image is then converted into a 3D model through an image-to-3D pipeline.

For 3D synthesis, Compos3D first voxelizes the rough collage assembled by the user, producing a structured volumetric scaffold that encodes the overall form and proportions of the design. In parallel, the system renders multiple views of the collage and refines

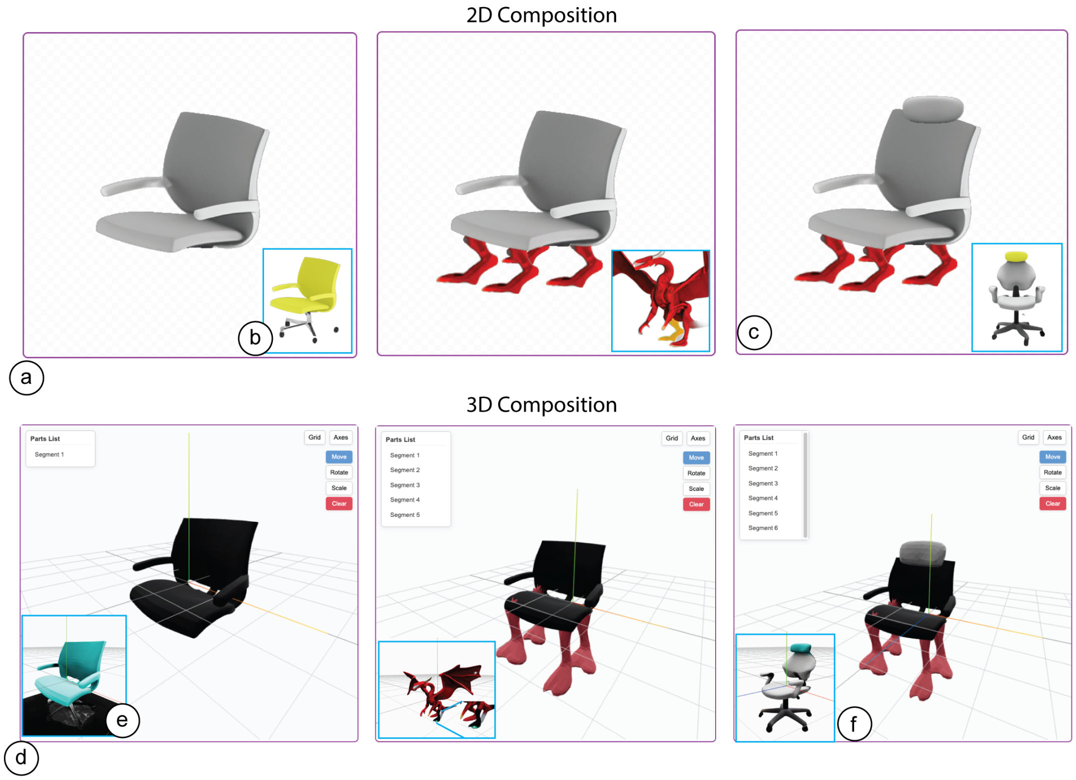

Figure 5: Composition in Compos3D. (a–c) 2D composition: users select regions from source models (examples shown in inset (b)) and place them on a 2D canvas, where they can scale, rotate, and overlap elements to build collages. (d–f) 3D composition: users extract parts from source models (inset (e)) and manipulate them as meshes in 3D space using translation, rotation, and scaling tools for precise alignment. In both modalities, users can revisit different source models and iteratively add parts until satisfied with their composition before moving to synthesis.

them with an image-to-image model (optionally guided by user text prompts), ensuring that the synthesized appearance reflects the intended structure and semantic details. These multi-view images, together with the voxelized geometry, are then passed into the TRELLIS pipeline [51], which operates on a Structured Latent (SLAT) representation to jointly generate geometry and appearance. By fixing the voxel structure, and combining with dense multi-view features, Compos3D uses TRELLIS to generate a coherent 3D mesh that preserves the high-level composition specified by the user while resolving seams, topology, and surface detail. The refined 3D model is then presented back to the user.

3.5 Summary

Across both 2D and 3D, Compos3D implements a unified workflow of selection, composition, and synthesis. In each case, users first isolate parts of interest, then arrange them into a rough collage, and finally refine the assembly into a coherent 3D model. While the underlying process remains consistent, the modalities highlight different affordances: 2D offers speed and flexibility for rapid collage-making, whereas 3D affords precise spatial control and proportion. This design enables users to choose the modality that best fits their goals and aligns with their creative process. It also provides a basis for our user study, allowing us to directly compare how 2D and 3D workflows shape creative outcomes and user experience.

3.6 Technical Implementation

Compos3D integrates multiple generative AI models (Segment Anything Model (SAM) [23], Point-SAM [58], TRELLIS [51], OpenAI's `gpt-image-1`) into a unified workflow for segmentation, composition, and synthesis. Given the rapid progress in generative AI research, we designed our system in a modular fashion to allow easy replacement of any component with an improved version. For segmentation, we employ the Segment Anything Model (SAM) [23] for 2D images and Point-SAM [58] for 3D meshes. Both support positive/negative click interactions, allowing users to iteratively refine highlighted regions with real-time feedback.

For 2D composition, collages assembled by the user are refined with OpenAI's `gpt-image-1` model, which refines the image. The refined image is then converted into a 3D model using an image-to-3D pipeline based on TRELLIS [51]. This pipeline takes the image as input, predicts consistent multi-view renderings, and reconstructs a structured 3D mesh from them.

For 3D composition, mesh segments arranged in the 3D workspace are first voxelized into a latent volumetric scaffold that encodes their

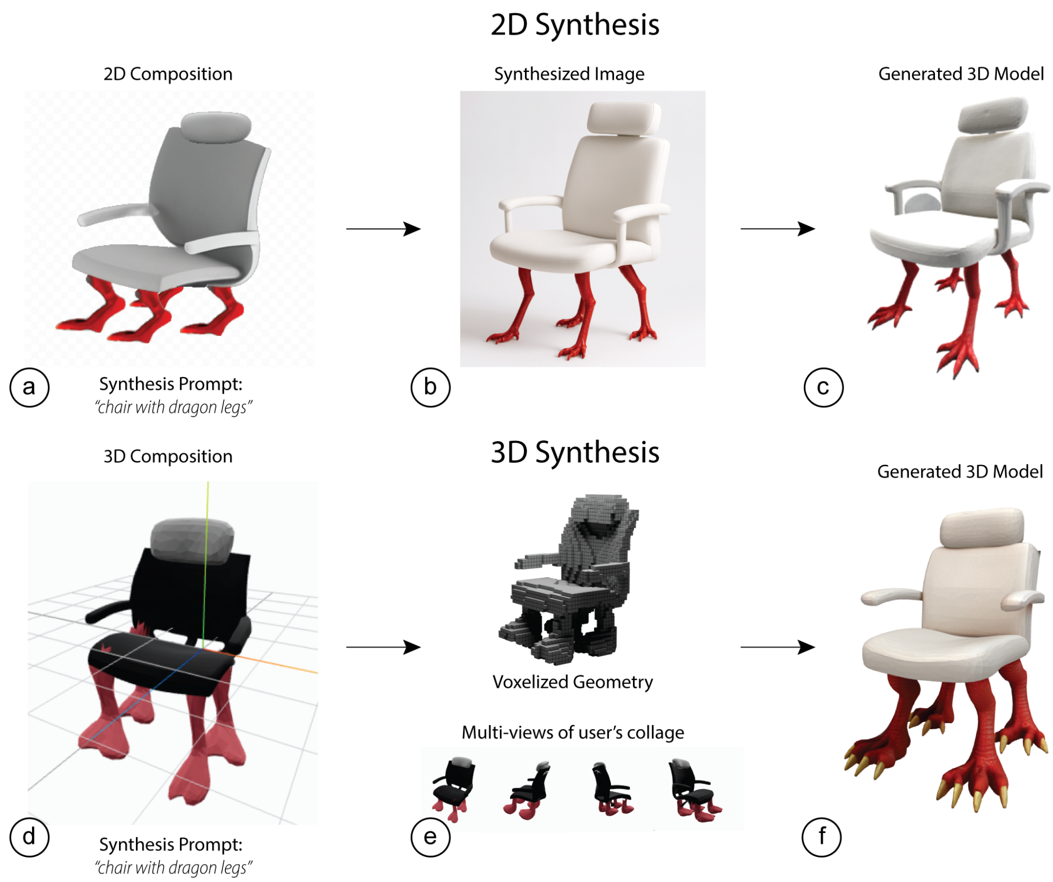

Figure 6: Synthesis with Compos3D: (Top row) 2D synthesis: user creates a composite image, which is refined by an image-generation model into a coherent design; if the user is satisfied with the image, it is converted into a 3D model with an image-to-3D model. (Bottom row) 3D synthesis: users arrange mesh segments into a rough composition, which is voxelized and used as latent volume by a generative model guided by multi-view images of the collage and an optional textual prompt. The resulting mesh resolves seams and topology while retaining semantic intent, producing a coherent 3D model. Both workflows allow users to transform rough collages into finalized, high-quality outputs.

global structure and proportions. In parallel, the system renders multiple 2D views of the rough assembly, which are refined with image-to-image generation optionally guided by user text prompts, if provided. These multi-view images, together with the voxel scaffold, are passed into the TRELLIS pipeline, which operates on a structured latent (SLAT) representation. This representation fuses volumetric and multi-view features, ensuring watertight topology while preserving surface detail and semantic coherence. On an NVIDIA L4 GPU, end-to-end synthesis runs in an average 42.3 seconds (std dev=7.2 secs) per model, enabling interactive generation.

4 User Study

Compos3D's key contribution lies in enabling part-based remixing for generative 3D modeling, where users create new hybrid models from multiple generated models rather than repeated regeneration, a trial-and-error process. Our approach aims to provide users with precise control over the final output, while leveraging the expressiveness and diversity of 3D generation models as a tool for proposing candidates.

To compare our proposed workflow with the current regeneration approach, we conducted a controlled user study comparing Compos3D's remixing workflow with a baseline regeneration workflow. Our goal was to assess whether remixing provides users with greater creative control, expressiveness, and satisfaction when constructing 3D models. We also explore how different interaction modalities—2D image-based and 3D mesh-based—shape the design process. We centered the study around the following research questions:

(1) **RQ1 (Workflow Preference):** Does remixing provide better support creative control, expressiveness, and alignment with user intent compared to regeneration?
(2) **RQ2 (Modality Preference):** How do 2D and 3D modalities differ in terms of usability, effort, and perceived effectiveness?

(3) **RQ3 (Trade-offs):** What trade-offs do users perceive across workflows and modalities in terms of effort, control, and satisfaction?

The study employed a 2×2 within-subjects design, crossing Tool (Remixing vs. Regeneration) with Modality (2D image- vs. 3D segment-based interaction). In the following subsections, we describe the study design and baseline workflow, participant demographics, and the tasks and procedure used in the evaluation.

4.1 Study Design & Baseline Workflow

The study followed a within-subjects 2×2 factorial design with two independent variables: **Tool** (Remixing vs. Regeneration) and **Modality** (2D image-based vs. 3D segment-based interaction). Each participant therefore completed four conditions: Remixing–2D, Remixing–3D, Regeneration–2D, and Regeneration–3D, in the study that ran for 2 hours. Task order was counterbalanced using a Latin Square. With eight participants, we completed two full rotations of the Latin Square, ensuring that each condition appeared equally often in each order position.

Baseline: Regeneration Workflow. As a baseline, we implemented a *regeneration workflow* in both 2D and 3D, mirroring the interaction paradigm of existing generative tools (Fig 7). To ensure that participants interacted with a consistent interface and underlying generation pipeline, we implemented this workflow within the Compos3D framework. In this workflow, users first generated the image or 3D model as described in Section 3.1. They then select a local region of interest either by drawing a bounding box on a 2D image or by highlighting a surface region on a 3D mesh, depending on the modality. Then, they provide a textual instruction describing the desired change (e.g., "replace legs with dragon legs" or "add a tray console"). The system regenerated the selected region while preserving the rest of the model, allowing users to progressively refine their design across multiple iterations until satisfied.

4.2 Participants

We recruited eight participants (4 male, 4 female), aged 21–39 years (M = 27.0, SD = 5.76), through snowball sampling and institute mailing lists. All participants were beginners with 3D modeling tools. Participants' familiarity with AI generative tools (e.g., ChatGPT, DALL-E) was moderate (M = 4.38, SD = 1.60, on a 7-point scale), ranging from beginner (1) to expert (7). The entire user study lasted for 2 hours, and the participants were provided with a 120 USD gift card for their time.

4.3 Tasks and Procedure

Each participant completed four design tasks, one in each condition of the 2×2 design (Remixing–2D, Remixing–3D, Regeneration–2D, Regeneration–3D). To allow direct within-subject comparisons while holding task requirements constant, all tasks were based on the *same* design brief:

> **Design a hybrid work chair.**
> *Backrest:* Tall, arched frame.
> *Armrests:* Jointed supports as branches, with one armrest featuring a swing-out tray console and control screen.
> *Base:* Rotating foundation with a footrest platform in the front.
> *Headrest:* Adjustable head support with lighting.
> *Seat Cushion:* Lattice structure with a soft material.

A design brief provided a controlled task, where functional and aesthetic requirements are held constant across conditions, while allowing for creative interpretation and freedom. Design briefs also mirror real-world design practice, where requirements containing both function and style are specified as written descriptions [12, 18].

Procedure. At the start of each session, participants completed a background questionnaire on demographics, prior 3D modeling experience, familiarity with generative AI tools, and exposure to remixing workflows. They then received a short tutorial and completed a warm-up design task using a simplified 3-part brief (i.e., designing a mug) to familiarize themselves with the interface.

Each participant then completed the four experimental conditions in counterbalanced order using a Latin Square. In each condition, they (1) reviewed the design brief, (2) completed the modeling task with the assigned tool and modality, and (3) answered a post-task questionnaire rating task outcome, perceived control, cognitive demand/effort, enjoyment, and tool responsiveness, with optional open-ended feedback.

After completing all tasks, participants filled out a post-study questionnaire on overall preferences (tool and modality), perceived control, and confidence in meeting the design brief. Finally, we conducted a semi-structured interview to capture deeper reflections on creative strategies, sense of control, and envisioned roles for generative AI in future design practice.

5 Results

To compare the workflows, we measured both quantitative and qualitative outcomes. Likert-scale ratings from post-task questionnaires were aggregated by participant and condition, and we report means, standard deviations, and 95% confidence intervals. Open-ended responses from questionnaires, reflections, and interviews were thematically coded to capture recurring strategies, perceived strengths and limitations, and participants' reflections on creative control and expressiveness. Results are organized by research question.

5.1 RQ1: Preference - Remixing vs. Regeneration

To compare the two workflows, we analyzed participants' responses to the post-task questionnaire across three dimensions: *Task Outcome* (Q1–Q4), *Cognitive Demand & Effort* (Q5–Q7), and *Tool Responsiveness* (Q8–Q10). Figure 8 summarizes these quantitative results.

Task Outcome. Participants rated Remixing consistently higher than Regeneration across all outcome measures. They more strongly agreed that they were able to create the design they had in mind (Q1: Remix $M = 5.81$, $SD = 1.05$ vs. Regeneration $M = 4.44$, $SD = 0.81$) and that the output matched the design brief (Q2: Remix $M = 6.12$, $SD = 0.89$ vs. Regeneration $M = 4.69$, $SD = 0.79$). They also reported greater support for creative expression (Q3: Remix $M = 5.88$, $SD = 0.89$ vs. Regeneration $M = 4.56$, $SD = 1.26$) and

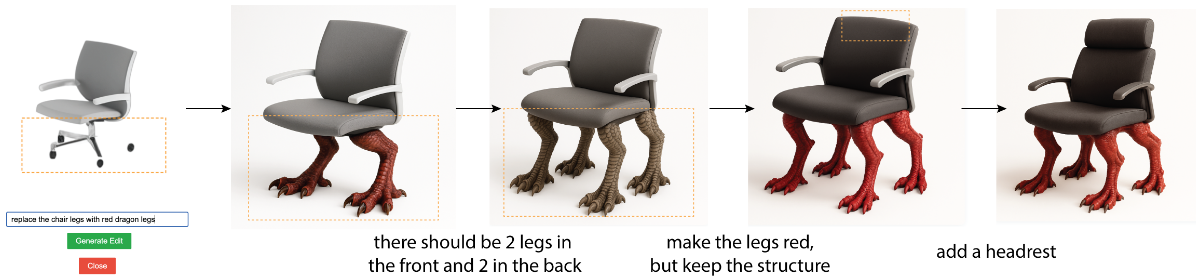

Figure 7: Regeneration workflow in Compos3D. Users highlight a local region (via bounding box in 2D or surface selection in 3D) and provide a textual instruction to specify the desired change. The system regenerates the highlighted region while preserving the rest of the design, allowing users to progressively refine their model across multiple iterations.

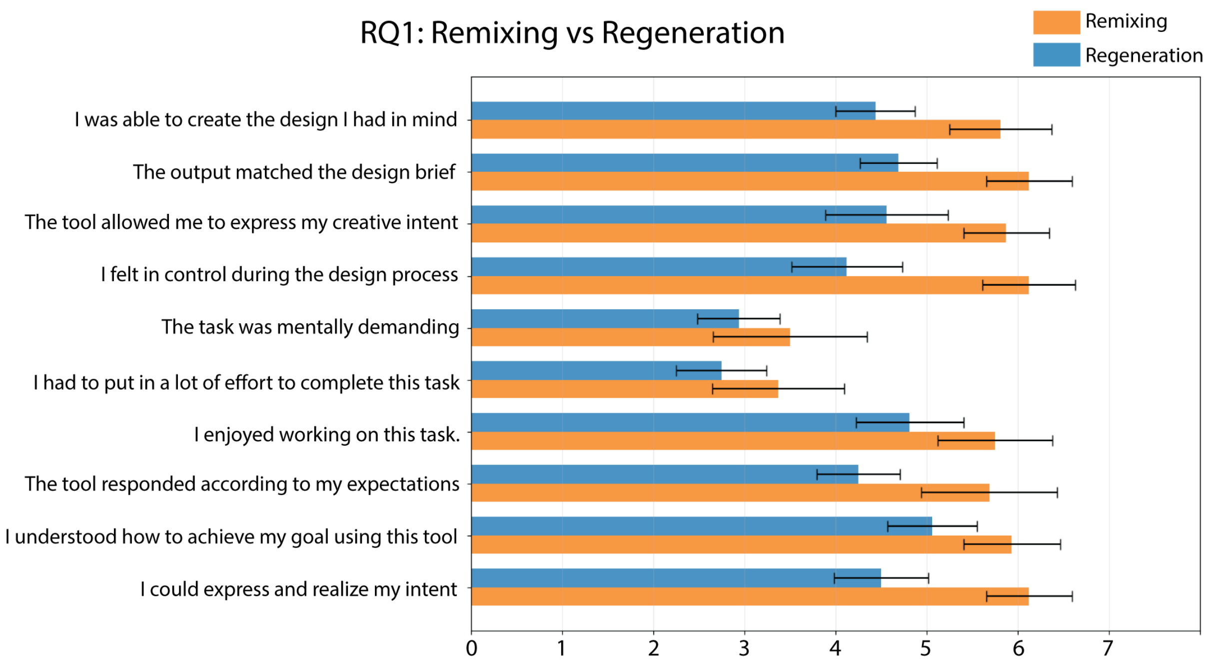

Figure 8: Mean participant ratings (Likert 1–7; 95% CIs) for Remixing (orange) vs. Regeneration (blue) across *Task Outcome* (Q1–Q4), *Cognitive Demand & Effort* (Q5–Q7), and *Tool Responsiveness* (Q8–Q10). Remixing trends higher on creative expression, perceived control, and responsiveness, with a modest increase in reported effort.

felt more in control of the design process (Q4: Remix $M = 6.12$, $SD = 0.96$ vs. Regeneration $M = 4.12$, $SD = 1.15$).

Cognitive Demand & Effort. Ratings here were more mixed. While participants reported slightly higher mental demand and effort in the Remixing condition (Q5: $M = 3.50$, $SD = 1.59$ vs. Regeneration $M = 2.94$, $SD = 0.85$; Q6: $M = 3.38$, $SD = 1.36$ vs. $M = 2.75$, $SD = 0.93$), they nonetheless found Remixing more enjoyable (Q7: $M = 5.75$, $SD = 1.18$ vs. $M = 4.81$, $SD = 1.11$). This suggests that the increased effort of Remixing was offset by greater satisfaction and engagement.

Tool Responsiveness. Participants rated Remixing as significantly more responsive and aligned with their goals. They reported that the tool responded according to expectations (Q8: Remix $M = 5.69$, $SD = 1.40$ vs. Regeneration $M = 4.25$, $SD = 0.86$), that they understood how to achieve their goals more clearly (Q9: Remix $M = 5.94$, $SD = 1.00$ vs. Regeneration $M = 5.06$, $SD = 0.93$), and that they could express their intent more effectively (Q10: Remix $M = 6.12$, $SD = 0.89$ vs. Regeneration $M = 4.50$, $SD = 0.97$).

Taken together, these results indicate a consistent advantage for Remixing: despite requiring slightly more effort, it provided participants with greater creative control, stronger alignment with their design intent, and a more enjoyable overall experience.

Qualitative Findings: Participants consistently emphasized that *Remixing* gave them a stronger sense of creative control than Regeneration. Several noted that Regeneration often removed or altered features they wanted to keep, whereas remixing allowed them to deliberately preserve and recombine elements: "*Regeneration sometimes removes features I want, but remixing gives me an option to bring those lost features back into the final model*" (P6). Others highlighted that Remixing better aligned with how they envisioned AI in the creative process: "*More aligned with how I wish to use AI — give high level description on what I want, and then use the tool to get refined features*" (P8).

Participants also framed their preference in terms of explicit control and reduced unintended changes: "*I felt remixing gave me more control over the final product [and] introduced less unintended changes than regeneration*" (P2). Another participant explained, "*It's hard to generate an image or a 3D model that is perfect, and remixing allows me to combine different parts which I think are perfect in their own setting. It makes me feel I am controlling the process*" (P3).

This theme of "control" extended beyond feature preservation to spatial manipulation. As one participant noted: "*I feel in more control on objects, including the position and scale*" (P4). Overall, participants described Remixing as a more intentional and reliable process.

5.2 RQ2: Modality Preference - 2D or 3D

To compare the two modalities, we analyzed post-task questionnaire responses across three dimensions—*Task Outcome* (Q1–Q4), *Cognitive Demand & Effort* (Q5–Q7), and *Tool Responsiveness* (Q8–Q10). Figure 9 presents these results. Overall, ratings for 2D and 3D were largely similar, with modest advantages for 3D in creative expression and enjoyment, and advantages for 2D in speed and selection clarity.

Task Outcome. Ratings for task outcome were generally comparable across modalities. Participants reported similar ability to create the design they had in mind (Q1: 2D M = 5.00, SD = 1.21 vs. 3D M = 5.25, SD = 1.13) and to match the design brief (Q2: 2D M = 5.44, SD = 1.15 vs. 3D M = 5.38, SD = 1.09). The 3D modality was rated somewhat higher for supporting creative expression (Q3: 3D M = 5.50, SD = 1.32 vs. 2D M = 4.94, SD = 1.18), whereas feelings of control were similar (Q4: 2D M = 5.19, SD = 1.22 vs. 3D M = 5.06, SD = 1.69).

Cognitive Demand & Effort. Ratings for mental demand and effort were nearly identical across 2D and 3D. For mental demand (Q5), scores were close (2D M = 3.25, SD = 1.29 vs. 3D M = 3.19, SD = 1.33), as with perceived effort (Q6: 2D M = 3.00, SD = 1.03 vs. 3D M = 3.12, SD = 1.36). Enjoyment was slightly higher in 3D (Q7: M = 5.50, SD = 1.15 vs. 2D M = 5.06, SD = 1.29).

Tool Responsiveness. Both modalities were rated similarly in responsiveness and clarity. Participants felt the tool responded according to expectations (Q8: 2D M = 4.88, SD = 1.15 vs. 3D M = 5.06, SD = 1.57), and reported equal understanding of how to achieve their goals (Q9: both M = 5.50). Expressing intent was also rated equally across modalities (Q10: both M = 5.31).

Overall, participants perceived 2D and 3D modalities as largely equivalent in effectiveness. The 3D modality offered modest advantages for supporting creative expression and enjoyment, while 2D remained slightly easier to use, especially for quickly aligning with the design brief.

Qualitative Findings. Most participants reported that the 3D modality felt more intuitive because it allowed them to directly manipulate geometry and align parts in space. As one participant put it, "*I found 3D modality more intuitive because it lets me more easily place objects where I want them and I can make very specific adjustments*" (P1). Another explained that working directly in 3D removed the need to mentally translate from 2D images: "*Working with 3D was more intuitive than working with 2D because when working with 2D I always had to think about how the 2D–3D translation will affect the shape and design*" (P5).

Several participants emphasized that 3D manipulation gave them more control over how parts came together: "For producing a 3D model, working directly in 3D and being able to align elements directly in 3D worked better for me. Probably because I did not have to imagine a 3D output from a 2D image" (P2). Others highlighted the ability to inspect models from multiple viewpoints: "*I can see and manipulate things with different angles*" (P4).

At the same time, a minority found 2D interactions simpler and less ambiguous, particularly for segmentation: "*I think the selection aspect of 2D was more intuitive than the view/angle part of 3D, and it seemed to respond better*" (P7). Another echoed, "*I had more control with 2D than with 3D with selectable areas, also lesser ambiguity for the AI model*" (P8).

Overall, participants valued 3D for precision and spatial reasoning, but recognized that 2D offered speed and simplicity in selection. This reflects a trade-off between the intuitiveness of direct 3D manipulation and the ease of quick 2D operations.

5.3 RQ3: What Perceived Trade-offs across workflows

To complement the per-task ratings, the post-study questionnaire asked participants to compare the four conditions directly—identifying their preferred tool and modality, the most accurate Tool × Modality combination, the condition that best supported freedom and control, the one that required the most effort, and where they felt most confident in meeting the brief.

Participants expressed a unanimous preference for Remixing over Regeneration (100%, 8/8). When comparing modalities, most found 3D more intuitive (75%, 6/8), while 25% (2/8) preferred 2D. These results come from single-choice questions, where participants were asked to select one preferred workflow or modality. For task accuracy, Remixing–3D was most often selected (62.5%, 5/8), followed by Remixing–2D (37.5%, 3/8). Perceptions of creative freedom and control also centered around Remixing–3D (75%, 6/8), with the remainder selecting Remixing–2D (25%, 2/8).

In contrast, effort was most strongly associated with Regeneration–3D (62.5%, 5/8), followed by Remixing–3D (37.5%, 3/8). Confidence in fulfilling the design brief was highest for Remixing–3D (62.5%, 5/8), with fewer participants selecting Remixing–2D (25%, 2/8) and Regeneration–3D (12.5%, 1/8).

Together, these results highlight a clear trade-off: while Remixing–3D demanded greater effort than 2D-based workflows, it was also seen as the condition providing the greatest control, creative freedom, and confidence. Participants described 2D as faster and more straightforward for initial segmentation, but 3D as offering stronger precision and spatial reasoning.

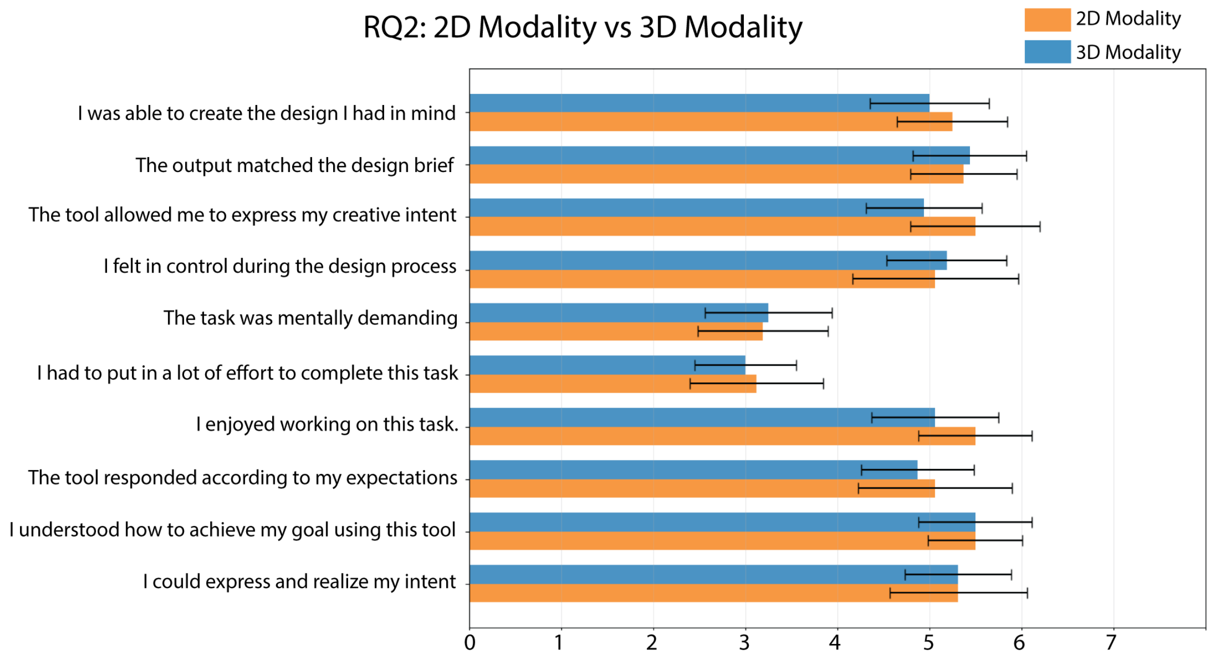

Figure 9: Mean participant ratings (Likert 1–7; 95% CIs) for 2D (green) vs. 3D (purple) across *Task Outcome* (Q1–Q4), *Cognitive Demand & Effort* (Q5–Q7), and *Tool Responsiveness* (Q8–Q10). Ratings were broadly similar across modalities.

5.4 Exploratory Reflections

In addition to addressing our research questions, we asked participants open-ended questions about their design strategies, attachment to outcomes, and perspectives on how AI-based 3D modeling tools should evolve. These reflections highlight how users envision the role of interactivity, agency, and modality in future design tools.

Interactive engagement is valued even if AI were 'perfect'. When asked in the post-study questionnaire, all participants (8/8) reported that they would still want to interactively design even if AI could generate a flawless result from a single prompt. They stressed that design intent often emerges during the process: *"I don't have a very clear idea of what I want at the start, and I need the process to find it out while designing"* (P1). Others emphasized that prompting alone felt limiting: "*A prompt always leaves a lot open in regards to execution*" (P3). One participant noted, "*I don't think I am capable of expressing what I want accurately in one go*" (P6). Another added, "*Creativity cannot be expressed in one step, but only in multiple iterations*" (P7).

Design thinking alternates between whole objects and parts. Participants described shifting fluidly between reasoning about the whole design and its individual components. For some, part-based thinking was primary: "*When designing this model, I was definitely thinking more in terms of parts or components*" (P5). Others alternated: "*I switched between parts and the whole a lot of times*" (P1). But some participants shared how thinking in whole parts helped them brainstorm the overall design: "*Thinking in terms of whole objects allowed me to brainstorm ideas and also extract interesting parts*" (P7).

2D and 3D modalities shape different modes of thinking. Reflections echoed the trade-offs identified in RQ2 (Sec 5.2). Participants described 3D as deliberate and precise: "*In 3D I have very explicit control... more effort but more precise results*" (P1), "*3D gave me more information and made it feel like I am actually designing"* (P2). By contrast, 2D was described as lighter-weight but less under the user's authorship: "*In 2D, AI kind of takes over the design… the results are nice but I don't feel as much that they are mine*" (P1). Still, a few found 2D segmentation easier: *"The selection aspect of 2D was more intuitive than the view/angle part of 3D"* (P7).

Ownership emerges from composing parts. Nearly all participants stated they felt more attached to designs composed from parts than to those generated all at once. As one put it, *"Generating didn't feel like I was creating, more like typing something and seeing what happens* (P2). P5 shared that composition led to ownership, *"I did feel more attached to designs composed from part... [as I] need to think about which parts to include how to arrange them,"* and *"It was my design intent that actually translated when I chose the components."* P2 summarized: *"You feel attached to what you create."*

Together, these reflections suggest that AI-assisted 3D modeling tools should prioritize *co-creation over automation*, supporting workflows where users iteratively shape, refine, and compose designs. Such approaches not only improve control and precision but also foster a stronger sense of ownership and creative engagement.

6 Applications

To illustrate the relevant use cases for Compos3D, we present four applications that highlight how remixing workflows extend beyond prompt-based generation. These examples demonstrate how users can assemble complex multi-part designs, reframe flaws as opportunities for creative repair, and combine multiple personal artifacts into new compositions, which can be either used as digital artifacts or be fabricated for use in the real world.

6.1 Part-Based Creative 3D Modeling with Generative AI

A central capability of Compos3D is supporting designs that draw from multiple sources. Rather than regenerating a whole object in hopes that the model preserves certain features while adding new ones, users can deliberately select and combine distinct components to construct complex artifacts. This approach also unlocks a larger space of 3D models than the AI model was originally trained on, enabling users to create out-of-distribution 3D designs. In the example shown in Figure 10, the user assembled five separate elements to create a unique design of a teapot. They extract a floral motif mug body, dove spout, vine handle, Roman pedestal base, and bamboo lid as source models. These parts, sourced from different generated models, were arranged into a collage (Fig.10b). Compos3D synthesized them into a coherent teapot design (Fig.10c), preserving both the recognizable form of each segment and the overall aesthetic harmony.

This example illustrates how multi-part remixing enables users to utilize intermediate generations as raw material for design and create novel, out-of-distribution 3D results.

6.2 Remixing as Repair

Generative 3D modeling often produces incomplete or broken geometry that requires post-processing or refinement [51]. While current systems treat such flaws as errors to be corrected with 3D modeling tools, we suggest they can instead serve as opportunities for creative intervention. We take inspiration from Japanese *Kintsugi* [37], where broken pottery is repaired with precious materials to highlight rather than conceal damage. Similarly, Compos3D allows users to approach these errors as sites of creative expression.

Figure 11 illustrates this idea. A generated mug (a–b) contained a faulty handle that would normally be treated as an error. Traditional correction would require complex 3D modeling operations, such as manually editing meshes and recoloring surfaces, skills that are inaccessible to most novices. Instead, Compos3D abstracts away these low-level operations and lets users intervene at the design level. Here, the user added a gear model in the 3D composition stage (Fig 11c), augmenting the faulty region with a new element. The system then synthesized a coherent output (Figure 11d), producing a distinctive aesthetic feature that emerged from remixing.

6.3 Composition with Existing Artifacts from the Environment

People often associate strong memories and personal meaning with everyday artifacts—old gifts, souvenirs, or toys—that they want to preserve and reimagine in new contexts. Prior work on memory reconstruction, such as IntelRecon [28], highlights how digitizing personal items allows them to serve as lasting memory triggers. However, existing digitization methods typically replicate objects as static forms, missing opportunities to creatively integrate them into new designs.

With Compos3D, users can upload photos of their personal artifacts, recreate them with generative 3D models, and then remix them into new compositions. This enables objects to live on not just as preserved scans, but as active design materials that carry forward their sentimental value.

We demonstrate this in Figure 12 with a lamp assembled from everyday items. The user takes a plush toy as the lampshade, a decorative plate as the base, a mug as the stem, and a toy chicken as a pull chain. These parts, digitized and synthesized with Compos3D, yield an object that is both unique and personally meaningful. This shows how generative remixing can bridge past memories and future creations, letting users embed personal history into the design of new artifacts.

6.4 From Generative Remixing to Personal Fabrication

A core vision in HCI is to empower end-users to create personally meaningful artifacts through digital fabrication [3, 14, 15]. Compos3D contributes to this space by enabling users to go beyond modification toward composition. Users can assemble hybrid artifacts from multiple generative outputs, repair flawed geometry through remixing, and embed personal objects into new designs.

As shown in Figure 12e, the lamp assembled from everyday items can also fabricated and used. This demonstrates how remixing workflows can contribute towards closing the loop for personal design: picking objects as inspiration from the real world, remixing them into new designs, and fabricating them back into objects to be used in the real world.

7 Discussion

Our study demonstrates that remixing enables new forms of interaction with generative 3D models, offering users greater control, expressiveness, and satisfaction than trial-and-error prompting alone. In this section, we translate our findings into design recommendations for future systems, reflect on remixing as a broader paradigm for generative 3D design, examine trade-offs between 2D and 3D modalities, and outline limitations and directions for future work.

7.1 Design Recommendations for AI-Assisted 3D Modeling

From our study findings and exploratory reflections, we distill three higher-level themes. These themes highlight how future AI-assisted 3D modeling systems can better support creative control, reduce effort, and foster user agency.

1. Prioritize Remixing over Regeneration. Participants repeatedly emphasized that they wanted agency over outcomes rather than passively accept model outputs. Remixing supported this by enabling part-based composition and preserving desired features: "*Remixing gave me more control over the final product and introduced fewer unintended changes*" (P2). To foster ownership, tools should let users present and control high-level intent through part-based composition, deliberately preserving and recombining elements. Interactive workflows help ensure outcomes reflect users' design intent and control for model-driven randomness.

Figure 10: Part-based creative modeling with Compos3D. (a) Source components selected from multiple generated models: a floral motif mug body, dove spout, vine handle, Roman pedestal base, and bamboo lid. (b) User-assembled collage combining the selected parts. (c) Final synthesized teapot, where the system integrates the disparate elements into a coherent design.

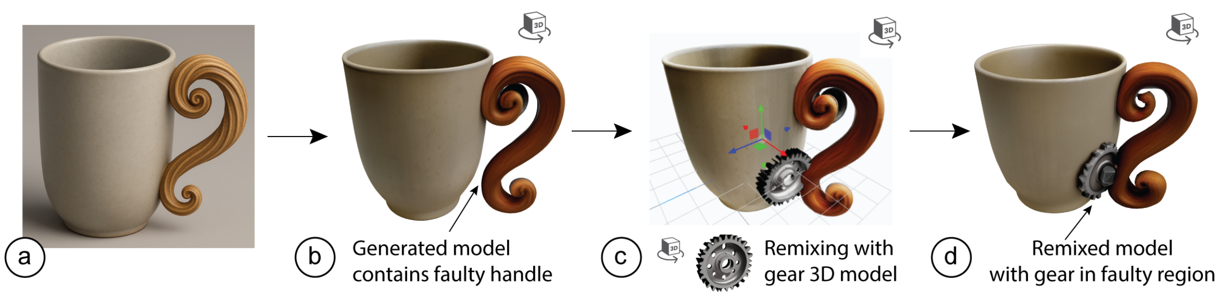

Figure 11: Remixing as repair with Compos3D. (a) A generated mug. (b) The resulting 3D model contains a faulty handle. (c) In the 3D composition stage, the user remixes the design by inserting a gear model into the faulty region. (d) The final synthesized model integrates the gear, transforming a defect into a distinctive design feature.

2. Balance 2D Speed with 3D Precision. Our results highlight a complementary relationship between 2D and 3D interactions. 2D provided speed and intuitive selection, while 3D afforded spatial precision but demanded more effort. As P7 noted, "*The selection aspect of 2D was more intuitive than the view/angle part of 3D*," whereas P2 stressed, "*3D gave me more information and made it feel like I am actually designing*." To integrate these strengths, systems should allow fluid switching between 2D canvas-based exploration and 3D spatial manipulation while scaffolding laborious tasks with aids such as segmentation, alignment snapping tools, etc.

3. Frame AI as a Transparent and Steerable Collaborator. Participants distinguished between the kinds of work they wanted AI to automate and the parts they wanted to control. As P2 reflected, "*In 3D it's actually taking over the labor I wouldn't want to do anyway, while in 2D it was taking over things I wanted more control over.*" This suggests that AI should be framed as a collaborator that automates low-level refinements (e.g., mesh cleanup, texture details) while leaving high-level decisions to the designer. Systems should provide transparency and steerable feedback loops, so users can see what the AI is modifying, regulate its influence, and override changes when needed.

Together, these three themes highlight that the value of generative AI in 3D design lies not only in producing high-quality outputs

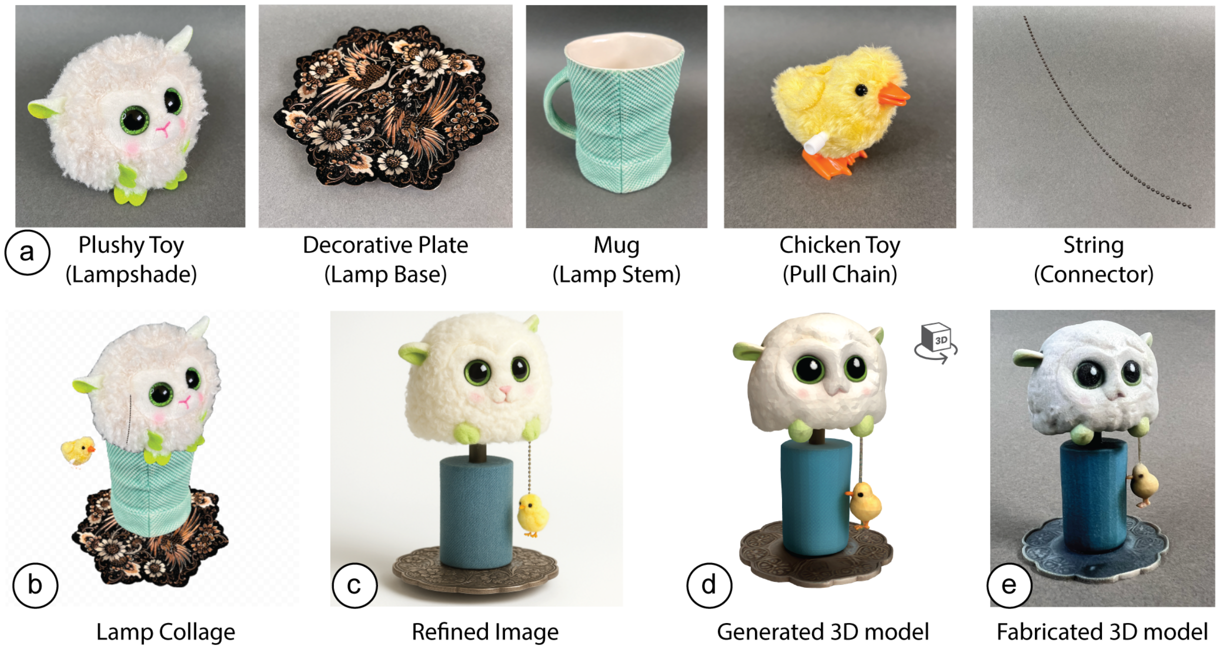

Figure 12: Composition with existing artifacts in Compos3D. (a) Everyday items used as source parts: plush toy (lampshade), decorative plate (base), mug (stem), chicken toy (pull chain), and string (connector). (b) Collage assembled from these components. (c) Refined image showing the composite lamp design. (d) Final generated 3D model, transforming familiar personal objects into a new composed artifact, that can then be e) fabricated and used in the real-world.

but in enabling workflows that balance efficiency, precision, and user agency.

7.2 Hybrid Workflows: Combining Remixing and Regeneration

While our study compared remixing and regeneration as distinct paradigms, our findings suggest that future systems could benefit from integrating them. Each workflow has complementary strengths: remixing affords deliberate preservation and recombination of parts, while regeneration supports rapid exploration of stylistic alternatives. Similarly, 2D and 3D modalities offer complementary affordances of speed and precision. Compos3D provides a first step toward understanding these axes in isolation, but did not explore how they might be combined. An important opportunity for future work is to design hybrid workflows that allow users to fluidly move between remixing and regeneration, or between 2D and 3D, within the same creative session. Such systems could give users layered forms of control—sketching quickly in 2D, refining precisely in 3D, remixing to preserve key parts, and regenerating to explore broader stylistic diversity—thereby supporting both efficiency and depth in generative 3D modeling.

7.3 Limitations and Future Work

Our study focused on a single design task as a first step toward exploring remixing workflows in 3D generation. Future work could expand the scope of the study to include larger-scale tasks and more intricate assemblies, such as multi-part mechanical objects, which may introduce additional requirements. Additionally, our current system recombines 3D geometry but does not yet account for fabrication constraints such as printability or material properties. Future work could integrate physics-aware generation, collaborative remixing, or adaptive interfaces that recommend compositional edits based on user history.

8 Conclusion

In this work, we introduced Compos3D, a system that supports a compositional remixing workflow for 3D generation. Rather than treating models as indivisible outputs, Compos3D enables users to select parts from multiple candidates through both 2D and 3D modalities, and compose them into a new coherent design. Our controlled user study demonstrates that this novel remixing-based approach requires slightly more effort, but provides greater creative control, alignment with user intent, and enjoyment than regeneration. We further highlight trade-offs between 2D and 3D modalities: while 2D offers speed and simplicity, 3D affords precision and spatial reasoning. From these findings, we distilled design guidelines that emphasize collaborative creation with AI, support for part-based thinking, and integration of multiple modalities.

References

[1] Shm Garanganao Almeda, JD Zamfirescu-Pereira, Kyu Won Kim, Pradeep Mani Rathnam, and Bjoern Hartmann. 2024. Prompting for discovery: Flexible sense-making for ai art-making with dreamsheets. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–17.

[2] Tyler Angert, Miroslav Suzara, Jenny Han, Christopher Pondoc, and Hariharan Subramonyam. 2023. Spellburst: A node-based interface for exploratory creative

coding with natural language prompts. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–22.
[3] Patrick Baudisch, Stefanie Mueller, et al. 2017. Personal fabrication. *Foundations and Trends® in Human–Computer Interaction* 10, 3–4 (2017), 165–293.
[4] Karim Benharrak and Amy Pavel. 2025. HistoryPalette: Supporting Exploration and Reuse of Past Alternatives in Image Generation and Editing. *arXiv preprint arXiv:2501.04163* (2025).
[5] Mark Boss, Zixuan Huang, Aaryaman Vasishta, and Varun Jampani. 2024. Sf3d: Stable fast 3d mesh reconstruction with uv-unwrapping and illumination disentanglement. *arXiv preprint arXiv:2408.00653* (2024).
[6] Stephen Brade, Bryan Wang, Mauricio Sousa, Sageev Oore, and Tovi Grossman. 2023. Promptify: Text-to-image generation through interactive prompt exploration with large language models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–14.
[7] Ruojin Cai, Guandao Yang, Hadar Averbuch-Elor, Zekun Hao, Serge Belongie, Noah Snavely, and Bharath Hariharan. 2020. Learning gradient fields for shape generation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*. Springer, 364–381.
[8] Tracy Diane Cassidy. 2008. Mood boards: Current practice in learning and teaching strategies and students' understanding of the process. *International journal of fashion design* 1, 1 (2008), 43–54.
[9] Siddhartha Chaudhuri, Evangelos Kalogerakis, Stephen Giguere, and Thomas Funkhouser. 2013. Attribit: Content Creation with Semantic Attributes. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology* (St. Andrews, Scotland, United Kingdom) *(UIST '13)*. Association for Computing Machinery, New York, NY, USA, 193–202. doi:10.1145/2501988.2502008
[10] Gene Chou, Yuval Bahat, and Felix Heide. 2023. Diffusion-sdf: Conditional generative modeling of signed distance functions. In *Proceedings of the IEEE/CVF international conference on computer vision*. 2262–2272.
[11] John Joon Young Chung and Eytan Adar. 2023. Promptpaint: Steering text-to-image generation through paint medium-like interactions. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–17.
[12] Nigel Cross. 2006. *Designerly ways of knowing*. Springer.
[13] Hai Dang, Frederik Brudy, George Fitzmaurice, and Fraser Anderson. 2023. Worldsmith: Iterative and expressive prompting for world building with a generative ai. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–17.
[14] Faraz Faruqi, Ahmed Katary, Tarik Hasic, Amira Abdel-Rahman, Nayeemur Rahman, Leandra Tejedor, Mackenzie Leake, Megan Hofmann, and Stefanie Mueller. 2023. Style2Fab: Functionality-Aware Segmentation for Fabricating Personalized 3D Models with Generative AI. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–13.
[15] Faraz Faruqi, Maxine Perroni-Scharf, Jaskaran Singh Walia, Yunyi Zhu, Shuyue Feng, Donald Degraen, and Stefanie Mueller. 2025. TactStyle: Generating Tactile Textures with Generative AI for Digital Fabrication. *arXiv preprint arXiv:2503.02007* (2025).
[16] Emma Frid, Celso Gomes, and Zeyu Jin. 2020. Music creation by example. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13.
[17] Jun Gao, Tianchang Shen, Zian Wang, Wenzheng Chen, Kangxue Yin, Daiqing Li, Or Litany, Zan Gojcic, and Sanja Fidler. 2022. GET3D: A Generative Model of High Quality 3D Textured Shapes Learned from Images. In *Advances In Neural Information Processing Systems*.
[18] Lena Hegemann and Antti Oulasvirta. 2024. Palette, purpose, prototype: The three ps of color design and how designers navigate them. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–19.
[19] Megan Hofmann, Gabriella Hann, Scott E. Hudson, and Jennifer Mankoff. 2018. Greater than the Sum of Its PARTs: Expressing and Reusing Design Intent in 3D Models. 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. doi:10.1145/3173574.3173875
[20] Yicong Hong, Kai Zhang, Jiuxiang Gu, Sai Bi, Yang Zhou, Difan Liu, Feng Liu, Kalyan Sunkavalli, Trung Bui, and Hao Tan. 2023. Lrm: Large reconstruction model for single image to 3d. *arXiv preprint arXiv:2311.04400* (2023).
[21] Heewoo Jun and Alex Nichol. 2023. Shap-e: Generating conditional 3d implicit functions. *arXiv preprint arXiv:2305.02463* (2023).
[22] Tae Soo Kim, Yoonjoo Lee, Jamin Shin, Young-Ho Kim, and Juho Kim. 2024. Evallm: Interactive evaluation of large language model prompts on user-defined criteria. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–21.
[23] 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. In *Proceedings of the IEEE/CVF international conference on computer vision*. 4015–4026.
[24] Roman Klokov, Edmond Boyer, and Jakob Verbeek. 2020. Discrete point flow networks for efficient point cloud generation. In *European Conference on Computer Vision*. Springer, 694–710.
[25] Janin Koch, Nicolas Taffin, Andrés Lucero, and Wendy E Mackay. 2020. SemanticCollage: Enriching digital mood board design with semantic labels. In *Proceedings of the 2020 ACM designing interactive systems conference*. 407–418.
[26] Yuki Koyama, Shinjiro Sueda, Emma Steinhardt, Takeo Igarashi, Ariel Shamir, and Wojciech Matusik. 2015. AutoConnect: Computational Design of 3D-Printable Connectors. *ACM Trans. Graph.* 34, 6, Article 231 (nov 2015), 11 pages. doi:10.1145/2816795.2818060
[27] Brian Lee, Savil Srivastava, Ranjitha Kumar, Ronen Brafman, and Scott R Klemmer. 2010. Designing with interactive example galleries. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2257–2266.
[28] Zisu Li, Jiawei Li, Zeyu Xiong, Shumeng Zhang, Faraz Faruqi, Stefanie Mueller, Chen Liang, Xiaojuan Ma, and Mingming Fan. 2025. InteRecon: Towards Reconstructing Interactivity of Personal Memorable Items in Mixed Reality. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–19.
[29] Sean J Liu, Nupur Kumari, Ariel Shamir, and Jun-Yan Zhu. 2025. Generative Photomontage. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 7931–7941.
[30] Vivian Liu, Jo Vermeulen, George Fitzmaurice, and Justin Matejka. 2023. 3DALL-E: Integrating text-to-image AI in 3D design workflows. In *Proceedings of the 2023 ACM designing interactive systems conference*. 1955–1977.
[31] Damien Masson, Sylvain Malacria, Géry Casiez, and Daniel Vogel. 2024. Directgpt: A direct manipulation interface to interact with large language models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–16.
[32] Paritosh Mittal, Yen-Chi Cheng, Maneesh Singh, and Shubham Tulsiani. 2022. Autosdf: Shape priors for 3d completion, reconstruction and generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 306–315.
[33] Xiaohan Peng, Janin Koch, and Wendy E Mackay. 2024. Designprompt: Using multimodal interaction for design exploration with generative ai. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*. 804–818.
[34] Xiaohan Peng, Janin Koch, and Wendy E Mackay. 2025. FusAIn: Composing Generative AI Visual Prompts Using Pen-based Interaction. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–20.
[35] Savvas Petridis, Nicholas Diakopoulos, Kevin Crowston, Mark Hansen, Keren Henderson, Stan Jastrzebski, Jeffrey V Nickerson, and Lydia B Chilton. 2023. Anglekindling: Supporting journalistic angle ideation with large language models. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–16.
[36] Thijs Jan Roumen, Willi Müller, and Patrick Baudisch. 2018. Grafter: Remixing 3D-Printed Machines. 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. doi:10.1145/3173574.3173637
[37] Céline Santini. 2019. *Kintsugi: Finding strength in imperfection*. Andrews McMeel Publishing.
[38] Alexander Scarlatos, Yusong Wu, Ian Simon, Adam Roberts, Tim Cooijmans, Natasha Jaques, Cassie Tarakajian, and Anna Huang. 2025. ReaLJam: Real-Time Human-AI Music Jamming with Reinforcement Learning-Tuned Transformers. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–9.
[39] Ryan Schmidt and Karan Singh. 2010. Meshmixer: An Interface for Rapid Mesh Composition. In *ACM SIGGRAPH 2010 Talks*. Association for Computing Machinery, New York, NY, USA. doi:10.1145/1837026.1837034
[40] Adriana Schulz, Ariel Shamir, David I. W. Levin, Pitchaya Sitthi-amorn, and Wojciech Matusik. 2014. Design and Fabrication by Example. *ACM Trans. Graph.* 33, 4, Article 62 (jul 2014), 11 pages. doi:10.1145/2601097.2601127
[41] 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 2024 CHI Conference on Human Factors in Computing Systems*. 1–18.
[42] Maria Shugrina, Ariel Shamir, and Wojciech Matusik. 2015. Fab Forms: Customizable Objects for Fabrication with Validity and Geometry Caching. *ACM Trans. Graph.* 34, 4, Article 100 (jul 2015), 12 pages. doi:10.1145/2766994
[43] Blair Subbaraman, Shenna Shim, and Nadya Peek. 2023. Forking a sketch: How the openprocessing community uses remixing to collect, annotate, tune, and extend creative code. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. 326–342.
[44] Hariharan Subramonyam, Roy Pea, Christopher Lawrence Pondoc, Maneesh Agrawala, and Colleen Seifert. 2023. Bridging the Gulf of envisioning: Cognitive design challenges in LLM interfaces. *arXiv preprint arXiv:2309.14459* (2023).
[45] Sangho Suh, Bryan Min, Srishti Palani, and Haijun Xia. 2023. Sensecape: Enabling multilevel exploration and sensemaking with large language models. In *Proceedings of the 36th annual ACM symposium on user interface software and technology*. 1–18.
[46] Jiaxiang Tang, Zhaoxi Chen, Xiaokang Chen, Tengfei Wang, Gang Zeng, and Ziwei Liu. 2024. Lgm: Large multi-view gaussian model for high-resolution 3d content creation. In *European Conference on Computer Vision*. Springer, 1–18.

[47] Tom Veuskens, Florian Heller, and Raf Ramakers. 2021. CODA: A Design Assistant to Facilitate Specifying Constraints and Parametric Behavior in CAD Models. In *Graphics Interface 2021*. https://openreview.net/forum?id=1dLDPJeafRZ

[48] Zhijie Wang, Yuheng Huang, Da Song, Lei Ma, and Tianyi Zhang. 2024. Promptcharm: Text-to-image generation through multi-modal prompting and refinement. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–21.

[49] Tongshuang Wu, Michael Terry, and Carrie Jun Cai. 2022. Ai chains: Transparent and controllable human-ai interaction by chaining large language model prompts. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–22.

[50] Zhijie Wu, Xiang Wang, Di Lin, Dani Lischinski, Daniel Cohen-Or, and Hui Huang. 2019. Sagnet: Structure-aware generative network for 3d-shape modeling. *ACM Transactions on Graphics (TOG)* 38, 4 (2019), 1–14.

[51] Jianfeng Xiang, Zelong Lv, Sicheng Xu, Yu Deng, Ruicheng Wang, Bowen Zhang, Dong Chen, Xin Tong, and Jiaolong Yang. 2025. Structured 3d latents for scalable and versatile 3d generation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 21469–21480.

[52] Jiale Xu, Weihao Cheng, Yiming Gao, Xintao Wang, Shenghua Gao, and Ying Shan. 2024. Instantmesh: Efficient 3d mesh generation from a single image with sparse-view large reconstruction models. *arXiv preprint arXiv:2404.07191* (2024).

[53] Zihan Yan, Chunxu Yang, Qihao Liang, and Xiang'Anthony' Chen. 2023. XCreation: A graph-based crossmodal generative creativity support tool. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–15.

[54] Guandao Yang, Xun Huang, Zekun Hao, Ming-Yu Liu, Serge Belongie, and Bharath Hariharan. 2019. Pointflow: 3d point cloud generation with continuous normalizing flows. In *Proceedings of the IEEE/CVF international conference on computer vision*. 4541–4550.

[55] Junran Yang, Andrew M McNutt, and Leilani Battle. 2024. Considering Visualization Example Galleries. In *2024 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. IEEE, 329–343.

[56] J Diego Zamfirescu-Pereira, Richmond Y Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny can't prompt: how non-AI experts try (and fail) to design LLM prompts. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–21.

[57] Lei Zhang, Jin Pan, Jacob Gettig, Steve Oney, and Anhong Guo. 2024. Vrcopilot: Authoring 3d layouts with generative ai models in vr. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–13.

[58] Yuchen Zhou, Jiayuan Gu, Tung Yen Chiang, Fanbo Xiang, and Hao Su. 2024. Point-sam: Promptable 3d segmentation model for point clouds. *arXiv preprint arXiv:2406.17741* (2024).