FusionProtor: A Mixed-Prototype Tool for Component-level Physical-to-Virtual 3D Transition and Simulation

Hongbo Zhang
Zhejiang University
Hangzhou, China
hongbozhang@zju.edu.cn

Pei Chen*
Zhejiang University
Hangzhou, China
chenpei@zju.edu.cn

Xuelong Xie
Zhejiang University
Hangzhou, China
xuelongxie@zju.edu.cn

Zhaoqu Jiang
Zhejiang University
Hangzhou, China
zhaoqujiang@zju.edu.cn

Yifei Wu
Zhejiang University
Hangzhou, China
yfwu0501@zju.edu.cn

Zejian Li
Zhejiang University
Hangzhou, China
zejianlee@zju.edu.cn

Xiaoyu Chen
Zhejiang University
Hangzhou, China
rainchard@zju.edu.cn

Lingyun Sun
Zhejiang University
Hangzhou, China
sunly@zju.edu.cn

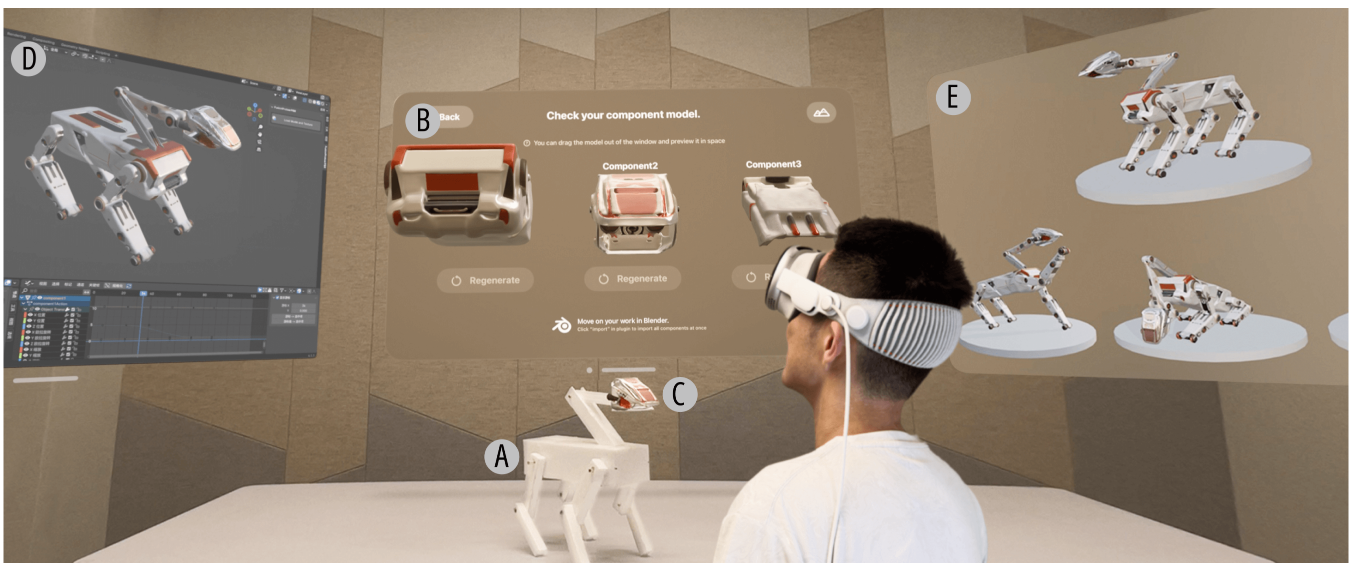

Figure 1: A prototype process supported by FusionProtor. Ⓐ: The designer constructed and interacted with a physical prototype embodiedly. Ⓑ: He utilized FusionProtor to generate component-level 3D prototypes based on the low-fidelity physical prototype. Ⓒ: He coupled the physical prototype with virtual components for mixed iteration. Ⓓ: He assembled 3D components and simulated their interaction according to physical motion logic. Ⓔ: FusionProtor's 3D design outcome and simulation presentation.

*The corresponding author.

CHI '25, Yokohama, Japan

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713686

Abstract

Developing and simulating 3D prototypes is crucial in product conceptual design for ideation and presentation. Traditional methods often keep physical and virtual prototypes separate, leading to a disjointed prototype workflow. In addition, acquiring high-fidelity prototypes is time-consuming and resource-intensive, distracting designers from creative exploration. Recent advancements in generative artificial intelligence (GAI) and extended reality (XR) provided new solutions for rapid prototype transition and mixed simulation. We conducted a formative study to understand current challenges in the traditional prototype process and explore how to effectively utilize GAI and XR ability in prototype. Then we introduced FusionProtor, a mixed-prototype tool for component-level 3D prototype transition and simulation. We proposed a step-by-step generation pipeline in FusionProtor, effectively transiting 3D prototypes from physical to virtual and low- to high-fidelity for rapid ideation and iteration. We also innovated a component-level 3D creation method and applied it in XR environment for the mixed-prototype presentation and interaction. We conducted technical and user experiments to verify FusionProtor's usability in supporting diverse designs. Our results verified that it achieved a seamless workflow between physical and virtual domains, enhancing efficiency and promoting ideation. We also explored the effect of mixed interaction on design and critically discussed its best practices for HCI community.

CCS Concepts

• **Human-centered computing → Interactive systems and tools**; • **Applied computing → Computer-aided design**; • **Computing methodologies → Artificial intelligence**.

Keywords

conceptual design, 3D prototype, generative AI

ACM Reference Format:

Hongbo Zhang, Pei Chen, Xuelong Xie, Zhaoqu Jiang, Yifei Wu, Zejian Li, Xiaoyu Chen, and Lingyun Sun. 2025. FusionProtor: A Mixed-Prototype Tool for Component-level Physical-to-Virtual 3D Transition and Simulation. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26-May 1, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3706598.3713686

1 Introduction

Developing and simulating 3D prototypes is essential during product design and development. These prototypes are integral for exploring the design space, testing feasibility, communicating ideas to stakeholders, and furnishing actionable implementation details [10]. Given that conceptual design determines up to 70-80% of a product's lifetime cost [12, 22], exploring novel and efficient prototype tools is indeed valuable.

Various prototypes have inherent strengths and application scope in conceptual design [40]. The physical prototype supports the embodied interaction and tangible testing, helping to reveal potential flaws in advance compared with the sketch [64]. As these early physical prototypes are hard to manage versions [26] and lack fidelity [15, 40], once designers determine the basic physical scale, shape, and structures, they generally transition to virtual forms, which offer enough fidelity for further refinement and presentation [45]. In the traditional transition process, acquiring high-fidelity prototypes is time-consuming and resource-intensive, involving multiple manual adjustments [44]. Besides, the previous transition usually makes the physical and virtual domains operate independently, leading to a disjointed workflow [25].

Recent advancements in generative artificial intelligence (GAI) and extended reality (XR) have provided new solutions for rapid prototype transition and bridging interaction between physical and virtual prototypes. GAI enables designers to derive various design schemes from rough expressions such as textual descriptions or sketches [19, 80]. This capability can leverage initial design information for refinement, supporting rapid iteration, fidelity enhancement, and productivity improvement in conceptual design [62]. Additionally, the XR development has further expanded the prototype space, enabling designers to connect physical and virtual space for more intuitive and immersive ideation.

Aiming to effectively utilize GAI and XR ability in prototyping, we conducted a formative study with design experts to understand current challenges in the traditional prototype process and consider how to combine GAI and XR to assist the rapid physical-to-virtual 3D prototype transition and simulation task. Our key finding involved that an ideal 3D prototype tool in conceptual design needs to support 1) mixed prototype between physical and virtual domains, 2) rapid transition from physical-to-virtual and low- to high-fidelity, and 3) component-level creation and interaction for flexible iteration and simulation.

In this context, we introduce FusionProtor, a mixed-prototype tool combining GAI and XR for rapid component-level prototype transition and simulation. FusionProtor supports conceptual design in an unprecedented way. It supports the rapid physical-to-virtual 3D transition, fully using early design information in physical prototypes and realizing the joint workflow from low- to high-fidelity. In addition, FusionProtor achieves component-level 3D generation and interaction. It not only allows local component refinement and iteration during the generation but also provides mixed interaction modes with both physical and virtual components in XR.

Following the system development, we initiated a technical evaluation study to validate FusionProtor's technical features. Additionally, we carried out a user evaluation study with 16 designers, confirming FusionProtor's usability and robustness in supporting diverse design tasks. We also clarified FusionProtor's unique strengths and ideation support modes in conceptual design.

This paper makes the following contributions:

- We introduced FusionProtor, a novel mixed-prototype tool integrating GAI and XR technologies. We proposed a step-by-step generation pipeline for physical-to-virtual 3D transition in it, effectively transiting 3D prototypes from physical to virtual, as well as low- to high-fidelity.
- We innovated a component-level 3D creation method and applied it in the XR environment for the mixed prototype. This supports flexible component-level fusion and simulation.
- We conducted comprehensive technical and user evaluation studies, verifying FusionProtor's usability and critically clarifying its strengths, interaction modes, and application scope.

2 Related Work

2.1 Virtual Creation and Prototype in Design

The 3D virtual prototype is a commonly used design representation. It promotes externalizing and conveying detailed intentions and supports motion simulation and engineering calculation for further design implementation [25]. The virtual prototype creation is often time-consuming and resource-intensive, involving multiple iterations and refined adjustments [44]. In design practices, computer-aided freeform surface modeling is one of the mainstream creation methods of 3D virtual prototypes, which is often high-cost, time-consuming, and skill-threshold [40]. The HCI community has also explored other 3D creation methods. For example, Stemasov et al. [67] developed a parametric design tool in XR for personalization artifacts. Similarly, Alcaide-Marzal et al. [1] introduced a modeling method based on deformation rules and grammar. In addition to parametric modeling, Farqui et al. [27] proposed a modeling tool based on point cloud generation for 3D local modification. Moreover, retrieval and refinement based on existing 3D repositories also support more efficient 3D creation [2, 43, 68].

Most 3D creation methods or tools have high learning curves and often distract designers from the creative exploration process [40]. Another significant challenge in prototyping is the inability to effectively utilize preliminary low-fidelity design information for detailed modeling, resulting in a disjointed process. For example, although designers often externalize ideas in a more efficient way (e.g., sketching or physical prototyping) before moving on to virtual modeling, they typically cannot leverage this low-fidelity design information and must start modeling from scratch. This gap hinders effective 3D prototype transitions, making designers hesitant to develop 3D virtual prototypes early in the design process. It leads to missed opportunities for concrete display, interaction simulation, and effective communication in conceptual design [25].

2.2 Physical Prototype and XR in Design

The physical prototype is a common design representation in conceptual design due to its distinct strengths. Initially, as low-fidelity representations, physical prototypes introduce intentional ambiguity, encouraging broader interpretation and reasoning within the design space [64]. In addition, compared with other common prototypes, such as sketches, physical prototypes provide multi-sensory engagement—sight, touch, and even smell—that enhances intuition and interactivity, thereby reducing cognitive load and fostering creativity [30]. This tangible nature allows designers to conduct structural tests and uncover potential flaws, facilitating a more thorough function evaluation [40]. The intuitive dissection and simulation foster deep structural thinking and iterative improvement [75]. However, the long fabrication time, lack of details, and version management are the inherent challenges of early-stage physical prototypes [40].

The HCI community also focuses on integrating physical prototypes into virtual design tools [32]. Among them, utilizing XR technology and using the head-mounted display (HMD) equipment is a mainstream method to construct a hybrid design space for mixed display or interaction, superposing the inherent strengths of both physical and virtual characteristics [36, 52, 53]. For example, Peng et al. [53] proposed an interactive design system supporting digital editing on 3D printing prototypes. Barbieri et al. [11] developed a mixed prototyping tool with physical prototypes for virtual evaluation. However, in previous studies, most of the virtual prototypes were rough sketches or retrieved models from the existing model base, and few of them use generation technology to create prototypes in the mixed world.

2.3 GAI for 3D Creation

The recent surge of GAI has sparked increasing interest across a multitude of design tools integrating GAI. Most of them can offer 2D design schemes [48, 62, 80] or assist design reasoning by text [18, 66], supporting ideation [66], accelerating prototype [29], and advising iteration [46]. Other output representations include the icon [77], UI scheme [38, 41], and animation [78]. However, 3D generation in design has been explored less due to the previous limitations of quality and time. In this context, previous HCI studies adopted some alternative methods. For example, Zhang et al. [82] proposed a mixed prototype containing physical models, but they implemented the 2D generation based on the captured images of physical models instead of direct 3D transition. Similarly, some studies introduced 3D design tools supporting spatial expression, but they achieved stereo image generation rather than a 3D model [35, 61].

The recent dramatic development of GAI, such as the advancements in upscaled model parameters, increased computational power, and greater or larger datasets, has significantly accelerated progress in 3D generation. In this context, the current mainstream 3D generation pipeline can be mainly summarized into two categories: feedforward 3D generation (e.g., LRM [34], DMV3D [79], TripoSR [74], CLAY [49]) and optimization-based 3D generation (e.g., DreamFusion [55], Magic3D [47], MVDream [63], DreamCraft3D [69]). These GAI advancements have enabled real-time 3D generation in design. However, although these 3D advances support content generation, they struggle to fully meet the intricate demands of specific design tasks. For instance, generating complete 3D models without layering or component segmentation poses challenges in refining and iterating designs, as well as in simulating motion and interaction [82]. In this paper, we confront the practical design needs and strive to integrate advanced GAI capabilities to achieve component-level 3D creation from HCI perspective.

3 Formative Study

We conducted a formative study and invited four professional industrial designers to participate in remote interviews. We aim to gain insights into industrial designers' practical prototype process and challenges, as well as understand how to combine GAI and XR to assist the rapid physical-to-virtual 3D prototype transition and simulation task.

3.1 Participants and Procedure

All participants are experts with rich theoretical and practical design experience, the expert information presented in Appendix A. We invited them to participate in a 60-minute semi-structured interview. The guiding questions focus on 1) *the typical prototype workflow in the conceptual design of a product*, 2) *existing design tools and materials in each stage and their strengths and shortcomings*, and 3) *challenges in design practices and desired features in a prototyping tool*.

3.2 Prototype Process

From experts' responses, we summarized a typical prototype workflow (Figure 2). It consists of three phases: the intra-team communication phase discussing the design problem, requirement, and style; the prototype phase exploring and testing various initial concepts; and the off-team communication phase presenting candidate schemes or solutions to stakeholders. Designers will be free to move

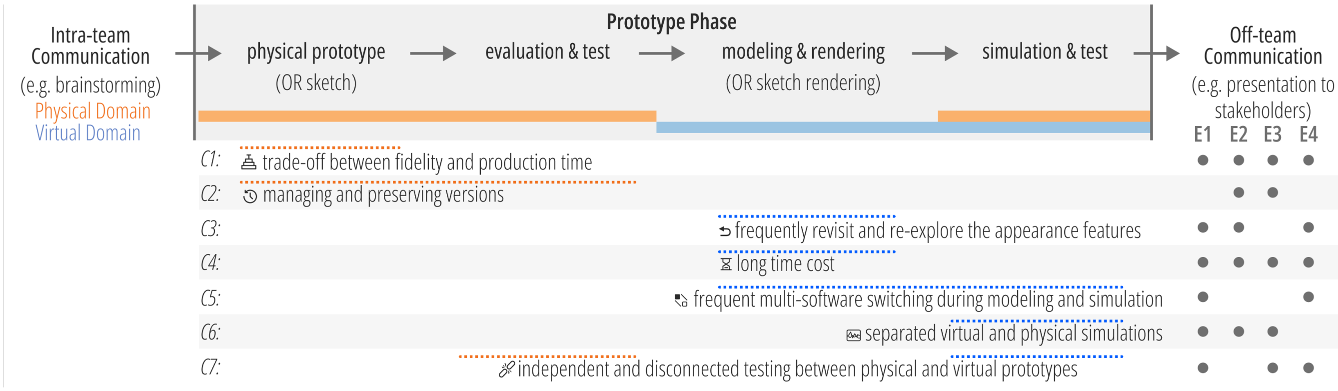

Figure 2: A typical prototype workflow and distinct challenges (C) in product conceptual design, extracted from expert interviews. • indicates the challenge is mentioned by corresponding experts.

forward and backward in these design activities. The three phases might be repeated many times before the next implementation.

We focused on the prototype phase and clarified the designers' operation domain during prototyping in Figure 2. Designers often determine the basic scale and shape through low-fidelity prototypes. They prototype various concepts for further evaluation and tangible structure testing. After multiple iterations, designers select the most suitable low-fidelity prototypes and proceed to the modeling and rendering stage. In this stage, they refine design schemes and produce multi-view renderings for presentation. Additionally, this model information can be used for simulation and testing, such as creating interactive or dynamic simulations for intuitive display. After the prototype phase, the design solution, including prototypes, high-fidelity renderings, and simulations, will be presented to clients or stakeholders for further communication and iteration. Experts also reported they sometimes use sketches or sketch renderings instead of physical prototypes and modeling for more efficient expression. Since this study focuses on 3D design representation, sketches are beyond our discussion scope.

3.3 Existing Design Materials and Challenges

We collected experts' feedback on existing design tools or materials and identified corresponding challenges (Figure 2). During the physical prototype stage, designers usually use shapeable and economical materials, such as foam board, plywood, and clay, for rapid externalization. Sometimes, designers iterate over the ready-made 3D models with these physical materials. A significant challenge is *the trade-off between fidelity and production time* (*C1*). For example, E2 pointed out that "*High-fidelity prototypes often need to invest more manual labor, but the rough prototype is not enough to clearly convey all details of the scheme*". Besides, during the early evaluation & test stage, designers directly interact with tangible models, intuitively evaluating them through observation and touch. *Managing and preserving versions of physical prototypes* (*C2*) presents another challenge. E3 indicated that "*It is difficult to save and share physical versions, especially when it comes to remote cooperation*".

During the modeling and simulation stages, common tools include AutoCAD [6], SolidWorks [24], Rhino [58], Fusion 360 [7], and Grasshopper [59] for modeling; V-Ray [31], KeyShot [50], and Blender [28] for rendering; and Maya [8], Blender [28], Cinema 4D [51], and 3ds Max [5] for simulation. Due to the typically limited details in physical prototypes, designers *frequently need to revisit and re-explore the appearance features during the modeling process* (*C3*). E1 reported that "*During digital modeling, the physical prototype can only provide me with basic design information such as scale and shape. I need to rethink and design when it comes to details*". In addition, other challenges in that stage included the *long-time cost* (*C4*) and *frequent multi-software switching during modeling and simulation* (*C5*). E4 said "*My workflow needs to switch between different software and different design representations repeatedly. I have to go back to the physical prototype to confirm the structural relationship when virtually modeling. I do hope I can work in one platform* ". The *separated virtual and physical simulations* (*C6*) and *the independent and disconnected testing between physical and virtual prototypes* (*C7*) were also challenging. For example, E3 reported that "*I want a tool to help me iterate over parts of a 3D model. With the existing tools, it is difficult for me to split different parts flexibly after modeling a 3D model, but this function is essential in the conceptual design stage. For example, we may want to make this part bigger or change its structure. But with the current modeling tools and GAI tools, it is difficult for me to adjust the parts and keep the whole unchanged flexibly*". E1 also indicated that "*I hope to touch the physical prototype while iterating virtual components. After modeling in the virtual space, I want to feel whether these components are reasonable, such as whether the virtual components can be 'projected' onto physical entities*".

3.4 Design Consideration and Goals

Based on the interviews and extracted challenges in our formative study, we summarized the main design goals that an ideal 3D prototype tool in conceptual design should meet.

- **G1: Allow mixed prototype** (*facing C2 & C7*). It can allow designers to see and utilize physical prototypes and virtual prototypes in a mixed way for seamless prototyping, ideation, and simulation.
- **G2: Offer rapid 3D transition and refinement** (*facing C1, C3, & C4*). It can offer high-fidelity 3D schemes based on preliminary low-fidelity physical prototypes, not just to achieve a 3D reconstruction of the physical prototype but

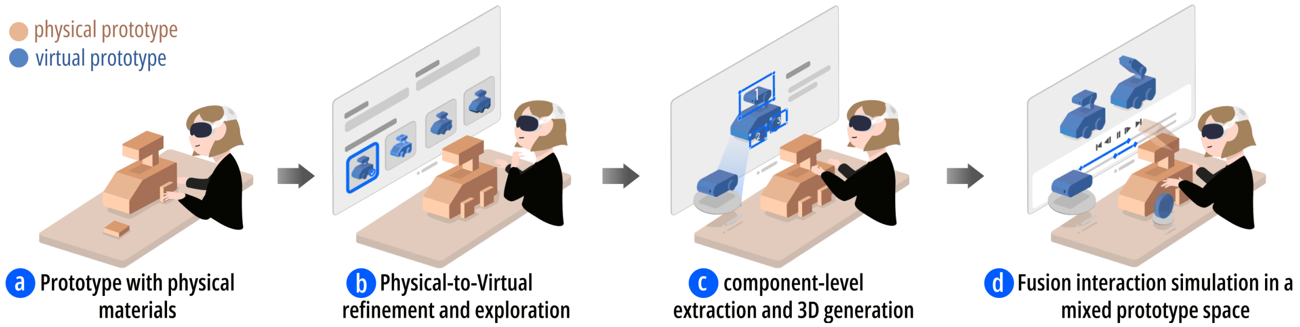

Figure 3: Interaction workflow using FusionProtor in the prototyping process.

to create various high-quality design schemes that meet the design intention within the physical prototype. We subdivide G2 into two sub-goals:
  - **G2.1: Support rapid physical-to-virtual 3D transition**.
  - **G2.2: Support refining low-fidelity prototypes and exploring appearance possibilities**.
- **G3: Support component-level interaction** (*facing C5 & C6*). It can support designers in efficiently verifying the interaction according to the motion logic in physical prototypes after 3D creation, supporting 3D creation and simulation presentation in one platform. We subdivide G3 into two sub-goals:
  - **G3.1: Support component-level generation**.
  - **G3.2: Support components assembly and interaction**.

4 FusionProtor

Following the design goals we identified in the formative study, we developed FusionProtor. It integrated multiple generation models for a component-level physical-to-virtual 3D transition pipeline and employed XR technique to support mixed prototype and dynamic interaction simulation. We designed FusionProtor in Apple Vision Pro that provides mixed work areas and multiple operation windows for complex creative tasks.

4.1 Interaction Flow

We provide a user journey to show FusionProtor's capabilities and interaction flow (Figure 3). A product designer conducted a conceptual design for a *Mobile Detection Robot*. Initially, she built the basic scale and shape with the low-fidelity physical prototype (Figure 3Ⓐ). After sketching local details based on the captured physical prototype and selecting the intended design style, she obtained various generated high-fidelity renderings (Figure 3Ⓑ). She then considered component relationships and extracted individual components from the generated complete scheme, leading to the component-level 3D generation (Figure 3Ⓒ). Next, she coupled virtual 3D components with physical prototypes and simulated their interaction, allowing the mixed component-level ideation (Figure 3Ⓓ). The linear workflow also allows the designers to move back and forth freely during prototyping according to their needs, allowing physical and virtual iteration at any design stage.

4.2 Mixed Prototype in FusionProtor (for *G1*)

4.2.1 Mixed Workspace in FusionProtor. To meet designers' needs for seamlessly transitioning between physical and virtual prototyping, and by leveraging the XR interaction capabilities provided by Apple Vision Pro, we incorporate three working areas in FusionProtor (Figure 4Ⓐ). The *Physical Prototype Work Area* offers various physical materials and tools for tangible creation in the real world. The *Prototype Generation Work Area* shows virtual prototypes corresponding to physical prototypes. Designers can directly drag the virtual 3D prototype into the physical space and mix them with physical prototypes. In addition, if designers aim to simulate complex motion or make simulation animation to further refine the product's structure and behavior, they can use the *Simulation and Animation Work Area*, an optional work area in FusionProtor. FusionProtor's main user interfaces are presented in Figure 4, including capturing and sketching based on physical prototypes (Figure 4(B1)), controlling generation (Figure 4(B2)), extracting components (Figure 4(B3)), browsing generated component (Figure 4(B4)).

4.2.2 Mixed Interaction in FusionProtor. We designed multiple interaction modes in FusionProtor to fully leverage XR capabilities to support mixed prototyping, as shown in Figure 5. First, FusionProtor supports the physical & virtual comparison. In that mode, designers compare different representations (Figure 5(A3)), modify physical model's details according to generated virtual prototype (Figure 5(A2), and make virtual animation according to physical model's motion logic (Figure 5(A1)). Second, FusionProtor mixes the virtual and physical components. It can augment the generated component on the physical prototype (Figure 5(B1), (B2)) or physical environment (Figure 5(B3)). In that mode, designers can couple physical and virtual components tightly, replace and evaluate different candidate components, and even move virtual components based on physical prototypes to simulate motion. Third, with the Apple Vision Pro's support, FusionProtor can realize the full virtual scheme presentation (Figure 5Ⓒ). In that mode, designers can change the virtual background according to the design object, as well as present and evaluate their prototype scheme in an immersive space, taking the prototype beyond the design studio's limitation.

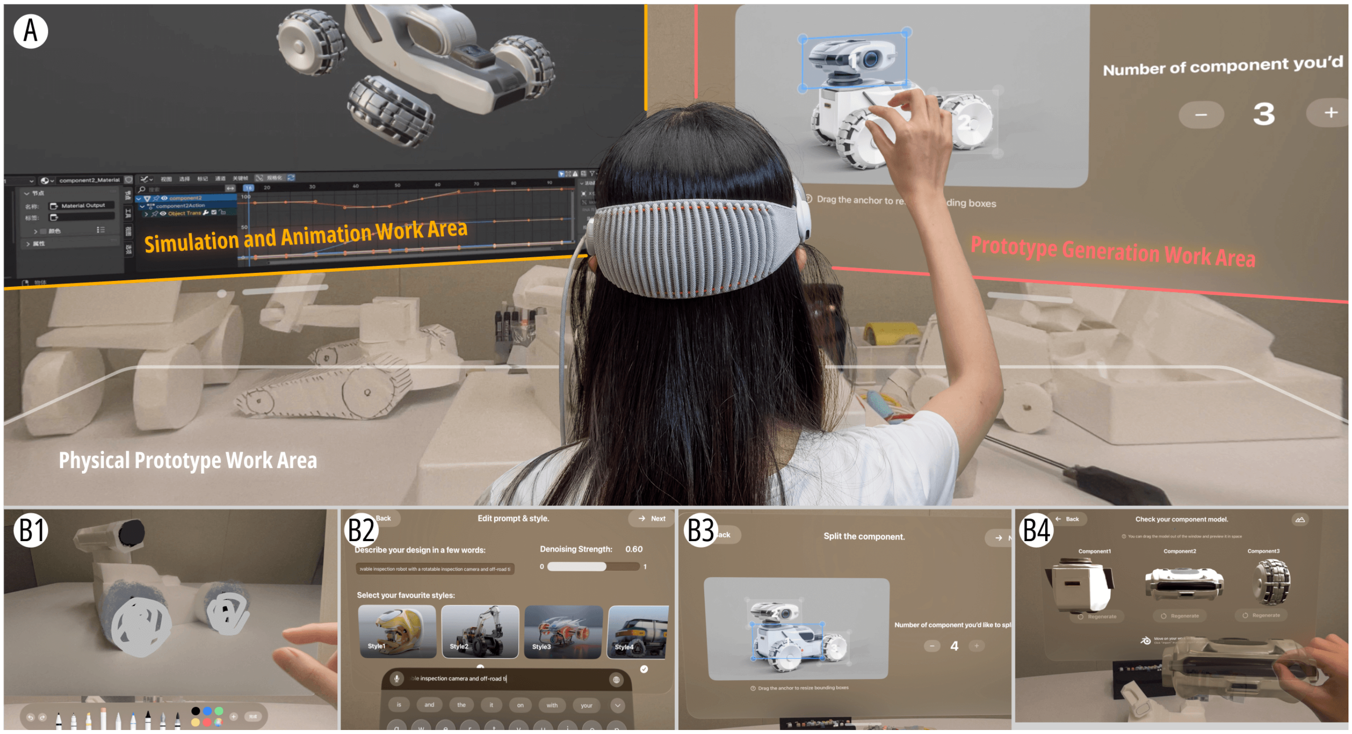

Figure 4: FusionProtor's mixed workspace (Ⓐ) and main user interfaces (Ⓑ). B1: capturing and sketching based on a physical prototype. B2: typing textual description and choosing reference style for generation. B3: browsing generated schemes and extracting components. B4: browsing generated 3D components.

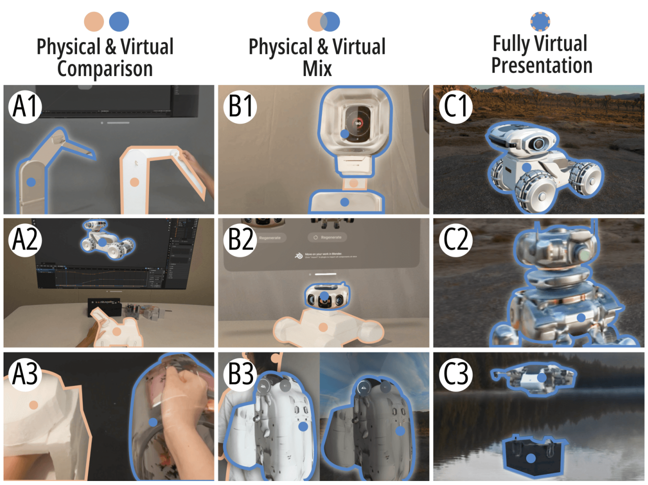

Figure 5: Designed mixed interaction modes in FusionProtor.

4.3 Rapid Physical-to-Virtual 3D Transition (for *G2.1*)

We proposed a step-by-step physical-to-virtual transition pipeline built on GAI technologies and multiple design representations (Figure 6). This pipeline accepts three types of input: an image of the physical model, a textual description of current design, and style-representative images. First, an image-to-image model integrated with an IP-Adapter merges these inputs to produce a high-fidelity, detail-rich image. Next, to enable component-level interaction, we introduce the Component-level Extraction and Generation (ComEG) method. ComEG extracts individual components from the generated image and uses an image-to-3D model to produce corresponding 3D models, allowing for flexible assembly. For designers who require dynamic testing, we provide an optional module integrating various 3D creation tools that support simulation and animation. Our pipeline is model-agnostic, allowing replacement of the image-to-image and image-to-3D generation models, as well as the 3D creation tools, with alternatives more familiar to designers. We released the plug-and-play pipeline and provided its application method in Appendix C. In the implemented pipeline of FusionProtor, the image-to-image translation and component extraction were run on a local GPU, it takes about 40 second for image-to-image translation and 1 min for component extraction. The image-to-3D translation was run on cloud service, it takes about 1 min for geometry and texture generation.

4.4 Prototype Refinement and Fidelity Optimization (for *G2.2*)

We developed *supplementary sketch* and *style transfer* functions (showcased in Figure 7) to support designers in adding details and controlling the style of prototypes, enabling prototype refinement and

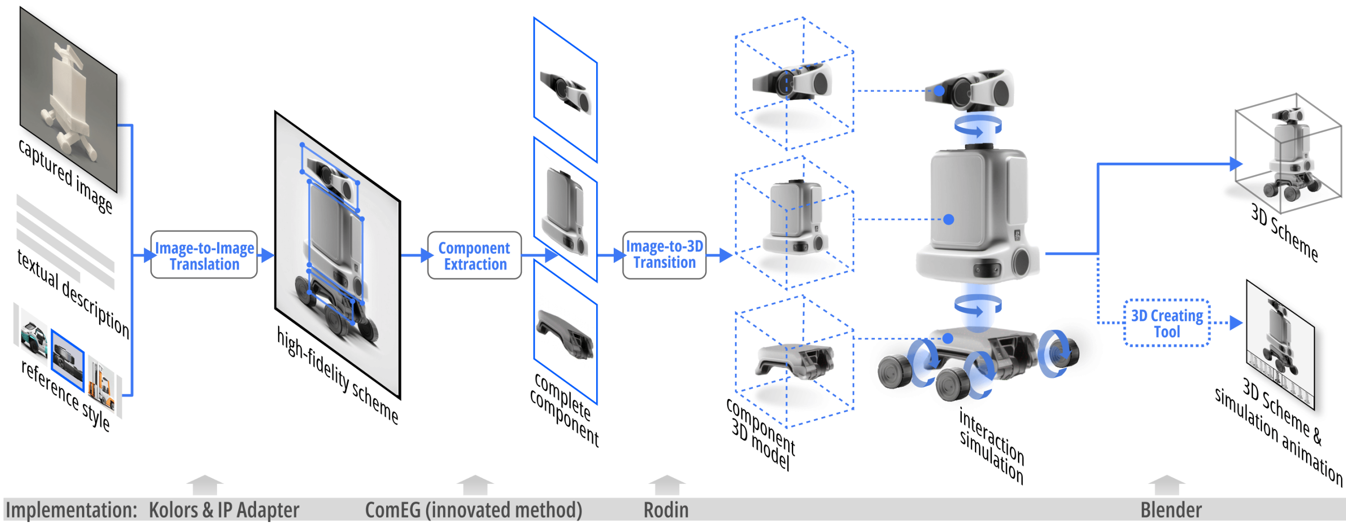

Figure 6: Overall generation pipeline of the rapid physical-to-virtual 3D transition.

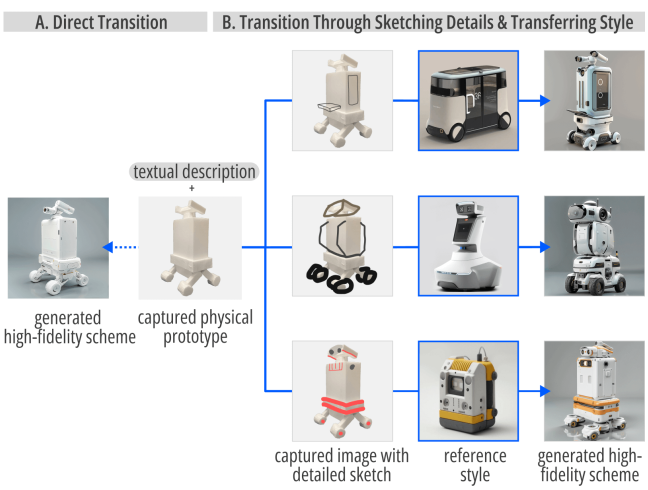

Figure 7: The function of prototype refinement in FusionProtor (B) and direct generation without this function for comparison (A).

optimization of fidelity. Specifically, The *supplementary sketch* function allows designers to add additional details to the captured image of the physical prototype. This eliminates the need for designers to construct every detail using physical materials, which can be time-consuming and challenging. In the style control module, designers can select the style based on their requirements or aesthetic preferences. We offer a variety of style reference images as candidates and also support users in uploading their own.

4.5 Component-level Extraction and Generation (for *G3.1*)

Component-level generation is a prerequisite for virtual component-level prototyping. However, obtaining individual components is challenging due to the complex spatial relationships often present in product design schemes. As Figure 8 shows, directly applying segmentation algorithms such as SAM [42] often results in components with "hole". In addition, current inpainting algorithms, such as SDXL-Inpainting [54], ControlNet-Inpainting [83], and PowerPaint [84] often introduce artifacts, or fails to separate components independently. To address these issues, we need an approach capable of both extracting components and "imagining" their obscured parts to achieve complete, artifact-free components. This ability is critical for generating 3D components in subsequent steps. Against this background, we innovated a Component-level Extraction and Generation (ComEG) method that integrates inpainting and completion techniques. Designers can specify different areas of a design scheme as individual components by simply drawing bounding boxes. The method then automatically extracts and generates the corresponding components based on their positions.

As illustrated in Figure 9, ComEG includes two phases. In Phase1, given an input image I and a designer-specified 2D bounding box b, we utilize SAM [42] to obtain the segmentation masks of the target component M_{target} and the entire product M_{entire}. Then we can obtain the add mask $M_{add} = M_{entire} - M_{target}$ for the regions that require inpainting and completion.

In Phase2, we perform inpainting and completion for the target extracted component by injecting the features of the target region M_{target} into the occluded or incomplete area M_{add}, enabling precise restoration and seamless integration of the missing content. Initially, the original image I is encoded into the latent representation z_0^{input}. Then DDIM inversion [65] is applied to transform z_0^{input} into the corresponding noisy latent vector z_T. Inspired by DesignEdit [37], then we integrate a self-attention mechanism into the decoder within the U-Net denoiser to inpaint and complete the area designated by M_{add} for the occluded regions of the extraction component. This approach ensures the preservation of the component's original integrity and consistency. In the self-attention layers of the U-Net denoiser, we extract the $1 - M_{add}$ region of

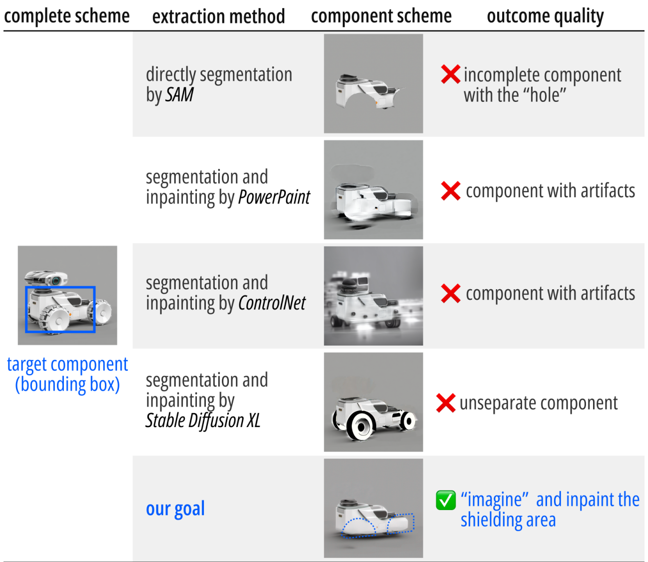

Figure 8: The necessity of innovation of component extraction method.

Algorithm 1 Component-level Extraction and Generation

```
Input: input image I, bounding box b
Output: extracted component image Output
M_entire, M_target, M_add ← SAM(I, b)   ▷ Generate masks for the entire image, target region, and additive region
z_0^input ← Encoder(I)   ▷ Encode the input image
z_T ← DDIMInversion(z_0^input)   ▷ Initialize latent representation
for t = T, T − 1, . . . , 1 do   ▷ Perform T-step denoising
    {Q, K, V} ← U-Net(z_t, t)   ▷ Project z_t onto Q, K, V
    if in the self-attention layer of the U-Net denoiser then
        {Q, K, V} ← {M_target ⊙ Q, M_target ⊙ K, V}   ▷ Apply target mask and add mask to Q and K respectively
    end if
    ε_t ← ε_θ(z_t, t; {Q, K, V})   ▷ Predict the noise at time step t
    z_{t−1} ← Sampler(z_t, ε_t)   ▷ Sample the next latent from noise prediction
    z_{t−1} ← z_{t−1} ⊙ M_add + z_T ⊙ (1 − M_add)   ▷ Update latent to retain component features
end for
Output ← Decoder(z_0)   ▷ Decode latent representation into output image
return Output
```

key features K and the M_{target} region of query features Q during the T denoising steps. Subsequently, each component is directly reconstructed using an image-to-3D method to produce a single 3D model. The whole method with Algorithm 1 is presented in Appendix B.

Additionally, designers can draw multiple bounding boxes at once to precisely control the extraction of different components. When multiple bounding boxes are present, the aforementioned process is repeated for each box. Moreover, there are cases where the component to be segmented does not experience any occlusion. In such instances, the SAM method can be directly utilized. Consequently, in FusionProtor, after drawing the bounding boxes, designers can obtain components segmented directly by the SAM, as well as those processed using our approach. Designers can then select the satisfied components based on their specific needs to proceed with subsequent 3D component generation.

4.6 Interaction Simulation and Animation (for *G3.2*)

We integrated mainstream 3D Creating Tools in FusionProtor to support designers realize complex motion and animation presentation through manual simple assembly and simulation. We adopted this approach for two main reasons. First, 3D interactive simulation and animation typically involve high operational complexity and steep learning thresholds. In the formative study, experts expressed a preference for working with familiar software and interaction paradigms rather than adapting to brand-new interfaces (E1). Second, precise control of operations such as timing, positioning, and motion paths is generally more achievable with 2D interfaces than with mid-air 3D interactions. While XR technology makes mid-air operations more intuitive, previous studies have shown that achieving precise control remains challenging [4, 16]. To address these considerations, we aim to enable designers to operate accurately in 2D interface of a professional 3D creation tool during animation production, while still displaying the final output in an intuitive 3D space after export.

Following this rationale, we integrated several mainstream 3D Creating Tools, such as Blender [28], Cinema 4D [51], and 3ds Max [5], in FusionProtor as an optional module. These tools runs on a local computer, and their screens are streamed to a workspace on the Apple Vision Pro via screen mirroring. We optimized the implementation for a seamless workflow, including one-click import of 3D components across devices. We developed a Python-based PBR Importer plug-in that automatically binds generated 3D components and their textures received from the back-end server. This setup streamlines the design process, enabling designers to focus on core prototyping and simulation tasks while FusionProtor handles the more complex and non-creative operations. In the following user evaluation study and associated figures, we take Blender in the implemented pipeline, but we also provided alternative 3D Creating Tools (see Appendix C), along with their configuration files and plug-ins for FusionProtor.

4.7 Implementation Details

In our system, we employed Kolors [71], an open-source text-to-image generation model built on latent diffusion, as the image-to-image generation model. Style control was implemented using the IP-Adapter model [81], which embeds target style features into the generated prototype to ensure consistency with the desired design style. For the component-level extraction, we enhanced SDXL-1.0 [54] with DDIM inversion [65]. To produce individual 3D components, we integrated Rodin's API [72] for image-to-3D translation. Rodin is a non-open-source but publicly accessible 3D generation model. Our pipeline is model-agnostic, allowing for the substitution of image-to-image, image-to-3D models, and 3D creation tools with alternatives familiar to designers. We released the plug-and-play pipeline and clarified its application in Appendix C. All implementations were conducted on RTX 3090 GPUs.

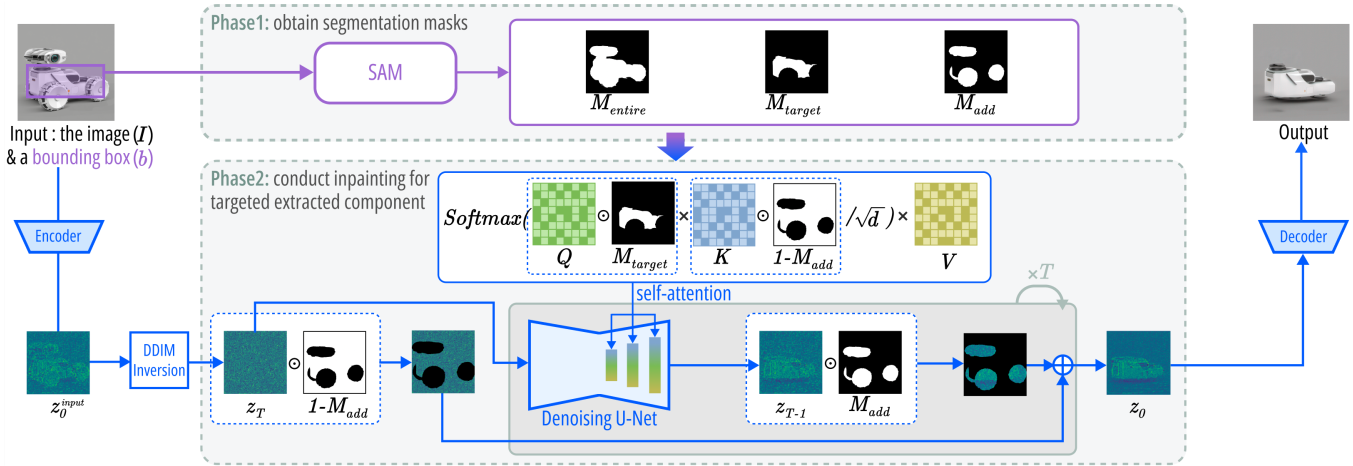

Figure 9: Technical framework of the proposed Component-level Extraction and Generation (ComEG) method.

5 Technical Evaluation

We conducted a technical evaluation to verify the technical performance. Our technical innovation in FusionProtor centers on two aspects: *a component extraction method for product design tasks* and *a 2D-to-3D transition method integrating component-level generation and assembly.* This led us to two technical research questions: **TRQ1**: *Can the proposed extraction method effectively extract product components?* **TRQ2**: *Does the proposed 2D-to-3D transition method produce higher-quality 3D models compared to direct 3D generation without component-level assembly ?*

5.1 Experimental Setups

Datasets. As produced schemes in FusionProtor require designer involvement, such as building physical prototypes and assembling components, we used schemes produced in our user study (described in Section 6) for technical evaluation. Specifically, for TRQ1, we randomly selected 20 2D complete schemes generated by FusionProtor integrated Kolors [71] as a test dataset. Each 2D scheme has a foreground mask chosen by designers through the bounding box and generated by SAM [42]. For TRQ2, we randomly selected 20 2D complete schemes that did not overlap with the dataset in TRQ1.

Baselines. For TRQ1, we compared our method with three mainstream image editing models, SDXL-Inpainting [54], ControlNet-Inpainting [83], and PowerPaint [84]. To answer TRQ2, our method involves manual participation, making it unsuitable and unfair for direct comparison with conventional 3D generation models. However, to verify our method's effectiveness, we compared it with Rodin [72], which we integrated into FusionProtor's technical pipeline.

Metrics. We conducted a human perceptual evaluation. We invited 10 users to compare our proposed methods with baselines. They were shown complete and extracted component images in TRQ1 and multi-view images of 3D schemes with the 2D input image for TRQ2. They completed a questionnaire to evaluate all datasets. Specifically, for TRQ1, we asked questions from the *image quality*, *extraction quality*, and *inpainting ability* perspectives. The questions involve *Do you think 1) this scheme has high image quality and fewer artifacts? 2) this scheme has been completely and individually extracted from the complete product? 3) the shielding area of the product has been imagined and inpainted in this component image?* For TRQ2, we asked questions from the *geometric quality*, *texture quality*, and *different view consistency* perspectives The questions involve *Do you think 1) this model has high geometric quality? 2) the details of the texture map have high quality? 3) this model has high consistency from different views?* An example questionnaire is presented in Appendix D. The answers include YES and NO. All data sets were presented randomly. We will report the average proportion of YES answers in each question, which represents users' recognition of each question.

5.2 Results

Qualitative Comparison. For TRQ1, as shown in Figure 10, our extraction method can accurately extract the target component with fewer artifacts and inpaint the shielding areas between multiple components. Other methods often struggle to isolate high-quality and complete components or generate various additional unreasonable content. For example, in the 4th row and 5th column of Figure 10, GAI adds irrelevant content. In the 3rd row and 6th column, GAI generates more artifacts.

For TRQ2, as shown in Figure 11, our 2D-to-3D transition method can produce a 3D scheme with high-quality details. For example, in the 3rd row, GAI generates better camera detail due to component-level generation. Because the image-to-3D model deals with each component rather than the whole product. Besides, due to the manual assembly participation, the final outcome had better structural relationships. For example, in the 2nd row, the structural relationship of components is better expressed than direct image-to-3D translation.

Quantitative Comparison. The questionnaire results are shown in Figure 12. Following the Shapiro-Wilk test and Levene's test were run for the normality test and variance homogeneity, we used t-test to compare our YES proportion with all baselines. For TRQ1, our method had significant advantages in *image quality*, *component*

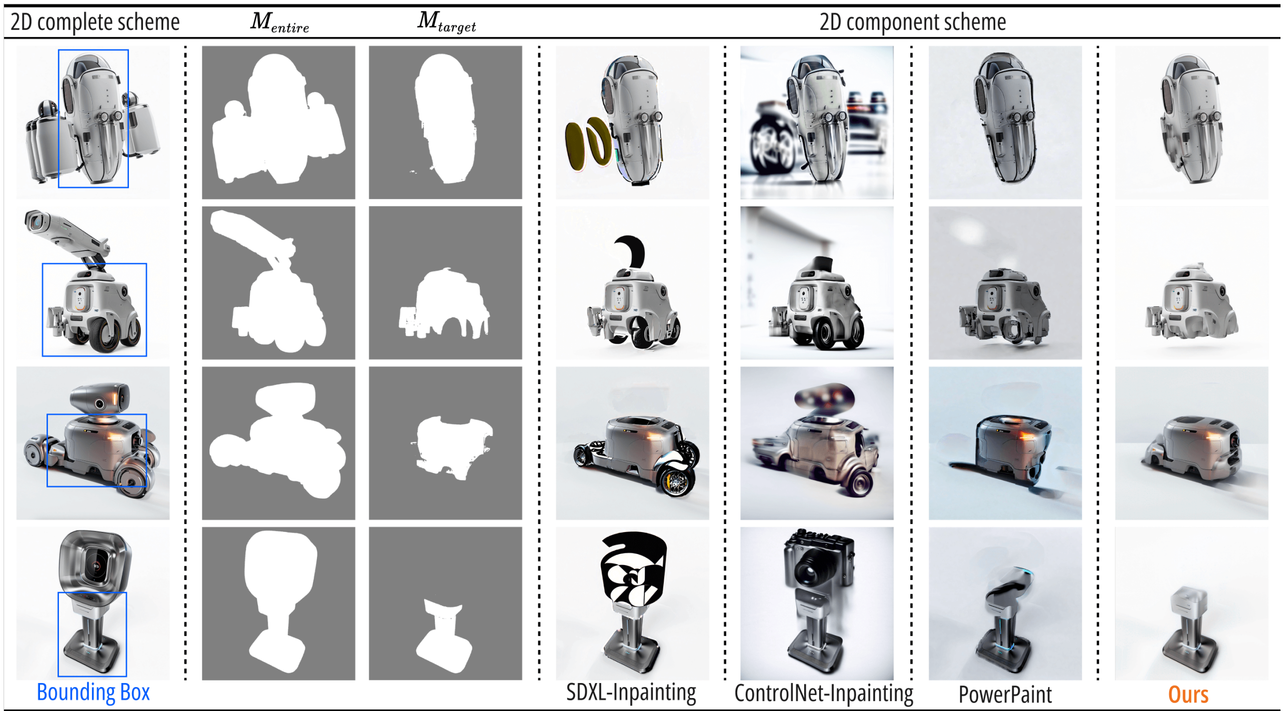

Figure 10: Visual comparison for our component extraction method.

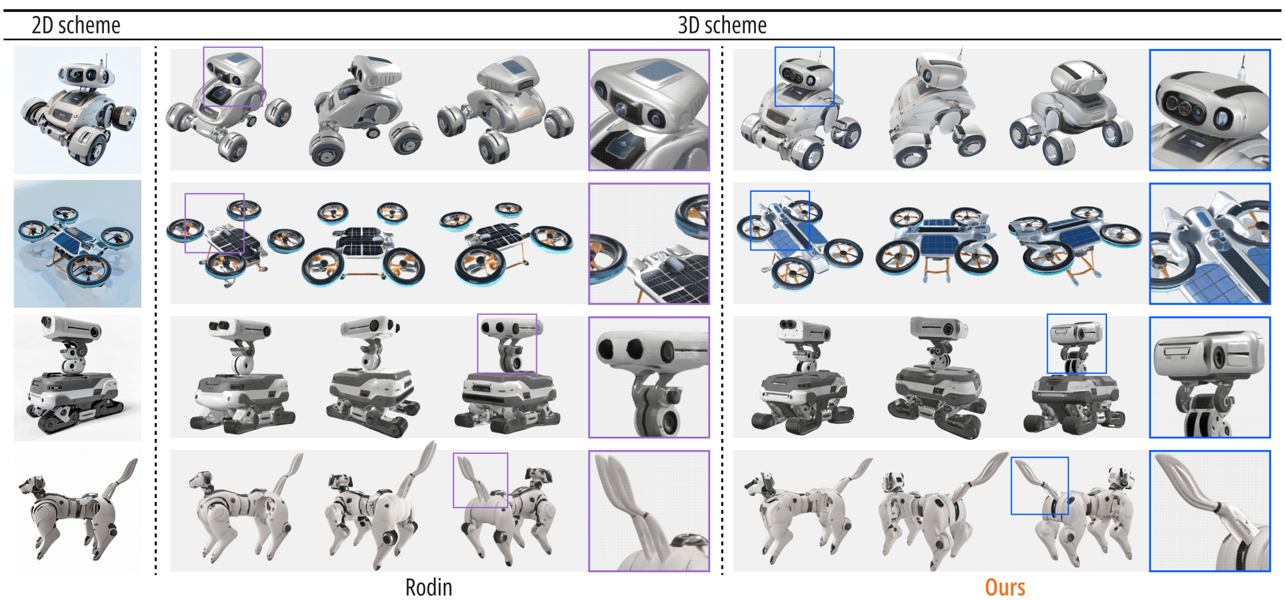

Figure 11: Visual comparison for our 2D-to-3D transition method integrating component-level generation and assembly.

extraction, and *inpainting for the shielding area* compared to SDXL-Inpainting (Q1: $t = 5.80, p < 0.001$, Q2: $t = 7.74, p < 0.001$, Q3: $t = 14.34, p < 0.001$), ControlNet-Inpainting (Q1: $t = 13.46, p < 0.001$, Q2: $t = 18.91, p < 0.001$, Q3: $t = 17.65, p < 0.001$), and PowerPaint (Q1: $t = 3.67, p < 0.001$, Q2: $t = 2.70, p = 0.010$, Q3: $t = 8.97, p < 0.001$). For the TRQ2, our method had significant advantages in the *texture map details* than direct 3D generation without component-level extraction and generation (Q5: $t = 4.84, p < 0.001$).

Component Extraction Task (for TRQ1)	SDXL-Inpainting	ControlNet-Inpainting	PowerPaint	Ours
Q1: this scheme has high image quality and less artifacts:	62.00% (SD=14.35)	29.00% (SD=13.75)	73.00% (SD=11.87)	**88.00% (SD=13.27)**
Q2: this scheme has been completely and individually extracted from the complete product:	56.50% (SD=18.51)	26.50% (SD=13.14)	82.00% (SD=19.39)	**89.50% (SD=12.03)**
Q3: the shielding area has been imagined and inpainted in this component image:	30.50% (SD=17.74)	12.50% (SD=13.37)	54.00% (SD=15.62)	**88.50% (SD=8.53)**
2D-to-3D Transition Task (for TRQ2)			Rodin	Ours
Q4: this model has high geometric quality:			84.50% (SD=8.65)	86.50% (SD=7.26)
Q5: the details of the texture map have high quality:			72.00% (SD=8.12)	**87.00% (SD=5.57)**
Q6: this model has high consistency from different views:			86.50% (SD=10.62)	85.00% (SD=6.71)

Figure 12: The proportion of recognition answers in questionnaire results. Bold indicates the statistical difference.

6 User Evaluation Study

We address following research questions in user evaluation study.

RQ1: Is FusionProtor usable? Can FusionProtor support diverse product design tasks?
RQ2: What is the influence of FusionProtor on design and creativity in conceptual design? What are the distinct advantages compared with traditional prototype tools?
RQ3: How do designers use FusionProtor for mixed ideation? What are the mixed workflow and interaction supported by FusionProtor?

6.1 Participants

We recruited 16 professional industrial designers (10 males and 6 females, an average age of 27) with at least three years of design experience. The detailed information is shown in Appendix A. To avoid the influence of participants' familiarity with the system, all participants in the user evaluation study were newly enrolled and had not participated in any prior research related to this study, including formative study and technical evaluation. They can skillfully use modeling software, especially Blender, and have experience in designing with GAI tools (e.g., Midjourney [21], Stable Diffusion [60], and DALL-E [56]). Given that designers were required to wear Apple Vision Pro to complete design tasks, we recruited participants with normal vision or those who wore contact lenses for the experiment. Participants in this study were recruited through online publicity and design forums and screened through a registration questionnaire. All participants signed a consent form approved by our institution, and there were no other ethical or privacy impacts. All participants were compensated according to the length of their participation.

6.2 Tasks

We set two design tasks in the user evaluation study. **Task 1** involves a practical design assignment, *a Mobile Robot for Detection*, sourced from an industrial design company. Its design requirements basically include *a compound robot*, *an intelligent detection device*, and *a movable chassis*. A design problem detailing the target scenario and exact expectations was created by the design company and provided to all participants. Designers were encouraged to explore as many concepts as possible based on the basic design requirements, extensively exploring the possibilities of function, structure, and appearance. The completion time for Task 1 was set at 50 minutes, after which designers submitted a complete design solution, including renderings, 3D models, and descriptions.

Task 2 is an open-ended design task to understand design diversity using FusionProtor. Designers were allowed to prototype freely without specific requirements but were urged to innovate across concept, function, structure, and appearance. We provided five potential themes for reference: the *intelligent appliance*, *robot*, *drone*, *mechanical vehicle*, and *wearable device*, though designers were not restricted to these themes. The task required designers to submit at least one design scheme within 30-50 minutes.

6.3 Procedure

The whole procedure consists of three parts. First, we introduced the orientation session and allowed designers to familiarize themselves with basic functions and operations, as well as the Think-aloud method. Second, participants were asked to complete two design tasks in the design process. We suggested all participants complete Task 1 in about 50 min, and complete Task 2 in 30-50 min. If participants are still not finished within 50 minutes, we will urge them to finish as soon as possible. The order of the two tasks was counterbalanced among participants. The use of the Think-aloud method was a requisite during the design process. Following the design process, participants were asked to complete several questionnaires and participate in interviews.

6.4 Measurement

As summarized in Table 1, we combined quantitative and qualitative analysis to answer three RQs. We invited three design leaders from a design company to evaluate the design outcome of Task 1 for RQ1. They were tasked with evaluating design solutions for practical conceptual design based on completeness, which represents whether the scheme is complete enough and whether its quality can be displayed to stakeholders. The *Design Completeness Rate* was calculated as the ratio of complete solutions to total solutions. Given the subjective nature of completeness evaluation, three design leaders independently assessed all design outcomes, and an average rate was reported.

Several questionnaires were collected. Specifically, the System Usability Scale (SUS) [14] was applied to evaluate FusionProtor's usability for RQ1. NASA Task Load Index (TLX) [33] was adopted to assess the workload of the design process for RQ2. It is an overall workload score based on weighted average ratings on *mental demand*, *physical demand*, *temporal demand*, *effort, performance*, and

Table 1: Measurement summary in the user study. □ indicates the quantitative analysis while ■ indicates qualitative analysis.

RQ	Standpoint	Research Method	Metrics
RQ1	system usability	□ SUS	SUS score
	design outcome	□ expert evaluation	Design Completeness Rate
	diverse support	■ showcase	/
RQ2	creativity support	□ CSI	CSI score, Collaboration, Enjoyment, Exploration, Expressiveness, Immersion, Results Worth Effort
	design workload	□ NASA TLX	TLX score, Mental Demand, Physical Demand, Temporal Demand, Effort, Performance, Frustration Level
	strength and limitation	■ semi-structured interview by thematic analysis	/
RQ3	ideation mode	■ behavior analysis by Think-aloud protocol	/
	prototype interaction		

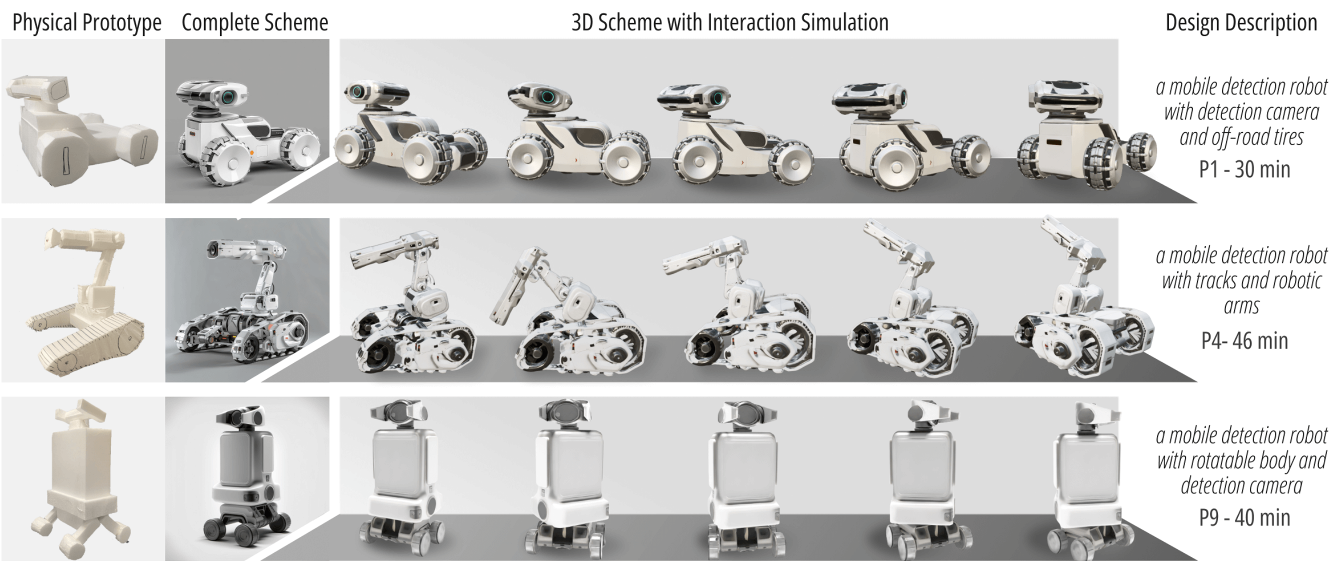

Figure 13: The presentation of Task 1's design outcomes under FusionProtor's support.

frustration level. Besides, the Creativity Support Index (CSI) [17] was utilized to evaluate the creativity support for RQ2. It measures six dimensions of creativity support: *collaboration*, *enjoyment*, *exploration*, *expressiveness*, *immersion*, and *results worth effort*. Similar to previous studies [57, 82], the *collaboration* index quantified the level of human-AI cooperation instead of human-human cooperation.

All participants were invited to semi-structured interviews for RQ2. The interview focused on four key issues, including *1) comparison with traditional prototype tools*, *2) GAI cooperation in design*, *3) ideation with the mixed-prototype*, *4) user experience and faced challenges*. Three researchers used thematic analysis [70] to extract commonly-mentioned codes. They independently coded all raw data and shared their codes, resolving disagreements and merging similar codes until they reached a consensus.

We adopted the Think-aloud protocol [3, 39] to answer RQ3 during the prototype process. Three researchers reviewed participants' design process. We focused on *1) how designers transform between the virtual and physical domain in design process* and *2) how designers use FusionProtor for mixed prototypes*.

Since FusionProtor introduces novel working modes and processes, using existing design tools as a baseline for comparison is both challenging and unfair. However, we still considered performance comparisons. First, from the design outcome perspective, design experts from the industry evaluated the outcomes for completeness against their practical standards. Second, from the design process perspective, participants were asked to compare FusionProtor's advantages and disadvantages with their traditional tools and processes according to their own design experience.

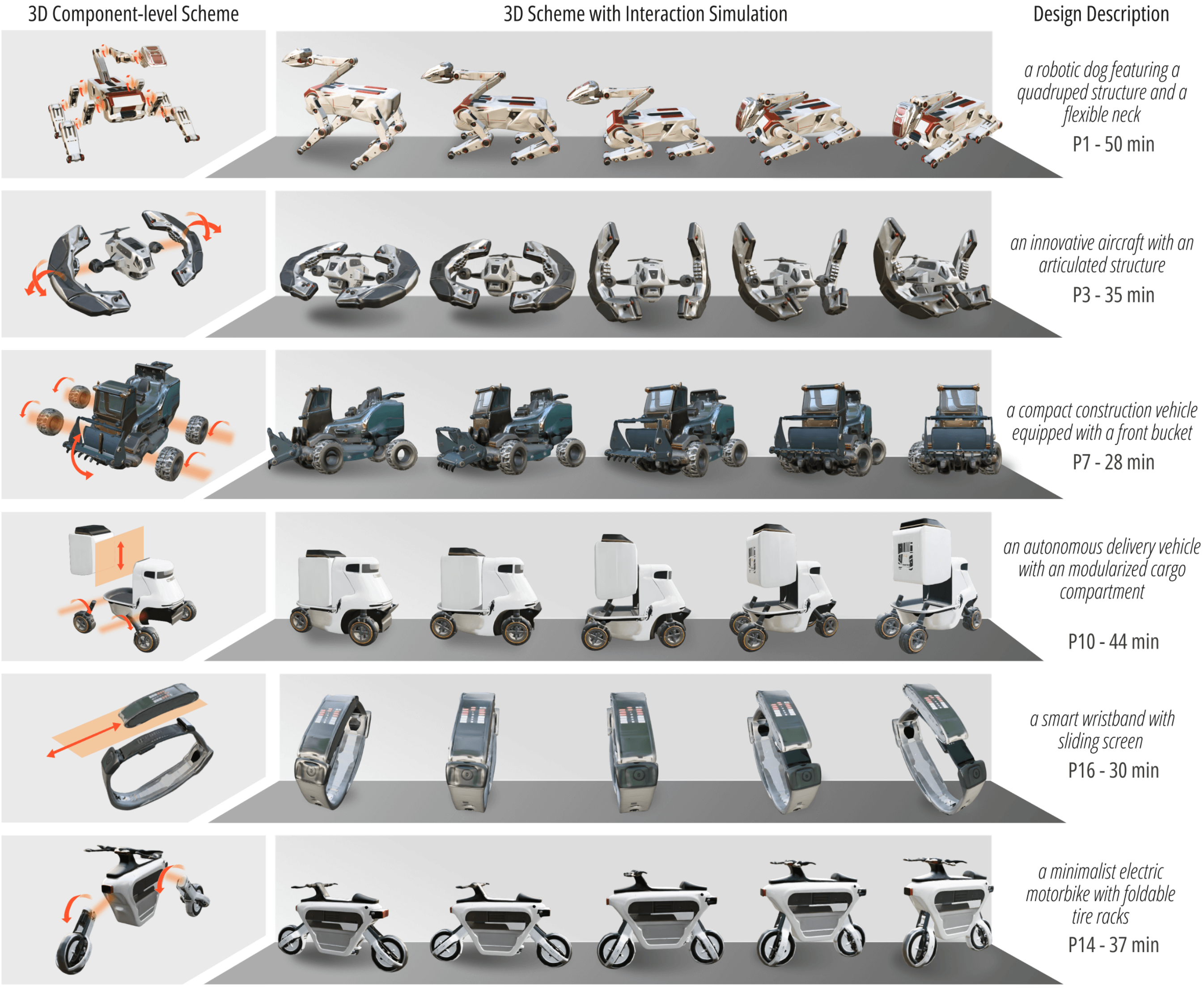

Figure 14: The presentation of Task 2's diverse design outcomes under FusionProtor's support.

6.5 Results and Findings

6.5.1 FusionProtor Is Highly Usable (for RQ1). ***FusionProtor has high usability.*** We collected 33 (16 for Task1 & 17 for Task2) design schemes during the user study. The *average prototyping duration* was 40.82 (SD = 7.52) min, with Task 1 lasting 43.00(SD = 5.73) min and Task 2 lasting 38.76(SD = 8.37) min. The *SUS score* of FusionProtor was 87.81 (SD = 11.59) , which was rated as "*acceptable*" and "*excellent*" according to Bangor's standard [9].

FusionProtor reaches the practical conceptual design standard. Figure 13 shows some randomly-selected schemes from Task 1. Three design leaders from a design company reviewed the design schemes of Task 1 according to their practical standards and evaluated the completeness of design outcomes. The average *Design Completeness Rate* of the three experts was 91.67% (SD = 2.95). Experts were surprised that participants could clearly complete the design schemes in such a short timeframe and deemed the quality sufficient to effectively apply in design practices.

FusionProtor supports diverse design. Figure 14 shows some randomly-selected schemes from Task 2. In Task 2, designers used FusionProtor to explore diverse design concepts, transition components from physical to virtual, refine structures, and simulate dynamic interaction. Participants reported that FusionProtor had a wide application scope, enabling them to accomplish diverse conceptual design tasks efficiently.

6.5.2 FusionProtor Presents Distinct Advantages in Conceptual Design (for RQ2). The CSI and TLX results are shown in Figure 15ⒶⒷ. The *CSI score* was 88.79(SD = 1.76), which indicated FusionProtor's excellent support for creative work. CSI results demonstrated FusionProtor's distinct advantages in *expressiveness* and *result worth effort*. In addition, the TLX score was 20.38(SD = 4.75), which

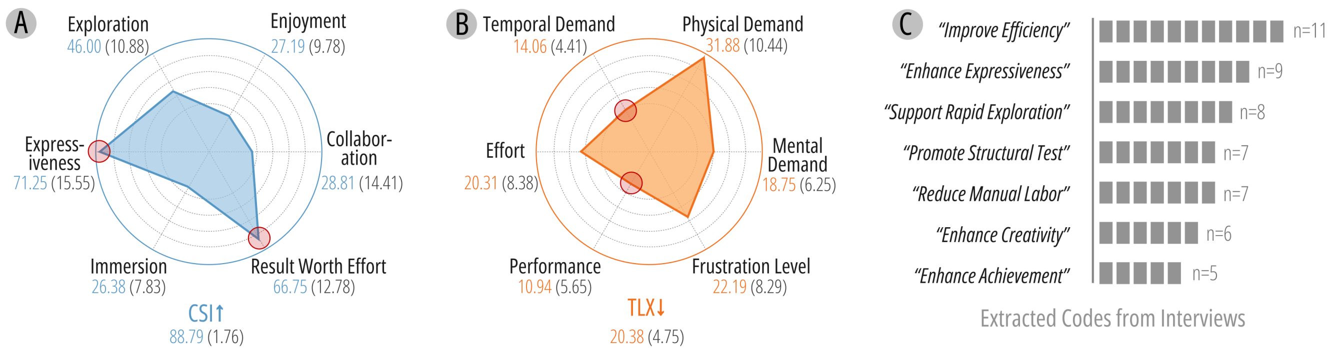

Figure 15: The CSI (A) and TLX (B) results, ↑ indicates higher value is better while↓ indicates lower is better. C shows the extracted codes from interviews.

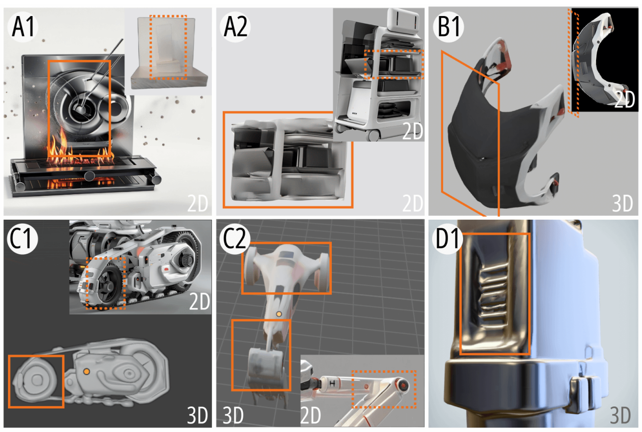

Figure 16: Problematic generated schemes in user study for critical discussion.

indicated that FusionProtor could help designers easily complete conceptual design tasks. TLX results highlighted FusionProtor's characteristics in low *temporal demand* and high *performance*.

In addition to quantitative results, we extracted and summarized commonly-mentioned codes related to the FusionProtor's strength in Figure 15C. The most frequently mentioned strength was "*Improve Efficiency*". Designers reported that "*it will take me about a week to model these schemes using traditional methods*" (P11) and "*FusionProtor integrated some design tools in one platform, allowing me to rapidly prototype seamlessly*" (P1). "*Enhance Expressiveness*" was another distinct strength. P2 indicated that "*FusionProtor enriched many details based on my low-fidelity physical model by integrating GAI capabilities*". Similarly, eight participants reported that FusionProtor can support design exploration. Specifically, P4 mentioned that "*GAI's participation has given me lots of inspiration, including details, structures, functions, and appearance styles*". Due to the GAI integration, designers considered that FusionProtor could take over the tedious expression tasks, reducing their manual labor for the prototype in the traditional design process. In addition, seven participants reported that the physical prototype and XR participation promoted their structural thinking and supported tangible tests.

6.5.3 Dissatisfied Cases and Mentioned Limitations (for RQ1&2). In addition to summarizing FusionProtor's strengths, we also critically reported the limitations and new challenges through the analysis of the design process and interview.

Dissatisfied Cases. We concentrated on scheme generation where designers might be dissatisfied or seek regeneration. Specifically, in generating the complete 2D scheme, a single textual description may not produce a result that meets the designer's expectations. And the capturing quality can influence generation quality (Figure 16A1). In extracting 2D components, our user study found that FusionProtor exhibited limitations in extracting components from structures with severe occlusion or tight integration, such as isolating a modular container from a *mobile delivery robot* (Figure 16A2). During the 2D-to-3D transition, the generation quality diminished when designers provided flat and orthographic 2D schemes without any perspective distortion (Figure 16B1).

Mentioned Limitations. We also summarized the commonly-mentioned limitations or challenges during the design process. First, more than half of the participants complained about the physical load in the process of using FusionProtor. P9 reported that "*The weight of HMD makes it difficult for me to apply it in design practice for a long time*". Second, GAI controllability was another widely reported challenge, especially during the image-to-image generation from the captured image to the generated scheme. P3 said that "*Although I can transfer the style of low-fidelity physical models, I don't have fine-grained control during generation, such as local designated generation. This might increase my regeneration times during the image-to-image process*". Third, participants critically discussed that GAI lacks an understanding of design knowledge, especially structural knowledge, leading to low structural rationality of the generated components. P1 said "*GAI can only understand my physical scheme from the appearance perspective and 'imitate my scheme to generate', but cannot understand my structure and logic*". For example, FusionProtor generated 3D components that lack structural coherence and assembly relationships in P4's design (Figure 16C1). Similarly, it generated two connected parts, yet the structural dimensions of the joints differed significantly in P1's design (Figure 16C2). Fourth, designers pointed out that although FusionProtor achieved remarkable 3D quality, these schemes still fall short of being directly applicable to engineering production. P9

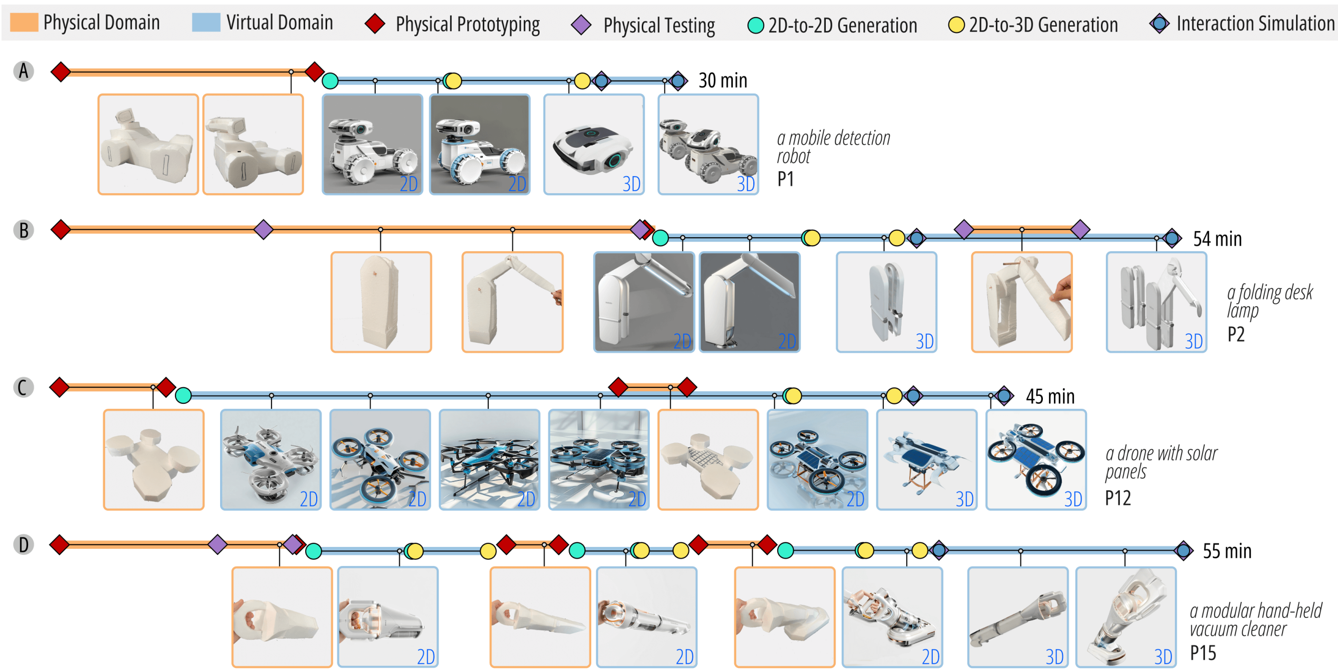

Figure 17: Ideation and prototype process in the user study, extracted through think-aloud and behavior analysis.

reported the imperfect and inaccurate geometry while P10 complained the texture details (Figure 16Ⓓ1). Fifth, some participants mentioned that GAI participation may lead to creativity fixation in the early design stage. We further critically discussed the GAI influence on the creative process in Section 7.3.

6.5.4 FusionProtor Supports Diverse Ideation and Creation Modes (for RQ3). We paid attention to FusionProtor's influence on the ideation process, in which we focused on how designers advance their design process. Although designers need to start with physical prototypes in FusionProtor, it does not mean that designers' thinking must develop linearly from physical to virtual. In the evaluation study, some designers elaborated on physical prototypes first and then transformed them into virtual forms. Some designers built simple physical prototypes at will to find inspiration in the virtual generation process. Some designers advanced their design process through repeated transformations between physical and virtual domains. FusionProtor allowed designers to freely move forth and back, as well as switch between physical and virtual domains.

Figure 17 presents some design processes using FusionProtor. For example, P1's design process was promoted in a linear way without any repeated operations (Figure 17 Ⓐ). This linear mode supported an efficient conceptual design task. Similarly, P2 also progressed linearly but dedicated significant time to constructing and testing physical prototypes (Figure 17 Ⓑ). P2 designed a *foldable desk lamp*, spending half the time on the physical domain to test the folding structure before transiting into the virtual form. On the contrary, P12 spent most of time in the virtual domain, especially on 2D scheme generation (Figure 17 Ⓒ). P12 made a basic physical prototype and then gained various inspirations through repeated generations. P12 reported that "*I can get various inspirations by adjusting the reference styles and parameters in GAI generation, and I constantly adjust my prototype*". In addition, P15's ideation process involved frequent transitions between physical and virtual domains (Figure 17 Ⓓ). Specifically, P15 designed a *modular vacuum cleaner*, prototyping various modular suction nozzles and repeatedly verifying and simulating them in both domains. P15 noted that "*I often verify the shape and structure in the physical world, such as the handle scale and inclination angle, and efficiently realize my ideas in the virtual world with generation*".

7 Discussion

7.1 The Component-level 3D Transition Pipeline on Prototype

In FusionProtor, we proposed a component-level 3D transition pipeline. To implement it, we innovated a component extraction method and integrated existing generative models to release a plug-and-play generation pipeline. First, from the technical perspective, this is the first component-level 3D generation pipeline for the industrial conceptual design field, which can realize high-quality product component generation. Our technical evaluation verified the effectiveness of our innovative component extraction method. Besides, compared to generating a complete 3D directly, our pipeline allowed the Img-to-3D model to deal with each component one by one, which can generate more detailed geometry and high-consistency texture of local parts. Second, from the perspective of the design workflow, our pipeline integrated multiple GAI models and design tools on one platform, providing a seamless prototype workflow. In the user evaluation study, participants commonly mentioned that FusionProtor improved the design efficiency and supported rapid exploration. The joint integration took over tedious operations such as data conversion, tool switching, format compatibility, and model deployment in the traditional design workflow, allowing designers to concentrate on their creative work. Third, from the perspective of creativity, designers pointed out that FusionProtor enabled them to focus on detail iteration because our pipeline emphasized the generation and iteration at the component level. Designers also considered that component-level design promoted their evaluation of component feasibility and adaptability, facilitating the exploration of modular structures and functions.

7.2 The Real-time Mixed Interaction on Prototype

While many HCI studies have used XR technology to create mixed environments for design, most virtual objects are retrieved from databases [36, 52] or obtained by mid-air modeling [67]. The preset artifacts limit the designer's freedom of conception while on-site modeling takes longer time with lower fidelity. FusionProtor integrated GAI in HMD equipment to enable the real-time mixed prototype, allowing designers to flexibly adjust and iterate high-fidelity virtual prototypes immediately. With FusionProtor's support, designers can operate the physical prototype embodied while seeing the high-fidelity virtual feedback in real-time. This brought new opportunities for mixed prototypes. It can support local rapid iteration and adaptive testing. For example, designers can replace different components based on physical prototypes (e.g., Figure 17 Ⓓ). They can even change different scenarios and browse products in different application spaces with the support of XR technology, supporting on-site design and immersive ideation.

7.3 Critical Discussion on the Influence of Using FusionProtor on Creativity

Although FusionProtor received positive feedback from designers in the evaluation study, we critically discussed the newly introduced potential problems related to design creativity and design fixation for reflection. On the one hand, in FusionProtor's design workflow, the high-fidelity prototype has been introduced prematurely during the physical-to-virtual, which might lead to design fixation in the early stage of conceptual design [76]. For example, in P3's design process, GAI randomly refined an ambiguous detail of the physical prototype into a concrete camera, and the camera design always existed in the subsequent prototype iterations. Although in the P3's original design idea, the camera may not exist or it is unnecessary. The rapid generation and iteration reduce designers' reflection on randomly provided high-fidelity details [73]. On the other hand, choosing styles in advance might limit the design possibilities and idea diversity in FusionProtor interaction. Previous studies indicated that designers should consider the overall form of a design before focusing attention on details, such as style, color, and material, to avoid early fixation [23]. Therefore, our workflow and interaction might enable designers to unconsciously focus on low-level design decisions early [82]. In addition, after observing participants' outputs, we found that the style transfer function might foster some sort of convergence towards a similar visual style. We critically pointed out that FusionProtor might lead to style fixation when it as a creativity support tool.

7.4 FusionProtor's Best Practice

Building upon our critical understanding of FusionProtor through the evaluation study, we proposed FusionProtor's best practices. First, we recommend using FusionProtor primarily for conceptual design ideation and presentation, as it is well-suited for high-fidelity presentations and effective stakeholder communication. It is not appropriate for engineering production due to the lack of rational structure understanding and generation. We particularly recommend FusionProtor for prototyping and simulating products with multiple independent components and rotational relationships, such as articulated products. This recommendation is based on our user study, where such products achieved the highest quality and satisfaction ratings. In addition, in order to avoid the influence of shooting quality, designers should avoid capturing physical prototypes from orthogonal angles to enhance the 3D generation quality. Besides, to alleviate the design fixation, we advise designers to input extensive textual descriptions and select diverse reference styles to generate various candidates that align with their intentions for further comparison and ideation instead of a single output.

7.5 FusionProtor's Optimization Space

We proposed FusionProtor's optimization space based on the limitations mentioned by participants. First, referring to HCI studies in the augmented reality domain [36, 52], we plan to couple the motion relationship between physical and virtual components, rather than just fusing them in visual presentation. For example, when designers move the physical prototype of a physical car, its virtual wheel can rotate with it. Second, we hope to give full play to the virtual creation and editing ability supported by XR technology. For example, FusionProtor only supports 2D sketches on the captured image of physical prototypes, and the simulation animations are also created in 2D interface, which limits the designers' mid-air interaction in 3D space. More mid-air interaction capabilities can be updated in FusionProtor, making 3D creation more natural and intuitive. Third, we are considering optimizing the scale and size of 3D genetation. Using multi-view captured images as inputs might make generated models more closely with physical prototype's scale. Besides, providing layout information of complete and component schemes to 3D generation models integrated with ControlNet [83] might allow for the component creation with more proportionate relative sizes.

7.6 Limitation

We discuss the limitations in this study. First, as FusionProtor integrates 3D creating tools to support animation production, it is not a complete desktop application. In the future work, the assembly function or animation simulation function adapted to FusionProtor should be further developed to make it a more full-fledged design software. Second, the integrated generative models can be updated. While we integrated the most advanced models available at the time, more powerful models could be replaced in the FusionProtor's technical pipeline for higher generation quality. Third, more comprehensive methods and quantitative metrics are needed. Comparative analyses with existing tools could serve as baselines to assess FusionProtor's effectiveness. Fourth, this study is limited by its lab setting, with restricted design time and a small participant pool. Future research should involve more practical studies with more complex design requirements and broader participation.

8 Conclusion

We introduced FusionProtor, a mixed-prototype tool for component-level physical-to-virtual 3D transition and simulation in conceptual design. FusionProtor incorporated GAI for rapid component-level generation while leveraging XR to enable designers to mix physical and virtual components for fusion ideation and presentation. A technical study validated the tool's robustness in extracting components and producing high-quality 3D models. A user study confirmed its practical usability in supporting diverse designs. Designers were able to create a satisfactory scheme in an average time of 40.82 minutes, greatly reducing both the time and manual labor of prototype and simulation. Additionally, designers demonstrated FusionProtor's distinct strengths in enhancing expressiveness, facilitating rapid exploration, and fostering creativity. Our findings verified that FusionProtor achieved a seamless workflow from physical to virtual and low- to high-fidelity, enhancing efficiency and promoting ideation. Finally, we critically discussed FusionProtor's best practices and explored the effect of mixed interaction on design for the HCI community.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (2022YFB3303301).

References

[1] Jorge Alcaide-Marzal, Jose Antonio Diego-Mas, and Gonzalo Acosta-Zazueta. 2020. A 3D shape generative method for aesthetic product design. *Design studies* 66 (2020), 144–176.
[2] Celena Alcock, Nathaniel Hudson, and Parmit K Chilana. 2016. Barriers to using, customizing, and Printing 3D designs on thingiverse. In *Proceedings of the 2016 ACM International Conference on Supporting Group Work*. 195–199.
[3] Obead Alhadreti and Pam Mayhew. 2018. Rethinking thinking aloud: A comparison of three think-aloud protocols. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–12.
[4] Rahul Arora, Rubaiat Habib Kazi, Fraser Anderson, Tovi Grossman, Karan Singh, and George W Fitzmaurice. 2017. Experimental Evaluation of Sketching on Surfaces in VR.. In *CHI*, Vol. 17. 5643–5654.
[5] Autodesk Inc. 2024. 3ds Max. Computer Software. https://www.autodesk.com/products/3ds-max
[6] Autodesk Inc. 2024. AutoCAD 2025. Computer Software. https://www.autodesk.com/products/autocad/overview
[7] Autodesk Inc. 2024. Fusion 360. Computer Software. https://www.autodesk.com/products/fusion-360/overview
[8] Autodesk Inc. 2024. Maya. Computer Software. https://www.autodesk.com/products/maya
[9] Aaron Bangor, Philip T Kortum, and James T Miller. 2008. An empirical evaluation of the system usability scale. *Intl. Journal of Human–Computer Interaction* 24, 6 (2008), 574–594.
[10] Qifang Bao, Daniela Faas, and Maria Yang. 2018. Interplay of sketching & prototyping in early stage product design. *International Journal of Design Creativity and Innovation* 6, 3-4 (2018), 146–168.
[11] Loris Barbieri, Agostino Angilica, Fabio Bruno, and Maurizio Muzzupappa. 2013. Mixed prototyping with configurable physical archetype for usability evaluation of product interfaces. *Computers in Industry* 64, 3 (2013), 310–323.
[12] W Beitz, G Pahl, and K Grote. 1996. Engineering design: a systematic approach. *Mrs Bulletin* 71 (1996), 30.
[13] black-forest labs. 2024. *FLUX*. https://github.com/black-forest-labs/flux
[14] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.
[15] Bradley A Camburn, Karen H Sng, K Blake Perez, Kevin Otto, Kristin L Wood, Daniel Jensen, and Richard Crawford. 2015. The way makers prototype: principles of DIY design. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Vol. 57175. American Society of Mechanical Engineers, V007T06A004.
[16] Noémie Chaniaud, Sylvain Fleury, Benjamin Poussard, Olivier Christmann, Thibaut Guitter, and Simon Richir. 2023. Is virtual reality so user-friendly for non-designers in early design activities? Comparing skills needed to traditional sketching versus virtual reality sketching. *Design Science* 9 (2023), e28.
[17] Erin Cherry and Celine Latulipe. 2014. Quantifying the creativity support of digital tools through the creativity support index. *ACM Transactions on Computer-Human Interaction (TOCHI)* 21, 4 (2014), 1–25.
[18] DaEun Choi, Sumin Hong, Jeongeon Park, John Joon Young Chung, and Juho Kim. 2024. CreativeConnect: Supporting Reference Recombination for Graphic Design Ideation with Generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–25.
[19] 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.
[20] Comfyanonymous. 2023. *ComfyUI*. https://github.com/comfyanonymous/ComfyUI
[21] Midjourney Company. 2022. Midjourney. https://www.midjourney.com.
[22] J Corbett and JR Crookall. 1986. Design for economic manufacture. *CIRP Annals* 35, 1 (1986), 93–97.
[23] Amod Damle and Philip J Smith. 2009. Biasing cognitive processes during design: the effects of color. *Design Studies* 30, 5 (2009), 521–540.
[24] Dassault Systèmes. 2024. SolidWorks 2025. Computer Software. https://www.solidworks.com
[25] Kristen M Edwards, Brandon Man, and Faez Ahmed. 2024. Sketch2Prototype: rapid conceptual design exploration and prototyping with generative AI. *Proceedings of the Design Society* 4 (2024), 1989–1998.
[26] Jorgen F Erichsen, Heikki Sjöman, Martin Steinert, and Torgeir Welo. 2021. Protobooth: gathering and analyzing data on prototyping in early-stage engineering design projects by digitally capturing physical prototypes. *AI EDAM* 35, 1 (2021), 65–80.
[27] 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.
[28] Blender Foundation. 2024. Blender. Computer Software. https://www.blender.org
[29] Frederic Gmeiner, Humphrey Yang, Lining Yao, Kenneth Holstein, and Nikolas Martelaro. 2023. Exploring challenges and opportunities to support designers in learning to co-create with AI-based manufacturing design tools. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–20.
[30] Matheus Galvão Gomes, André Ogliari, Rodrigo Bastos Fernandes, and Karuliny Oliveira Marques. 2022. Evaluation of physical models as creative stimuli in conceptual design of products. *Design Studies* 81 (2022), 101119.
[31] Chaos Group. 2022. V-Ray. Computer Software. https://www.chaos.com/vray
[32] Emrecan Gulay and Andrés Lucero. 2019. Integrated workflows: generating feedback between digital and physical realms. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–15.
[33] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.
[34] 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).
[35] Yihan Hou, Manling Yang, Hao Cui, Lei Wang, Jie Xu, and Wei Zeng. 2024. C2Ideas: Supporting Creative Interior Color Design Ideation with a Large Language Model. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–18.
[36] Rahul Jain, Jingyu Shi, Runlin Duan, Zhengzhe Zhu, Xun Qian, and Karthik Ramani. 2023. Ubi-TOUCH: Ubiquitous Tangible Object Utilization through Consistent Hand-object interaction in Augmented Reality. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–18.
[37] Yueru Jia, Yuhui Yuan, Aosong Cheng, Chuke Wang, Ji Li, Huizhu Jia, and Shanghang Zhang. 2024. DesignEdit: Multi-Layered Latent Decomposition and Fusion for Unified & Accurate Image Editing. *arXiv preprint arXiv:2403.14487* (2024).
[38] Qianzhi Jing, Tingting Zhou, Yixin Tsang, Liuqing Chen, Lingyun Sun, Yankun Zhen, and Yichun Du. 2023. Layout generation for various scenarios in mobile shopping applications. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–18.
[39] R Jourdenais. 1995. Does textual enhancement promote noticing? A think-aloud protocol analysis. *Attention and awareness in foreign language learning/Second Language Teaching & Curriculum Center, University of Hawai'i at Manoa* (1995).
[40] Lee Kent, Chris Snider, James Gopsill, and Ben Hicks. 2021. Mixed reality in design prototyping: A systematic review. *Design Studies* 77 (2021), 101046.
[41] Tae Soo Kim, DaEun Choi, Yoonseo Choi, and Juho Kim. 2022. Stylette: Styling the web with natural language. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–17.
[42] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr

Dollár, and Ross Girshick. 2023. Segment Anything. *arXiv:2304.02643* (2023).

[43] Elisa Kwon, Vivek Rao, and Kosa Goucher-Lambert. 2023. Understanding inspiration: Insights into how designers discover inspirational stimuli using an AI-enabled platform. *Design Studies* 88 (2023), 101202.

[44] Carlye Lauff, Jessica Menold, and Kristin L Wood. 2019. Prototyping canvas: Design tool for planning purposeful prototypes. In *Proceedings of the design society: international conference on engineering design*, Vol. 1. Cambridge University Press, 1563–1572.

[45] Glyn Lawson, Paul Herriotts, Louise Malcolm, Katharina Gabrecht, and Setia Hermawati. 2015. The use of virtual reality and physical tools in the development and validation of ease of entry and exit in passenger vehicles. *Applied ergonomics* 48 (2015), 240–251.

[46] Jie Li, Hancheng Cao, Laura Lin, Youyang Hou, Ruihao Zhu, and Abdallah El Ali. 2024. User experience design professionals' perceptions of generative artificial intelligence. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–18.

[47] Chen-Hsuan Lin, Jun Gao, Luming Tang, Towaki Takikawa, Xiaohui Zeng, Xun Huang, Karsten Kreis, Sanja Fidler, Ming-Yu Liu, and Tsung-Yi Lin. 2023. Magic3d: High-resolution text-to-3d content creation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 300–309.

[48] 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.

[49] Longwen Zhang, Ziyu Wang, Qixuan Zhang, Qiwei Qiuand Anqi Pang, Haoran Jiang, Wei Yang, Lan Xu, and Jingyi Yu. 2024. CLAY: A Controllable Large-scale Generative Model for Creating High-quality 3D Assets. *arXiv preprint arXiv:2406.13897* (2024).

[50] Luxion. 2024. KeyShot. Computer Software. https://www.keyshot.com

[51] Maxon. 2024. Cinema 4D. Computer Software. https://www.maxon.net/en/cinema-4d

[52] Kyzyl Monteiro, Ritik Vatsal, Neil Chulpongsatorn, Aman Parnami, and Ryo Suzuki. 2023. Teachable reality: Prototyping tangible augmented reality with everyday objects by leveraging interactive machine teaching. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15.

[53] Huaishu Peng, Jimmy Briggs, Cheng-Yao Wang, Kevin Guo, Joseph Kider, Stefanie Mueller, Patrick Baudisch, and François Guimbretière. 2018. RoMA: Interactive fabrication with augmented reality and a robotic 3D printer. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–12.

[54] Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. 2023. Sdxl: Improving latent diffusion models for high-resolution image synthesis. *arXiv preprint arXiv:2307.01952* (2023).

[55] Ben Poole, Ajay Jain, Jonathan T. Barron, and Ben Mildenhall. 2023. DreamFusion: Text-to-3D using 2D Diffusion. In *The Eleventh International Conference on Learning Representations*. https://openreview.net/forum?id=FjNys5c7VyY

[56] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. 2021. Zero-shot Text-to-image Generation. In *International Conference on Machine Learning (ICML)*. PMLR, 8821–8831.

[57] Jeba Rezwana and Mary Lou Maher. 2022. Understanding user perceptions, collaborative experience and user engagement in different human-AI interaction designs for co-creative systems. In *Proceedings of the 14th Conference on Creativity and Cognition*. 38–48.

[58] Robert McNeel & Associates. 2021. Rhinoceros 3D (Rhino) 8. Computer Software. https://www.rhino3d.com

[59] Robert McNeel & Associates. 2022. Grasshopper for Rhino. Graphical Algorithm Editor, Computer Software. https://www.grasshopper3d.com

[60] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10684–10695.

[61] Yulin Shen, Yifei Shen, Jiawen Cheng, Chutian Jiang, Mingming Fan, and Zeyu Wang. 2024. Neural Canvas: Supporting Scenic Design Prototyping by Integrating 3D Sketching and Generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–18.

[62] Yang Shi, Tian Gao, Xiaohan Jiao, and Nan Cao. 2023. Understanding design collaboration between designers and artificial intelligence: a systematic literature review. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–35.

[63] Yichun Shi, Peng Wang, Jianglong Ye, Mai Long, Kejie Li, and Xiao Yang. 2023. Mvdream: Multi-view diffusion for 3d generation. *arXiv preprint arXiv:2308.16512* (2023).

[64] Dorothé Smit, Bart Hengeveld, Martin Murer, and Manfred Tscheligi. 2022. Hybrid design tools for participatory, embodied sensemaking: an applied framework. In *Proceedings of the Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction*. 1–10.

[65] Jiaming Song, Chenlin Meng, and Stefano Ermon. 2020. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502* (2020).

[66] Francesco Stella, Cosimo Della Santina, and Josie Hughes. 2023. How can LLMs transform the robotic design process? *Nature Machine Intelligence* 5, 6 (2023), 561–564.

[67] Evgeny Stemasov, Simon Demharter, Max Rädler, Jan Gugenheimer, and Enrico Rukzio. 2024. pARam: Leveraging Parametric Design in Extended Reality to Support the Personalization of Artifacts for Personal Fabrication. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–22.

[68] Evgeny Stemasov, Enrico Rukzio, and Jan Gugenheimer. 2021. The road to ubiquitous personal fabrication: Modeling-free instead of increasingly simple. *IEEE Pervasive Computing* 20, 1 (2021), 19–27.

[69] Jingxiang Sun, Bo Zhang, Ruizhi Shao, Lizhen Wang, Wen Liu, Zhenda Xie, and Yebin Liu. 2023. Dreamcraft3d: Hierarchical 3d generation with bootstrapped diffusion prior. *arXiv preprint arXiv:2310.16818* (2023).

[70] Jon Swain. 2018. A hybrid approach to thematic analysis in qualitative research: Using a practical example. *Sage research methods* (2018).

[71] Kolors Team. 2024. Kolors: Effective Training of Diffusion Model for Photorealistic Text-to-Image Synthesis. *arXiv preprint* (2024).

[72] Deemos Tech. 2024. Rodin Gen-1. https://hyperhuman.deemos.com/rodin.

[73] Jakob Tholander and Martin Jonsson. 2023. Design ideation with ai-sketching, thinking and talking with Generative Machine Learning Models. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. 1930–1940.

[74] Dmitry Tochilkin, David Pankratz, Zexiang Liu, Zixuan Huang, Adam Letts, Yangguang Li, Ding Liang, Christian Laforte, Varun Jampani, and Yan-Pei Cao. 2024. Triposr: Fast 3d object reconstruction from a single image. *arXiv preprint arXiv:2403.02151* (2024).

[75] Christine A Toh and Scarlett R Miller. 2014. The impact of example modality and physical interactions on design creativity. *Journal of Mechanical Design* 136, 9 (2014), 091004.

[76] Vimal Viswanathan, Olufunmilola Atilola, Nicole Esposito, and Julie Linsey. 2014. A study on the role of physical models in the mitigation of design fixation. *Journal of Engineering Design* 25, 1-3 (2014), 25–43.

[77] Shishi Xiao, Liangwei Wang, Xiaojuan Ma, and Wei Zeng. 2024. TypeDance: Creating semantic typographic logos from image through personalized generation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–18.

[78] Liwenhan Xie, Zhaoyu Zhou, Kerun Yu, Yun Wang, Huamin Qu, and Siming Chen. 2023. Wakey-Wakey: Animate Text by Mimicking Characters in a GIF. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–14.

[79] Yinghao Xu, Hao Tan, Fujun Luan, Sai Bi, Peng Wang, Jiahao Li, Zifan Shi, Kalyan Sunkavalli, Gordon Wetzstein, Zexiang Xu, et al. 2023. Dmv3d: Denoising multi-view diffusion using 3d large reconstruction model. *arXiv preprint arXiv:2311.09217* (2023).

[80] 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.

[81] Hu Ye, Jun Zhang, Sibo Liu, Xiao Han, and Wei Yang. 2023. Ip-adapter: Text compatible image prompt adapter for text-to-image diffusion models. *arXiv preprint arXiv:2308.06721* (2023).

[82] Chen Zhang, Lin Xie, Li Liu, and Sun You. 2024. ProtoDreamer: A Mixed-prototype Tool Combining Physical Model and Generative AI to Support Conceptual Design. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–18.

[83] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2023. Adding Conditional Control to Text-to-Image Diffusion Models. https://github.com/lllyasviel/ControlNet-v1-1nightly

[84] Junhao Zhuang, Yanhong Zeng, Wenran Liu, Chun Yuan, and Kai Chen. 2023. A task is worth one word: Learning with task prompts for high-quality versatile image inpainting. *arXiv preprint arXiv:2312.03594* (2023).

A Participant Information in the User Study

The expert information in the formative study and the participant information in the evaluation study are shown in Table 2 and Table A, respectively.

Table 2: Overview of experts in the formative study.

ID	Age	Gender	Domain	Design Experience
E1	33	F	Industrial Design	10
E2	32	F	Industrial Design	10
E3	41	M	Computational Design	15
E4	29	M	Design Science	9

Table 3: Overview of participants in the evaluation study.

ID	Age	Gender	Domain	Design Experience
P1	29	M	Industrial Design	6
P2	24	M	Industrial Design	3
P3	24	F	Art Design	3
P4	24	M	Computational Design	3
P5	24	M	Industrial Design	3
P6	25	F	Industrial Design	3
P7	28	M	Computational Design	5
P8	28	F	Industrial Design	4
P9	27	F	Design Consulting	3
P10	27	M	Art Design	4
P11	26	M	Industrial Design	4
P12	26	M	Design Consulting	3
P13	27	M	Computational Design	4
P14	36	F	Industrial Design	10
P15	32	F	Computational Design	8
P16	25	M	Industrial Design	3

B Technical Details of the Component-level Extraction and Generation Method

The computation process within the self-attention mechanism is described as follows:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{(M_{target} \odot Q)\left((1 - M_{add}) \odot K\right)^T}{\sqrt{d}}\right) V \tag{1}$$

where the query Q, key K, and value V matrices are derived from the latent representation z_{T-k} during the denoising process. These matrices are projected using the weights W_Q, W_K, and W_V respectively. By querying the regions within M_{target} to match the features of the regions within M_{add}, the features within M_{target} can be injected into M_{add}, achieving completion. In addition, to preserve the areas outside the M_{add}, we replicate the component features from the latent z_T via the DDIM inversion path. In the k-th denoising step, we update the current latent z_{T-k} to retain the latest component features:

$$z_{T-k} = z_{T-k} \odot M_{add} + z_T \odot (1 - M_{add}) \tag{2}$$

After T denoising steps, we obtain z_0, which is then decoded through a VAE to restore the complete component image.

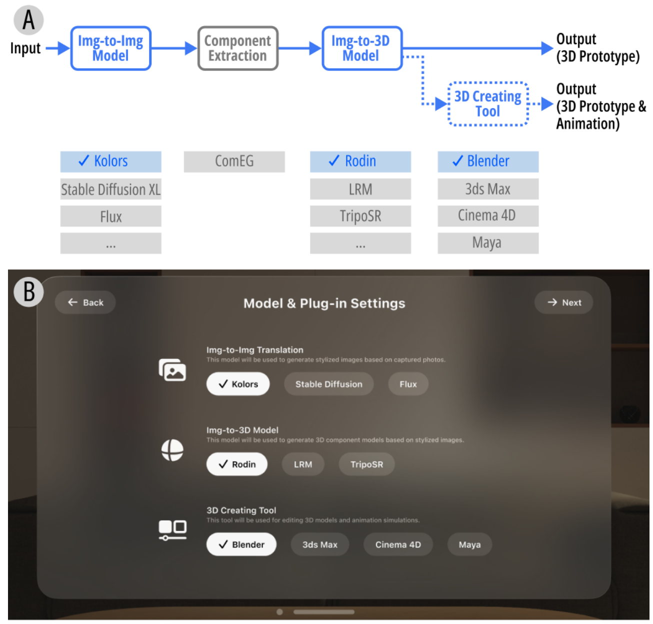

Figure 18: The plug-and-play pipeline (Ⓐ). The user interface for customizing the creation pipeline (Ⓑ).

C Plug-and-play Pipeline in FusionProtor

We designed and developed the pipeline in a plug-and-play manner in FusionProtor. On the one hand, designers can choose appropriate models and tools according to their own needs and familiarity. On FusionProtor's home page (Figure 18Ⓑ), users can customize their own creation pipeline, including changing generative models and 3D Creating Tools. On the other hand, considering the vigorous development of GAI technology, the plug-and-play manner allows the updating and integration of cutting-edge GAI models and deployment across various platforms.

In the plug-and-play framework implementation, the interchangeable components involve the image-to-image translation module, image-to-3D translation module, and 3D Creating Tools (Figure 18Ⓐ). First, for image-to-image translation, we have utilized a locally deployed ComfyUI [20] backend and provided the corresponding configuration files for several image-to-image models [13, 54, 71], enabling seamless switching between different image generation models by simply using various configuration files. Second, for image-to-3D translation, in addition to Rodin, we have also locally deployed TripoSR [74] and LRM [34] image-to-3D generation models as candidates. The component images extracted by ComEG can be directly fed into the image-to-3D models. Third, for 3D Creating Tools, we provided the corresponding PBR Importer plug-ins using Python to automatically bind generated 3D components and their textures received from the back-end server for several 3D Creating Tools Cinema 4D [51], 3ds Max [5] and Maya [8].