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

Hongbo Zhang
hongbozhang@zju.edu.cn
Zhejiang University

Pei Chen*
chenpei@zju.edu.cn
Zhejiang University

Xuelong Xie
xuelongxie@zju.edu.cn
Zhejiang University

Zhaoqu Jiang
zhaoqujiang@zju.edu.cn
Zhejiang University

Yifei Wu
yifeiwu@zju.edu.cn
Zhejiang University

Zejian Li
zejianlee@zju.edu.cn
Zhejiang University

Xiaoyu Chen
rachardchen@zju.edu.cn
Zhejiang University

Lingyun Sun
sunly@zju.edu.cn
Zhejiang University

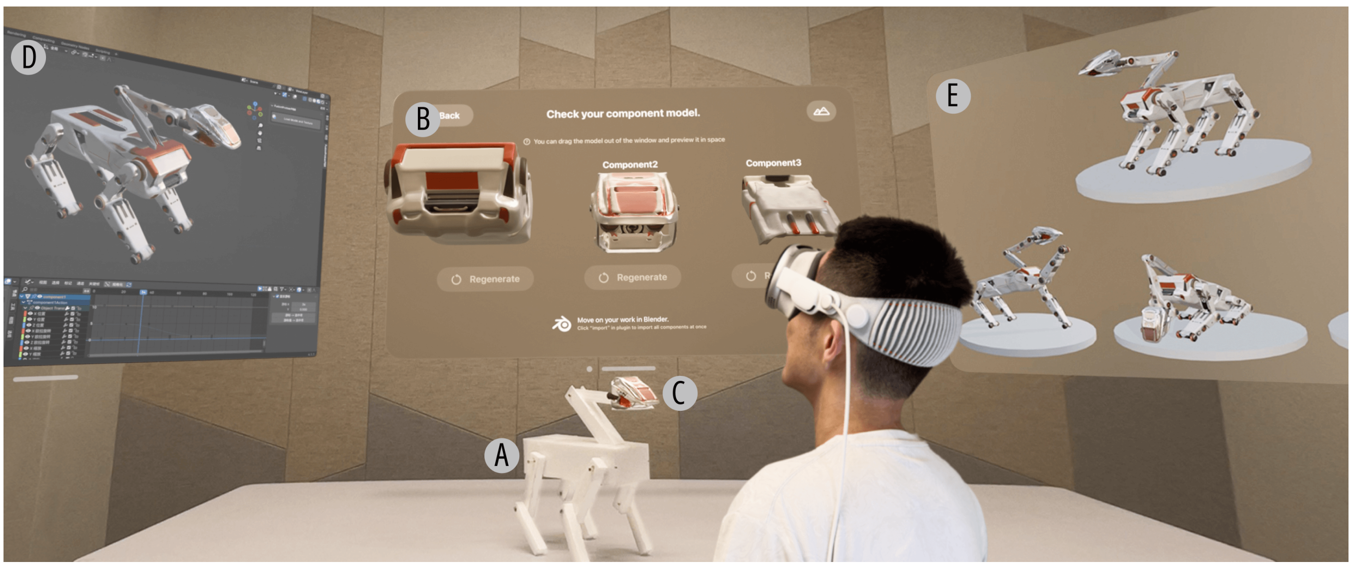

Figure 1: A prototype process supported by FusionProtor. Ⓐ: The designer constructed and interacted with a physical prototype embodiedly. Ⓑ: He utilized FusionProtor to generate component-level 3D prototypes based on the low-fidelity physical prototype. Ⓒ: He coupled the physical prototype with virtual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

Abstract

Developing and simulating 3D prototypes is crucial in conceptual design for ideation and presentation. Both physical and virtual prototypes offer distinct strengths and are integral to the conceptual design process. However, traditional methods often keep these two domains separate, leading to a disjointed prototype workflow. In addition, acquiring high-fidelity prototypes and simulations, whether physical or virtual, is time-consuming and resource-intensive, distracting designers from creative exploration. In this context, we introduce FusionProtor, a mixed-prototype tool for swift component-level physical-to-virtual 3D transition and simulation in conceptual design. It incorporates generative artificial intelligence (GAI) for component-level generation. Building upon generated independent components, it supports simulating dynamic interactions according to physical motion logic. It also leverages augmented reality (AR) to enable designers to mix physical and virtual components for fusion ideation and presentation. We conducted technical and user experiments to verify FusionProtor's robustness and usability in supporting diverse designs. Our results verified that FusionProtor

*The corresponding author.

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX

achieved a seamless workflow from physical to virtual and low- to high-fidelity, enhancing efficiency and promoting ideation. We also critically discussed its best practices and explored the effect of mixed interaction on design 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. 2018. FusionProtor: A Mixed-Prototype Tool for Component-level Physical-to-Virtual 3D Transition and Simulation. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX)*. ACM, New York, NY, USA, 19 pages. https://doi.org/XXXXXXX.XXXXXXX

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 [9]. Given that conceptual design determines up to 70-80% of a product's lifetime cost [11, 21], exploring novel and efficient 3D prototype tools is indeed valuable. Various 3D representations have inherent strengths and application scope in conceptual design [38]. For example, an early-stage physical prototype allows designers to repeatedly test and verify structures, helping to reveal potential flaws in advance [71]. However, these early physical prototypes are often hard to manage versions [24] and lack fidelity [13, 38]. Therefore, once designers determine the basic scale, shape, and structures, they generally transition to virtual prototypes to further refine, convey, and evaluate concrete concepts, which offer high fidelity and available data for further presentation [43].

Conceptual design is typical to progress from a low-fidelity prototype to a high-fidelity form in a linear fashion [23]. Designers face hurdles in this process. First, acquiring high-fidelity prototypes, whether virtual or physical, is time-consuming and resource-intensive, with each involving multiple iterations and manual adjustments to achieve the desired outcome [42]. This traps designers in tedious and laborious expression tasks, distracting them from the core aspect of conceptual design: the "exploration" [33]. Second, the lack of integration and fusion between physical and virtual prototypes makes the two domains operate independently and disconnected, leading to a disjointed workflow. Designers can neither effectively translate low-fidelity physical information into high-fidelity virtual formats nor can they jointly utilize the two domains for mixed simulation and presentation [23]. In this context, effectively leveraging initial design information for refinement and seamlessly integrating physical and virtual prototypes might greatly enhance efficiency and promote ideation.

Recent breakthroughs in generative artificial intelligence (GAI) enable designers to obtain various schemes from rough expressions such as textual descriptions or sketches [16, 90]. The GAI participation can streamline the traditional design process, leading to a burgeoning interest in GAI-assisted design within the HCI community [15, 17, 61, 67, 83, 91]. It opens the door for rapid iteration, fidelity enhancement, and productivity improvement in conceptual design [69]. However, current GAI-assisted design tools still face gaps and challenges in fully achieving the transition from physical to virtual prototypes. On the one hand, few design tools currently support high-fidelity 3D generation due to the previous limitations in 3D generation quality and time. On the other hand, most GAI-assisted design tools focus on generating complete schemes rather than supporting component-level or layered generation, which restricts designers from freely adjusting, iterating, and simulating individual parts like physical prototypes [32, 33]. In this context, addressing these gaps by leveraging GAI for component-level generation might enable the physical-to-virtual transition and support interaction simulations.

This study introduces FusionProtor, a mixed-prototype tool combining physical and virtual prototypes for rapid component-level prototype transition and simulation in conceptual design. FusionProtor supports conceptual design in an unprecedented way. First, by integrating multiple generation models, FusionProtor supports a rapid physical-to-virtual 3D transition pipeline, fully using preliminary design information and realizing the jointed workflow and seamless conversion from low- to high-fidelity. Second, by designer-AI interaction innovation, FusionProtor achieves component-level 3D generation, allowing local refinement, structural assembly, and interaction simulation. Third, by utilizing augmented reality (AR) technique, FusionProtor provides mixed interaction modes with both physical and virtual components, achieving fusion creation and testing in a hybrid world.

In this work, we designed FusionProtor based on a formative study and conducted a pilot study for iteration. Following the system development, we initiated a technical evaluation study to validate the proposed component extraction method and 2D-to-3D transition method involving component-level generation and assembly in FusionProtor. 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 its unique strengths and ideation support mode in conceptual design. This paper makes the following contributions:

- We design FusionProtor, a mixed-prototype tool that achieves a seamless workflow from physical to virtual and low- to high-fidelity, enhancing efficiency and promoting ideation.
- We propose a 3D creation method that involves component-level generation and manual assembly in FusionProtor, improving the generation quality of 3D schemes.
- We conduct technical and user evaluation studies, verifying FusionProtor's usability and critically clarifying its strengths, interaction modes, and application scope.

2 RELATED WORK

2.1 Physical Prototype in Design

The physical prototype is a common design representation in conceptual design due to its distinct strengths. First, the physical prototype offers sensorial stimuli, including sight, touch, and smell.

It makes prototypes more intuitive and interactive, reducing cognitive load in design and stimulating creativity [28]. Second, the low-fidelity physical prototype serves as an ambiguous design representation, providing more room for design interpretation and reasoning [71]. Third, its tangibility supports structural tests and defect exposure, allowing designers to effectively evaluate the shape and structure of their ideas [38]. The tangible dissection and simulation are considered a deep interaction, enhancing structural thinking [81]. However, the long fabrication time, lack of details, and version management are the inherent challenges of early-stage physical prototypes [38].

The HCI community also focuses on integrating physical prototypes into virtual design tools. Among them, utilizing AR equipment is a mainstream method to construct a mixed design space, superposing the inherent strengths of both physical and virtual characteristics [34, 51, 55]. Most previous studies paid attention to adding extra virtual details based on physical prototypes for mixed display or interaction. For example, Peng et al. [55] proposed an interactive design system supporting digital editing on 3D printing prototypes. Barbieri et al. [10] developed a mixed prototyping tool with physical prototypes for virtual evaluation. However, due to previous technical limitations, few studies focused on direct prototype transition and refinement based on constructed physical prototypes, which provides essential value in the seamless connection between physical and virtual prototypes.

2.2 Virtual Prototype in Design

The 3D virtual prototype is another commonly used design representation. It promotes externalizing and conveying detailed intentions and supports motion simulation and engineering calculation for further design implementation [23]. Compared with low-fidelity physical prototype creation, the virtual prototype creation is often more time-consuming and resource-intensive, involving multiple iterations and refined adjustments [42]. In design practices, computer-aided freeform surface modeling is one of the mainstream creation methods of 3D virtual prototypes, which are often high-cost, time-consuming, and skill-threshold [38]. The HCI community has also explored other 3D creation methods. For example, Stemasov et al. [74] developed a parametric design tool in AR for personalization artifacts. Similarly, Alcaide-Marzal et al. [1] introduced a modeling method based on deformation rules and grammars. In addition to parametric modeling, Farqui et al. [25] 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, 41, 75].

Every 3D creation method or tool has a high learning curve and often distracts designers from the creative exploration process [38]. 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 or convert physical prototypes into virtual ones early in the design process. It leads to missed opportunities for concrete display, interaction simulation, and effective communication in conceptual design[23].

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 [46, 69, 91] or assist design reasoning by text [15, 73], supporting ideation [73], accelerating prototype [27], and advising iteration [44]. Other output representations include the icon [87], UI scheme [36, 39], and animation [88]. 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. [94] 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 [32, 68].

The recent dramatic development of GAI, such as the advancements in upscaled model parameters, increased computational power, and greater and 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 [31], DMV3D [89], TripoSR [80], CLAY [47]) and optimization-based 3D generation (e.g., DreamFusion [57], Magic3D [45], MVDream [70], DreamCraft3D [76]). These AI 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 [94]. In this paper, we confront the practical design needs and strive to integrate advanced GAI capabilities to achieve component-level 3D creation from an 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 conceptual design process and challenges, as well as understand their desired features in a prototyping tool.

3.1 Participants and Procedure

All participants are experts with rich theoretical and practical design experience. E1 and E2 are professional designers in a hardware company. E3 is a design leader in an industrial design company. E4 is a PhD candidate majoring in design science. 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.*

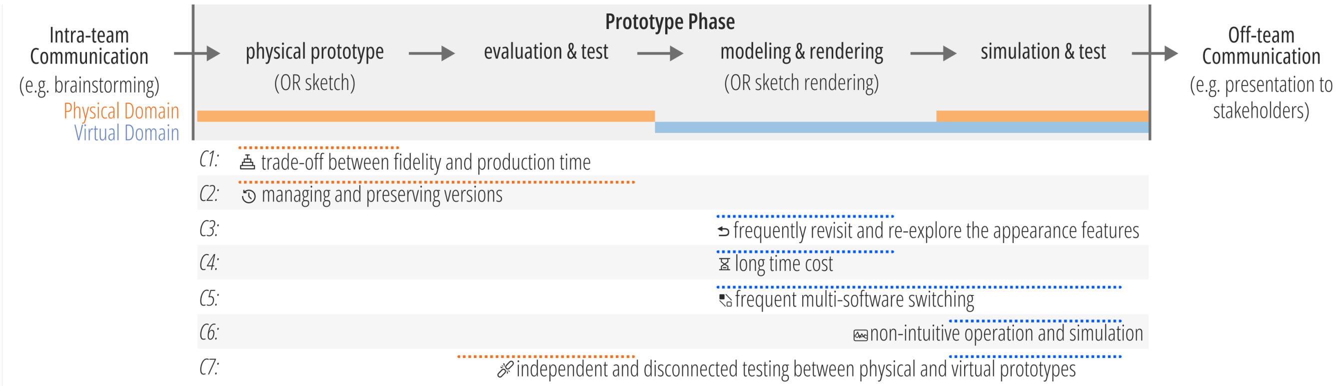

Figure 2: A typical prototype workflow and distinct challenges (C) in product conceptual design, extracted from expert interviews.

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. The three phases might be repeated many times in conceptual design 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, extensively exploring functional and structural possibilities. 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, is 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 challenges in them (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*). Efficient physical prototypes often lack design details, while high-fidelity prototypes require more manual labor. 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.

Designers often adopt software during the modeling and simulation stages. Common tools include AutoCAD [5], SolidWorks [22], Rhino [62], Fusion 360 [6], and Grasshopper [63] for modeling; V-Ray [29], KeyShot [48], and Blender [26] for rendering; and Maya [7], Blender [26], Cinema 4D [50], and 3ds Max [4] 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*). In addition, other challenges in that stage included the *long-time cost* (*C4*), *frequent multi-software switching* (*C5*), *non-intuitive operation and simulation on screen* (*C6*), and *the independent and disconnected testing between physical and virtual prototypes* (*C7*).

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**. 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**. 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 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 simulation**. It can support designers in efficiently verifying the interaction simulation according to the motion logic in physical prototypes, providing dynamic scheme presentation. We subdivide G3 into two sub-goals:

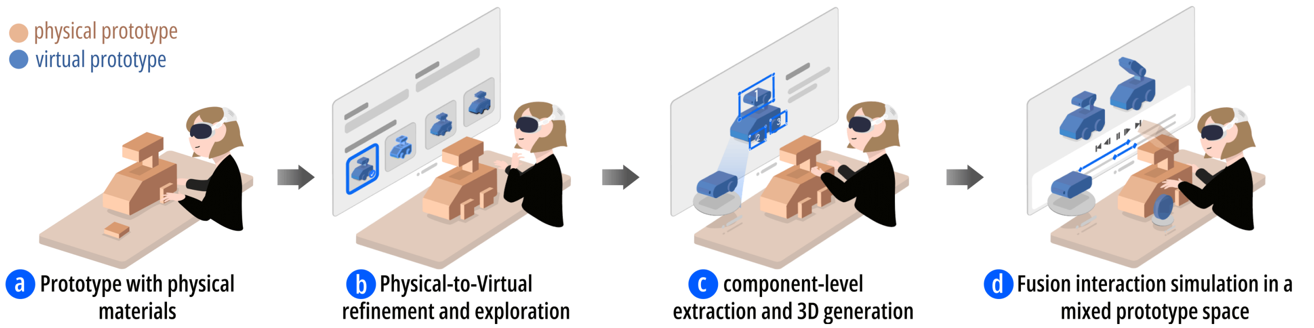

Figure 3: Interaction workflow using FusionProtor in the prototyping process.

- **G3.1: Support 3D scheme extraction and component-level generation**.
- **G3.2: Support 3D components assembly, interaction, and simulation**.

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 AR technique to support mixed prototype and dynamic interaction simulation. We designed FusionProtor in Apple Vision Pro that provides the hybrid 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 3a). After sketching local details and transferring her intended design style based on the captured physical prototype, she obtained various generated high-fidelity renderings that aligned with her design vision (Figure 3b). She then considered component relationships and extracted individual components from the generated complete scheme by adding bounding boxes, leading to the component-level 3D generation (Figure 3c). Next, she assembled these virtual 3D components and simulated their interaction based on the physical motion logic, allowing the mixed component-level ideation and simulation (Figure 3d). With FusionProtor's support, the designer efficiently explored and iterated her idea, significantly enhancing efficiency and promoting ideation.

4.2 Mixed Workspace in FusionProtor (for *G1*)

Designers create across three work areas in FusionProtor (Figure 4A). The *Physical Prototype Work Area* offers various physical materials and tools for tangible creation in the real world. The *Prototype Generation Work Area* supports component-level prototype transitions and generation. The *Prototype Simulation Work Area* enables dynamic component interaction simulation. The seamless coordination of multiple work areas was realized through Apple Vision Pro.

FusionProtor's main user interfaces are presented in Figure 4, including capturing and sketching based on physical prototypes (Figure 4B1), controlling generation (Figure 4B2), extracting components (Figure 4B3), browsing generated component (Figure 4B4) and simulating dynamic interaction (Figure 4B5). FusionProtor's AR characteristic allows designers to couple physical and virtual domains seamlessly. Specifically, FusionProtor supports adding virtual components directly to physical prototypes for fusion iteration (Figure 4B4). It allows designers to simultaneously simulate dynamic interactions while engaging with tangible prototypes (Figure 4B5) and also change environments for immersive ideation and presentation (Figure 4B6).

4.3 Component-level Prototype Transition and Production (for *G2* & *G3*)

4.3.1 Overall Prototype Production Pipeline (for G2.1). The overall prototype production pipeline of FusionProtor is shown in Figure 5. It treats the component-level physical-to-virtual transition as sequential tasks that move between multiple design representations: from physical prototype to captured image, then from captured image to high-fidelity complete scheme, then from complete to component-level schemes, then from 2D component to 3D form, and finally implementing component manual assembly and simulation.

For the captured image to high-fidelity complete schemes, we fed the captured image as input along with a textual description and a reference style image into the Img-to-Img translation Model, Kolors [78] and IP-Adapter [92]. For generating component-level schemes from a complete scheme, we proposed a component extraction method. We then fed component images and descriptions into an Img-to-3D model, Rodin [79]. Designers assembled components and simulated dynamic interactions within FusionProtor, which incorporates Blender capabilities. The introduced production pipeline functions in a plug-and-play manner, allowing the updating and integration of cutting-edge AI models and deployment across various platforms.

4.3.2 Prototype Refinement and Fidelity Optimization (for G2.2). Experts in the formative study pointed out that the low-fidelity physical prototypes lack appearance and style details for further modeling. To address this, we achieved prototype refinement and

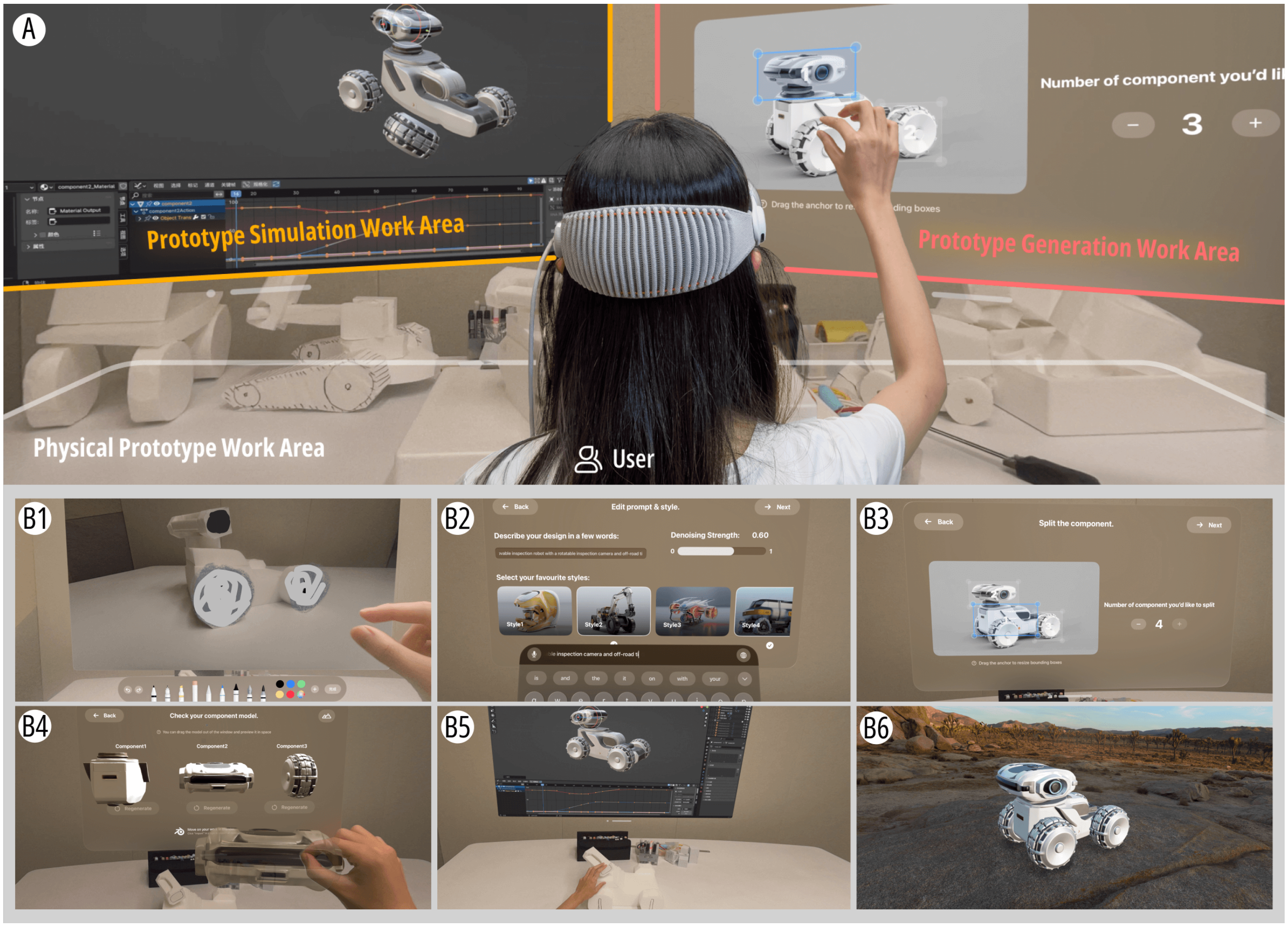

Figure 4: FusionProtor's mixed workspace (A) and main user interfaces (B). B1: capturing and sketching based on a physical prototype. B2: typing textual description and choosing reference style for generation. B3: browsing generated schemes and extracting components. B4: browsing generated 3D components. B5: testing tangible prototypes while assembling components and simulating interaction. B6: adjusting environments for immersive ideation and presentation.

fidelity optimization in FusionProtor. This enables designers to explore structural concepts with physical prototypes while refining style and appearance details using GAI's rapid generation and iteration capabilities. Specifically, we developed *supplementary sketch* and *style transfer* functions (showcased in Figure 6).

The *supplementary sketch* function allows designers to add details to captured images of physical prototypes, reducing the construction time of physical details. In addition, as designers often define style references (e.g., using Mood Board) before appearance design, FusionProtor provides style reference and utilizes IP-Adapter [92] to realize the *style transfer* function. It can transform low-fidelity prototypes into high-fidelity designs that conform to the intended style, enhancing control over AI-generated content and supporting creative appearance exploration.

4.3.3 Component-level extraction and generation (for G3.1). As component-level generation is a prerequisite for virtual 3D interaction simulations, we extract individual components from complete design schemes based on bounding boxes specified by designers. Current mainstream segmentation algorithm [40, 93] focuses on the complete segmentation of visible objects. However, product design schemes often present complex spatial relationships. Specifically, in a planar view of a product image, parts of components may overlap, creating occlusions. We need a segmentation algorithm capable of "imagining" the obscured parts of components and outputting these complete components. This ability is vital for the subsequent generation of 3D components. Against this background, we proposed a component extraction method that integrates inpainting and completion techniques.

As illustrated in Figure 7, our proposed component extraction method includes two phases. In Phase1, given an input image I and a designer specified 2D bounding box b, we utilize the SAM[40] to obtain the segmentation masks of the target component M_{target} and the entire product M_{entire}:

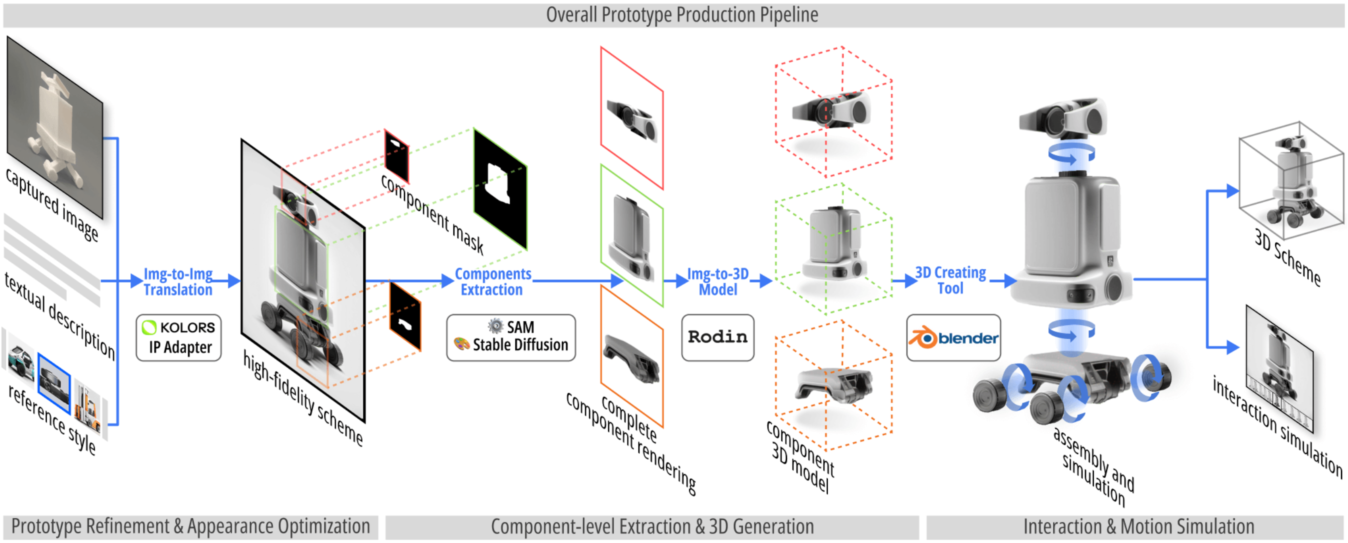

Figure 5: Overall prototype production pipeline.

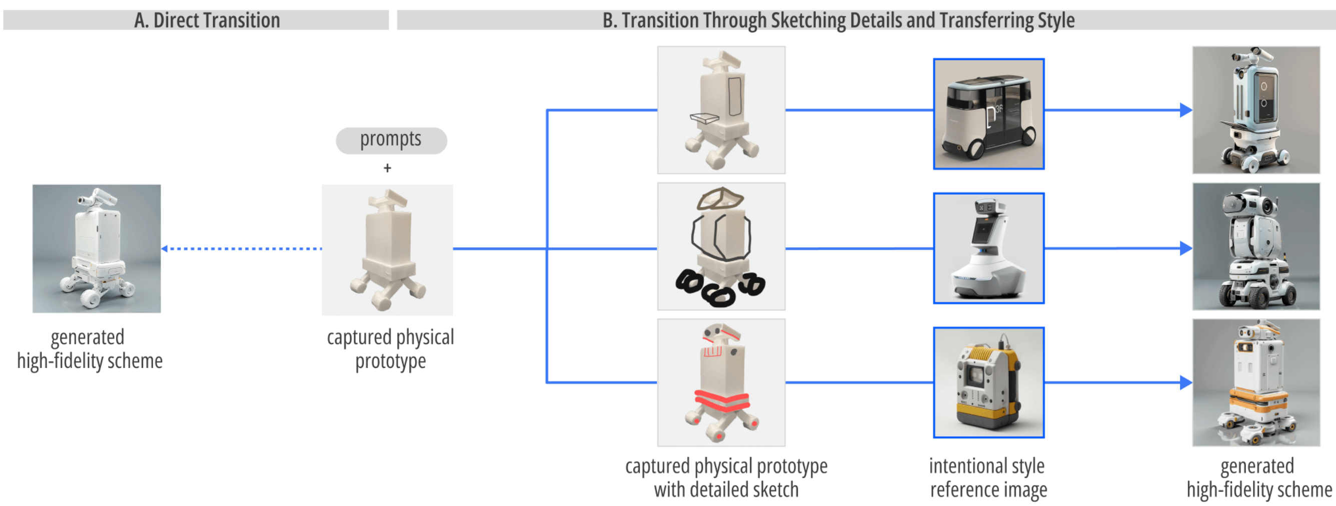

Figure 6: The Function of prototype refinement in FusionProtor (B) and direct generation without this function(A).

$$M_{entire}, M_{target} = \mathrm{SAM}(I, b) \tag{1}$$

then we can obtain the mask M_{remove} for the regions requiring inpainting and completion through M_{target} and M_{entire}:

$$M_{remove} = M_{entire} - M_{target} \tag{2}$$

In Phase2, we conduct inpainting for the target extracted component. Initially, we use DDIM inversion [72] to transform the original image I into a latent vector z_T. Inspired by DesignEdit[35], we integrate a self-attention mechanism into the decoder within the U-Net denoiser to inpaint and complete the area designated by M_{remove} 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 M_{target} region of key features K during the T denoising steps. This allows the regions outside M_{target} to be completed by querying the regions within M_{target}. The computation process within the self-attention mechanism is described as follows:

$$\mathrm{Attention}(Q, K, V) = \mathrm{Softmax}\left(\frac{Q\left((1 - M_{target}) \odot K\right)^T}{\sqrt{d}}\right)V \tag{3}$$

where the query Q, key K, and value V matrices are derived from the latent representation z_{T-k} during the denoising process. These

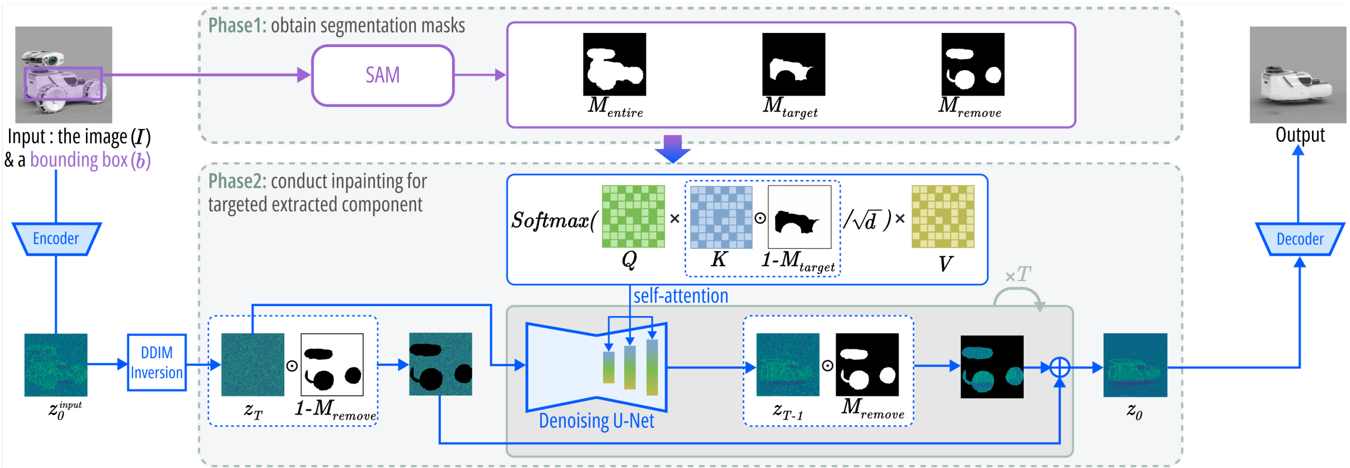

Figure 7: Technical framework of the proposed component extraction method.

matrices are projected using the weights W_Q, W_K and W_V respectively. Since attention weights are based on query-key matching, masking a key significantly reduces its match score, causing the regions outside M_{target} to be excluded from the weighted sum. By extracting the target component area of key features, we enable the query to the regions within M_{target} to complete the regions outside M_{target}. In addition, to preserve the areas outside the remove part, we replicate the component features from the latent z_T via the DDIM inversion path. At the k-th denoising step, we update the current latent z_{T-k} to retain the latest component features:

$$z_{T-k} = z_{T-k} \odot M_{remove} + z_T \odot (1 - M_{remove}) \tag{4}$$

After T denoising steps, we obtain z_0, which is then decoded through a VAE to restore the complete component image. Subsequently, each component is directly reconstructed using an Img-to-3D methods to produce single 3D model.

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.

4.3.4 Interaction and Motion Simulation (for G3.2). We integrated Blender in FusionProtor to support the assembly and interaction simulation of generated 3D components. Blender runs on a local computer whose screen is streamed to a work area on Apple Vision Pro via screen mirroring. We chose Blender for several reasons. It is a powerful and versatile 3D creation software widely used by professional designers for interaction simulation, aligning with their work habits and minimizing learning costs. Additionally, Blender is an open-source 3D creation suite whose open API supports the integration of custom features and the automation of workflows through scripting.

We developed some automation features into FusionProtor to simplify operations and enhance user experience. First, we realized the one-click import of 3D components across devices. Specifically, we developed a blender plug-in to automatically bind generated 3D components and their textures received from the back-end server. Second, we configured the simulation workspace according to professional designers' work habits, eliminating unnecessary work windows. Third, we enabled rapid cross-device export for 3D immersion preview. The aim of our above efforts is to streamline FusionProtor's design process, enabling designers to focus on essential prototype and simulation tasks while FusionProtor takes over complex and non-creative tasks.

4.4 System Architecture and Implementation

FusionProtor's system architecture is shown in Figure8, including an editor module, an AI module, and an integrated blender client. Our front-end interface was responsible for user interaction, visualization, and basic image processing, which was implemented mainly using Swift [66]. FusionProtor also integrated built-in functions of Blender for 3D simulation. We developed PBR Importer plug-in using Python for information transmission. AI functions and complex computations operated on a local server implemented with Nginx [60], Python, Flask [65], PyTorch [54], and other deep learning libraries (ComfyUI [18], Transformer [86], Diffuser [82]), which was equipped with a GTX 3090 GPU. The back-end server deployed some AI models and also integrated services provided by the cloud.

4.5 Pilot Study and Design Iteration

We conducted a pilot study with six designers to evaluate the user experience of FusionProtor in supporting conceptual design and to define its scope for iteration. We introduced the background and motivation behind FusionProtor and provided a three-day prototyping period during which designers could freely use FusionProtor in

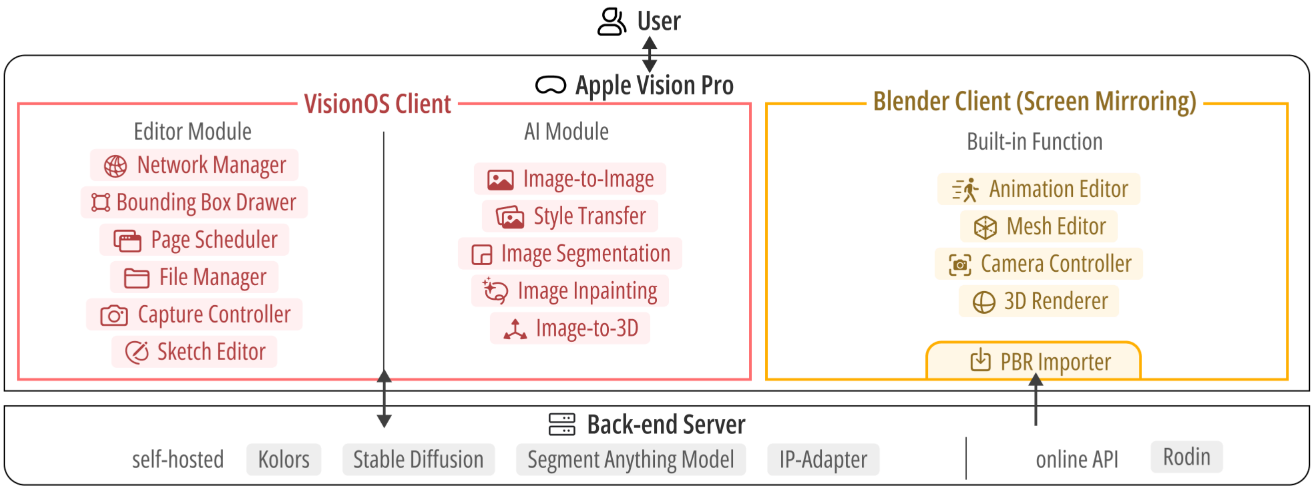

Figure 8: FusionProtor's system architecture.

their studios. Following this period, we conducted semi-structured interviews focusing on 1) *their design experiences with FusionProtor* and 2) *the challenges they faced during prototyping*.

Based on the designers' feedback, we iterated FusionProtor to enhance usability and user experience. First, since designers found typing prompts via a virtual keyboard in an AR environment inefficient, we implemented voice input for more intuitive and natural interaction. Second, to address the need for multiple generations to control design direction and refine details, we added a regenerate function for each generation stage, covering both 2D and 3D outputs. Third, responding to designers' desire to upload and customize reference styles instead of being limited to predefined options, we updated the upload and style customization feature.

5 TECHNICAL EVALUATION

We conducted a technical evaluation to verify the technology 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 Sec. 6) for technical evaluation. Specifically, for TRQ1, we randomly selected 20 2D complete schemes generated by FusionProtor integrated Kolors [78] as a test dataset. Each 2D scheme has a foreground mask chosen by designers through the bounding box and generated by SAM [40]. For TRQ2, we also randomly selected 20 2D complete schemes that did not overlap with dataset in TRQ1.

Baselines. For TRQ1, we compared our method with three mainstream image editing models, SDXL-Inpainting [56], ControlNet-Inpainting [95], and PowerPaint [97]. 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 [79], which we integrated into FusionProtor's technical pipeline.

Metrics. For TRQ1, we evaluated the performance of our component extraction method on five metrics. L1 was used to gauge the pixel-level difference between the generated image and the ground truth image. CLIP-I [58] was used to measure the image quality with the cosine similarity between the generated image and ground truth image. We evaluated the different aspects of image similarity including Peak signal-to-noise ratio (PSNR), the structural similarity (SSIM) [84], and perceptual similarity (LPIPS) [96]. Since component extraction task aims at recovering the masked regions according to reference image contexts, we used the original image as ground truths. For TRQ2, to evaluate the performance of the 2D-to-3D transition method, each generated 3D asset was rendered into 18 view images to compute quantitative metrics, including CLIP score, PSNR, SSIM and LPIPS. Additionally, we also conducted a human perceptual evaluation. We invited 10 users to compare our proposed methods with baselines and asked them to choose a better one with higher quality that matches the task requirements. They were shown complete and extracted component images in TRQ1 and multi-view images of 3D schemes with the 2D input image for TRQ2.

5.2 Results

Qualitative Comparison. For TRQ1, as shown in Figure 9, 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

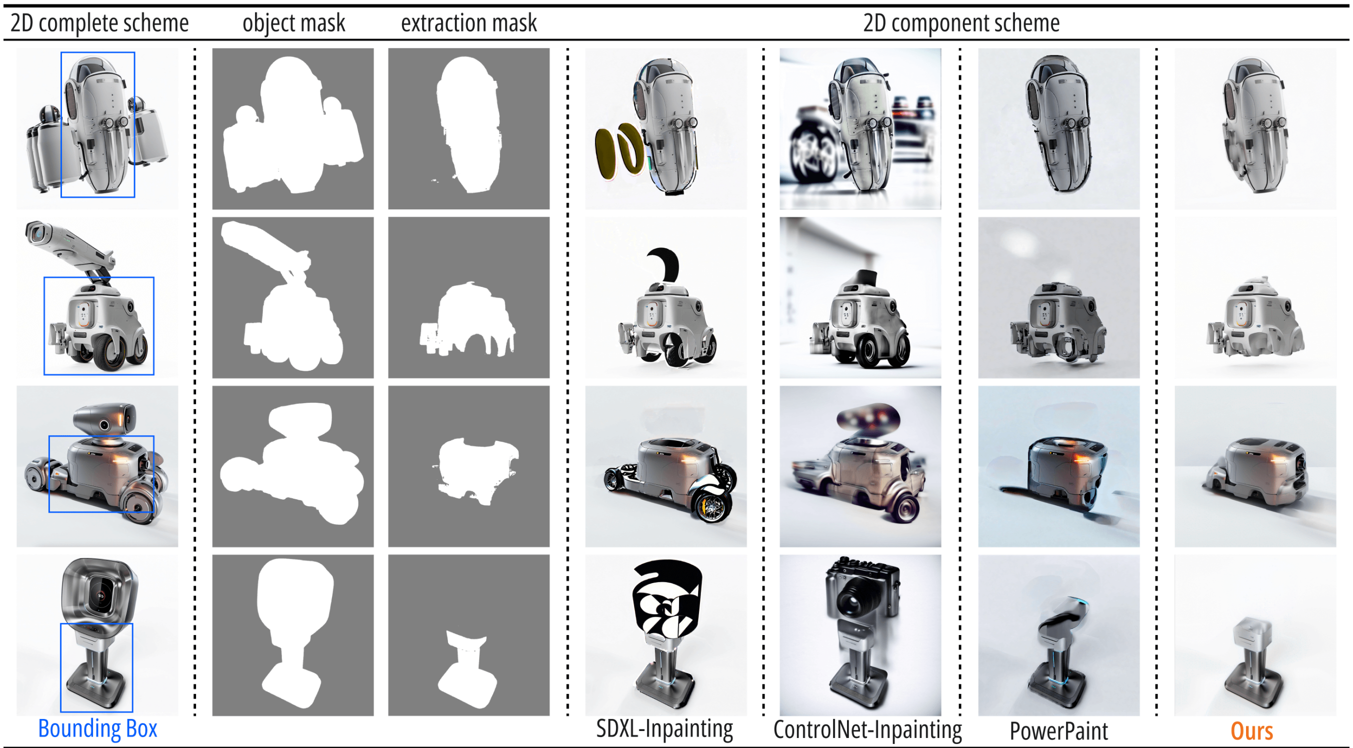

Figure 9: Visual comparison for our component extraction method.

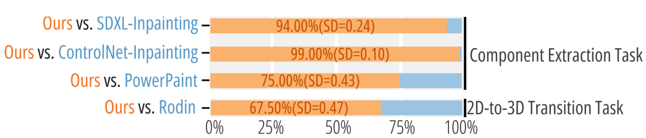

Figure 10: Human perceptual comparison of ours vs. baselines.

Figure 9, AI adds irrelevant content. In the 3rd row and 6th column, AI 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 and structural relationships. For example, in the 3rd row, AI generates better camera detail due to component-level generation. In the 2nd row, the structural relationship of components is better expressed due to manual assembly.

Quantitative Comparison. The quantitative results for TRQ1 and TRQ2 are presented in Table 1 and Table 2, respectively. For the component extraction task, our method outperformed comparison methods in L1, CLIP-I, SSIM, and PSNR. It validated that our component extraction method can effectively extract product components. For the 2D-to-3D transition task, our method resulted better in CLIP Score, SSIM, PSNR, and LPIPS compared to Rodin. It indicated that our 2D-to-3D transition method can produce higher-quality 3D models compared to direct 3D generation.

Human Perceptual Comparison. Figure 10 shows our methods outperform others in both the component extraction task and 2D-to-3D transition task. Specifically, in the component extraction task, our method secured 94.00%, 99.00%, and 75.00% votes when compared to SDXL-Inpainting [56], CortrolNet-Inpainting [95], and PowerPaint [97]. In the 2D-to-3D transition task, our method secured 67.50% votes when compared to Rodin direct generation [79].

Table 1: Quantitative comparison for proposed component extraction method. ↑ indicates higher is better.

Method	L1↓	CLIP-I↑	SSIM↑	PSNR↑	LPIPS↓
SDXL-Inpainting [56]	0.091	82.356	0.741	14.089	**0.227**
ControlNet-Inpainting [95]	0.092	82.019	0.693	15.226	0.385
PowerPaint [97]	0.155	82.852	0.737	13.504	0.355
Ours	**0.076**	**85.997**	**0.800**	**15.462**	0.243

1

Table 2: Quantitative comparison for proposed 2D-to-3D transition method. ↑ indicates higher is better.

Method	CLIP Score↑	SSIM↑	PSNR↑	LPIPS↓
Rodin [79]	81.239	0.555	10.073	0.581
Ours	**83.435**	**0.563**	**10.226**	**0.577**

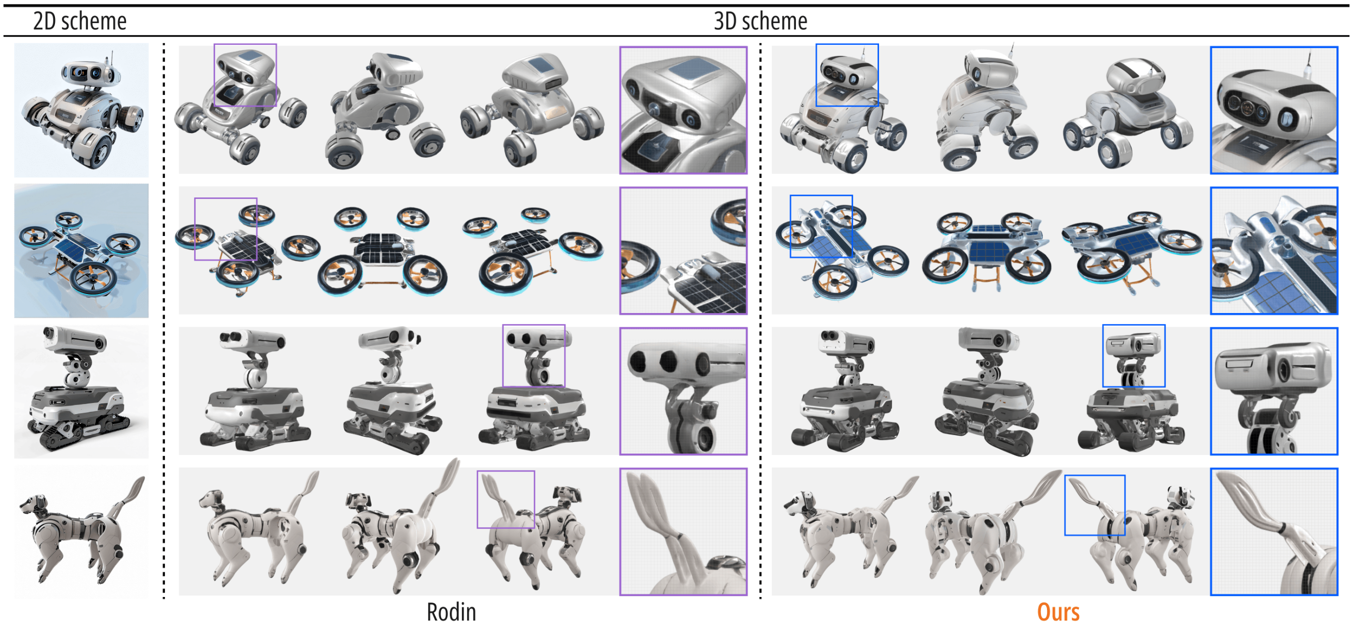

Figure 11: Visual comparison for our 2D-to-3D transition method integrating component-level generation and assembly.

6 USER EVALUATION STUDY

We address the following research questions in the user evaluation study.

RQ1: Is FusionProtor usable? Can FusionProtor support diverse product design tasks?
RQ2: What is the influence of FusionProtor on prototype 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. All participants in the user evaluation study were newly enrolled and had not participated in any prior research related to this study. They can skillfully use modeling software, especially Blender, and have experience in designing with GAI tools (e.g., Midjourney [20], Stable Diffusion [64], and DALL-E [59]). 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.

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 (Figure 12). 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. 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.

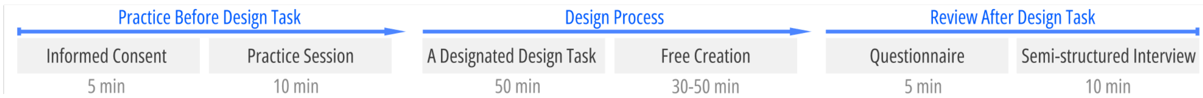

Figure 12: Procedure of the user evaluation study. The order of the two design tasks are counterbalanced among participants.

Table 3: 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, SUS usability, SUS learnability
	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
	distinct strength	■ semi-structured interview by thematic analysis	/
RQ3	ideation mode	■ behavior analysis by Think-aloud protocol	/
	prototype interaction		

6.4 Measurement

As summarized in Table 3, 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 clearly 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) [12] was applied to evaluate FusionProtor's usability for RQ1. NASA Task Load Index (TLX) [30] was adopted to assess the workload of the design process for RQ2. It is an overall workload score based on a weighted average ratings on *mental demand*, *physical demand*, *temporal demand*, *effort*, *performance*, and *frustration level*. Besides, the Creativity Support Index (CSI) [14] 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 [61, 94], 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 [77] to extract commonly-mentioned codes. They independently coded all raw data and shared their codes, discussing inconsistent codes to resolve disagreements and merging similar codes until they reached a consensus.

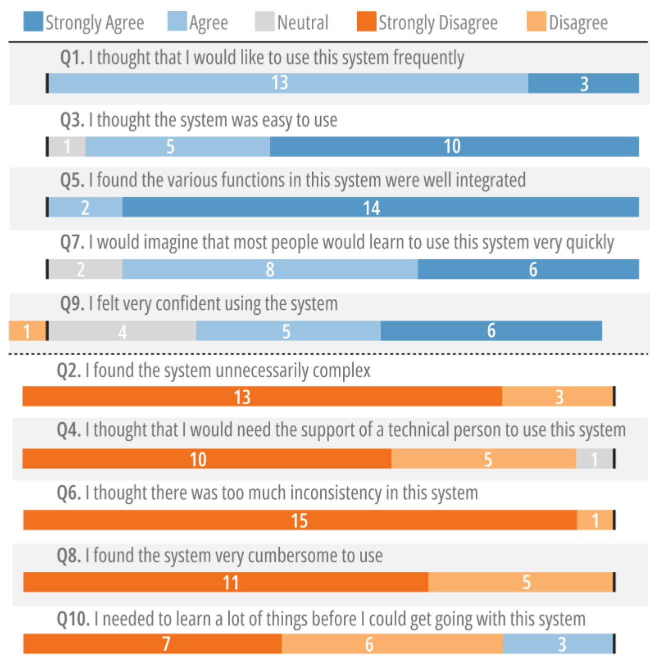

Figure 13: Results of the SUS questionnaire

We adopted the concurrent Think-aloud protocol [3, 37] to answer RQ3 during the prototype process. Three researchers reviewed participants' design process and design behavior. 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

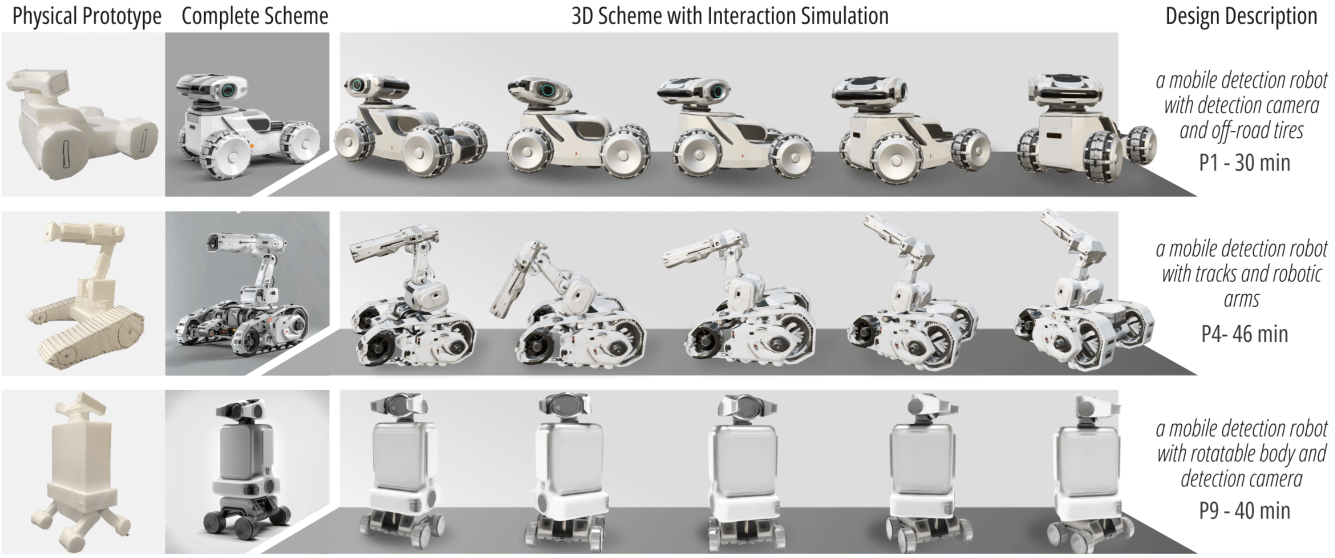

Figure 14: The presentation of Task 1's design outcomes under FusionProtor's support.

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.

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 ($SD = 11.59$) (Figure 13), in which the *learnability* and *usability* scores were 85.16 ($SD = 16.07$) and 88.48 ($SD = 10.58$). The SUS result was rated as "*acceptable*" and "*excellent*" according to Bangor's standard [8].

FusionProtor reaches the practical conceptual design standard. Figure 14 shows some randomly-selected schemes from Task 1. Three design leaders from a design company reviewed the design schemes of Task 1, including 2D renderings, 3D models, simulation video, and textual descriptions, 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. They deemed the quality of the schemes sufficient to effectively express concepts in actual conceptual design practices.

FusionProtor supports diverse design. Figure 15 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 high availability and 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 16ⒶⒷ. 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 indicated that FusionProtor can 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 in semi-structured interviews in Figure 16Ⓒ. The theme of these codes is the distinct advantages of FusionProtor compared with traditional design tools. The most frequently mentioned strength was "*Improve Efficiency*". Designers reported that "*FusionProtor supported me in obtaining high-quality conceptual design work quickly*" (P7) and "*it will take me about a week to model these schemes using traditional methods*" (P11). "*Enhance Expressiveness*" was another distinct strength. P2 indicated that "*FusionProtor enriched many details based on my low-fidelity physical model by integrating AI capabilities*". Similarly, eight participants reported that FusionProtor can support design exploration. Specifically, P4 mentioned that "*AI's participation has given me lots of inspiration, including details, structures, functions, and appearance styles*". Due to the AI integration, designers considered that FusionProtor could take over the tedious expression tasks, allowing them to focus on problem insight and design exploration, also reducing their manual labor for the 2D and 3D creation in the traditional design process. In addition, seven participants reported that the physical prototype and AR participation promoted their structural thinking and supported tangible tests.

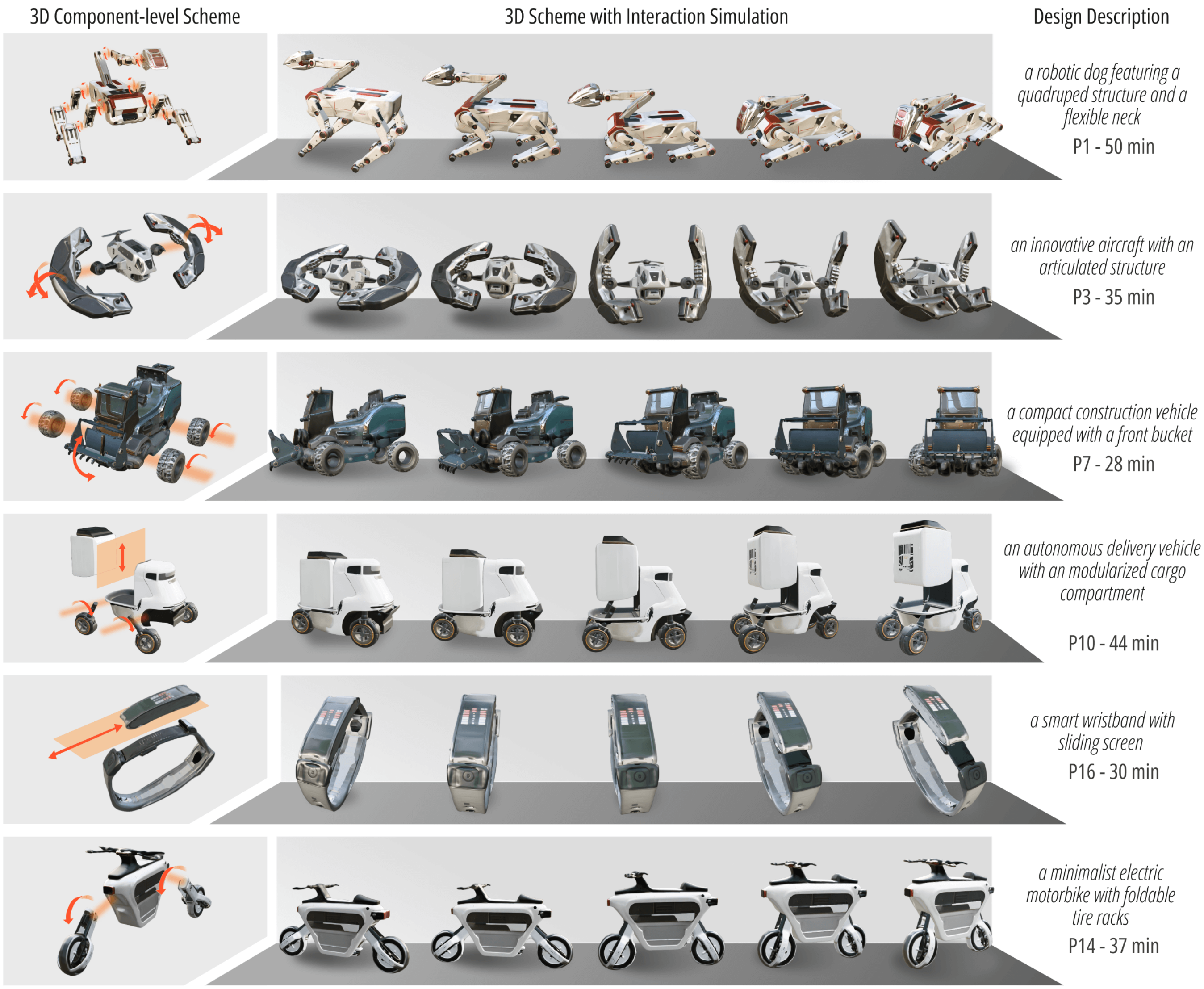

Figure 15: The presentation of Task 2's diverse design outcomes under FusionProtor's support.

6.5.3 FusionProtor Supports Diverse Mixed Ideation and Creation Modes (for RQ3). We used think-aloud and behavior analysis to explore FusionProtor's effect on mixed ideation and creation.

Influence on the ideation process. We first pay attention to its influence on the ideation process, in which we focus on how designers advance their design process. Figure17 presents some workflows using FusionProtor, clarifying participants' diverse ideation processes with FusionProtor. For example, the duration of each design stage and design domain was relatively average in P1's workflow (Figure17.P1). P1's design process was promoted in a linear way without any repeated operations. This linear mode supported an efficient conceptual design task. Similarly, P2 also progressed linearly but dedicated significant time to constructing and testing physical prototypes (Figure17.P2). 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. P2 indicated that "*FusionProtor makes my ideation more intuitive to test the structure and highlight the infeasibility compared with sketches*". On the contrary, P12 spent most of time in the virtual domain, especially on 2D scheme generation (Figure17.P12). P12 made a basic physical prototype and then gained various inspirations through repeated generations. P12 reported that "*I can get various inspiration by adjusting the reference styles and parameters in AI generation, and I constantly adjust my prototype*". In addition, P15's ideation process involved frequent transitions between physical and virtual domains (Figure17.P15). 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 AI generation*".

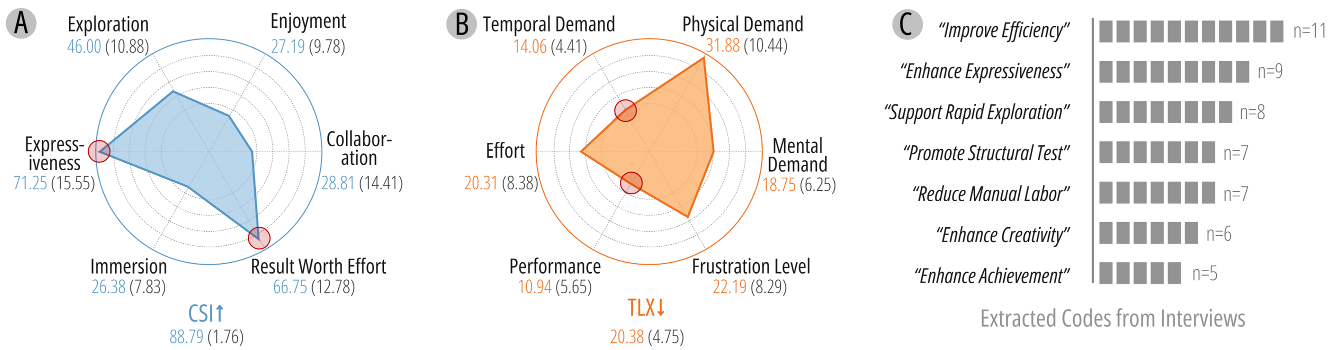

Figure 16: The CSI (Ⓐ) and TLX (Ⓑ) results, ↑ indicates higher value is better while↓ indicates lower is better. Ⓒ shows the extracted codes from interviews.

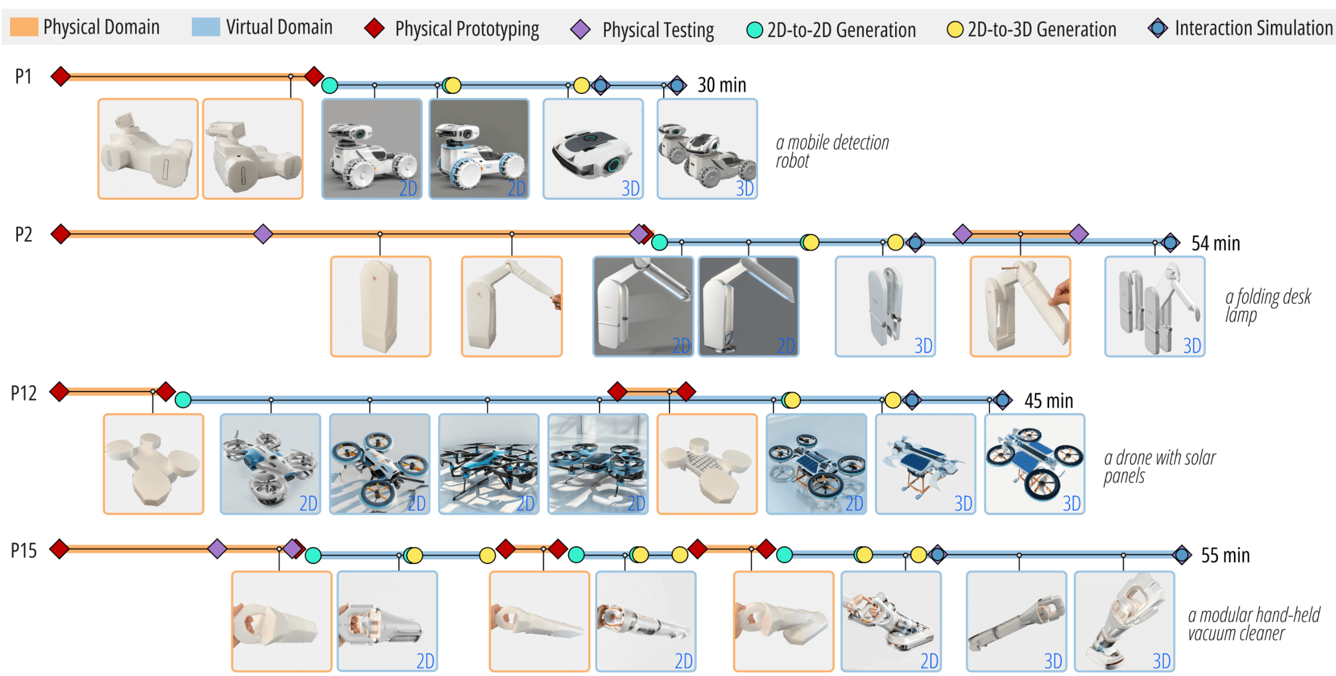

Figure 17: Ideation and prototype process in user study, which extracted through think-aloud and behavior analysis.

Influence on prototype interaction and simulation. We explored how designers prototype in mixed space and operate both the physical and virtual components. The majority of participants combined virtual and physical components for mixed presentation (Figure 18Ⓐ1Ⓐ2). Some also integrated components with the physical environment, notably in the on-site design of wearable components (Figure 18Ⓐ3Ⓐ4). Additionally, during the assembly and simulation stage, participants simulated component interaction based on the interaction logic in the physical world for mixed tests (Figure 18Ⓑ1Ⓑ2). Furthermore, participants leveraged the AR features of Apple Vision Pro to alter backgrounds and enhance immersion when reviewing and presenting the complete scheme (Figure 18Ⓒ1) or components (Figure 18Ⓒ2).

7 DISCUSSION

7.1 FusionProtor's Distinct Strengths

We first discuss FusionProtor's distinct strengths by comprehensively reviewing our results. From the efficiency perspective, FusionProtor revolutionized the traditional design process and expedited prototyping. Designers can not only externalize ideas rapidly with GAI's ability but also verify and simulate them instantly with AR's participation. Feedback from interviews indicates that FusionProtor enabled designers to complete their work in about one hour, a task that typically took several days or even a week in the traditional modeling way.

From the creativity perspective, FusionProtor supported creativity and prompted ideation. The GAI's randomness and variability

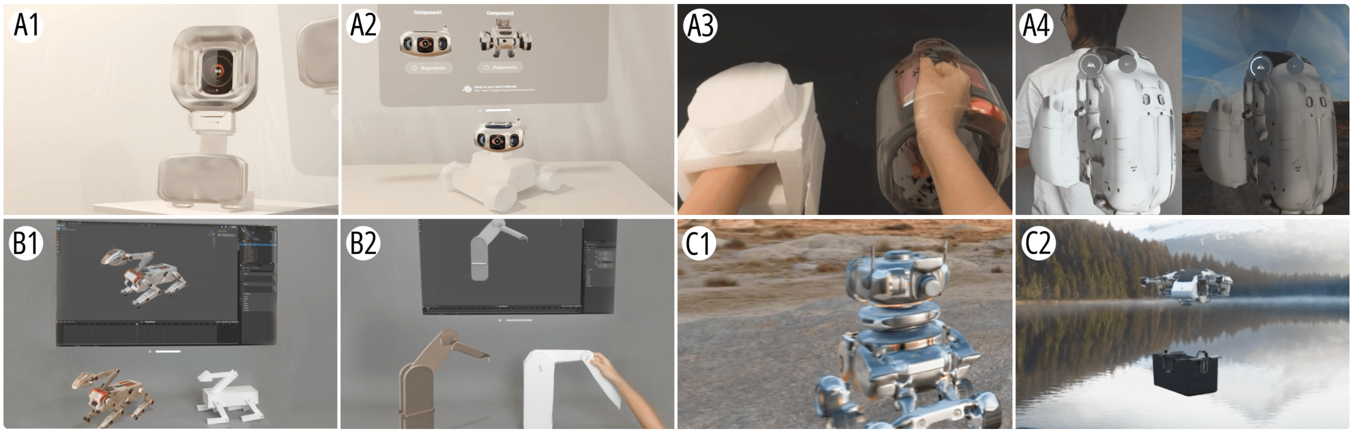

Figure 18: Mixed interaction and simulation with FusionProtor's support, which is extracted through the record of user study.

in FusionProtor greatly enriched designers' ideas and design details, opening imagination space and stimulating their creativity [85]. Additionally, the component-level design method improved structural thinking and supported the mixed simulation, helping designers focus on, consider, and test details and structures.

From the quality perspective, FusionProtor's technical pipeline can effectively produce high-quality design schemes that meet practical conceptual design standards. Our technical evaluation verified that the proposed 3D creation method involving component-level generation and manual assembly effectively improved the 3D design quality. Additionally, our user study showed that FusionProtor's generation pipeline, which integrates multiple generation models, aligns with the linear progression of conceptual design from low- to high-fidelity and from physical to virtual [23]. Our plug-and-play generation pipeline and component-level generation method offer insights into both the AI and design community from the HCI perspective.

7.2 The Effect of Novel Interaction in FusionProtor on Conceptual Design

As FusionProtor is a novel tool integrating AI and AR for component-level mixed prototype, we aimed to discuss the effect of this novel prototype method on conceptual design.

AR and GAI combination enabled the real-time mixed prototype. While many HCI studies have used AR to create mixed environments for design, most virtual objects are preset or retrieved from databases [34, 51]. FusionProtor integrated GAI and AR to enable the real-time mixed prototype, allowing designers to flexibly adjust and iterate high-fidelity virtual prototypes immediately. This capability supported diversified mixed-prototype simulations during design ideation. With FusionProtor's support, designers can feel the physical prototype embodiedly while seeing the high-fidelity virtual feedback immediately. They can also iterate prototypes in diverse environments in FusionProtor, supporting on-site design and immersive ideation.

Joint application of multi-modal generation models released manual labor and optimized the rhythm of the conceptual design. First, as the high-fidelity 3D expression often required additional manual labor and time in traditional way [23, 49], FusionProtor's rapid refinement function and 2D-to-3D transition pipeline significantly reduced the manual labor involved in 3D production. FusionProtor allows designers to focus more on creative contributions while GAI handles complex and tedious tasks, thereby optimizing the rhythm of the conceptual design process. Additionally, FusionProtor effectively integrated various design representations throughout the conceptual design process by the joint application of multi-modal generation models, utilizing preliminary low-fidelity information to guide high-fidelity expression. This promotes a seamless transition between prototype steps and reduces duplication of manual labor.

Component-level generation promoted detailed thinking and collaborating satisfaction with GAI. Most participants noted that the component-level generation and design is a clever and effective interaction innovation in prototyping. This approach provides a foundation for seamlessly coupling virtual and physical components rather than merely viewing prototypes in two domains separately. In addition, designers also considered that component-level design promotes their evaluation of component feasibility and adaptability, facilitating the exploration of modular structures and details. Moreover, component-level generation is advantageous for directing design and iterating details. During our user study, designers preferred iterating local parts based on satisfactory components rather than regenerating the entire scheme. They considered this designer-AI interaction, which is generated component-by-component and step-by-step, has enhanced their sense of control and satisfaction in collaborating with GAI.

7.3 FusionProtor's Critical Discussion, Best Practices, and Optimization Space

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 19A1). In extracting 2D components, our user study found that FusionProtor exhibited limitations in extracting components from structures with severe

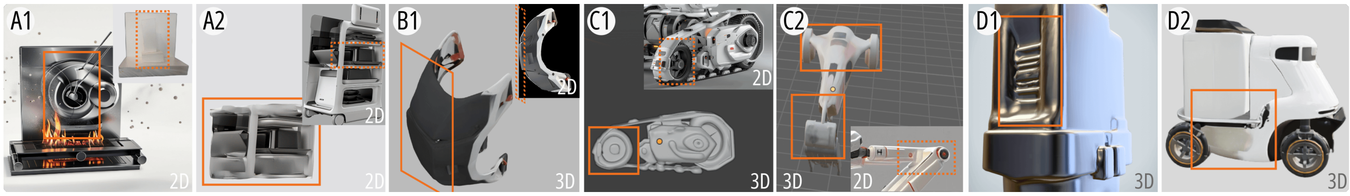

Figure 19: Problematic generated schemes in user study for critical discussion.

occlusion or tight integration, such as isolating a modular container from a *mobile delivery vehicle* (Figure 19Ⓐ2). During the 2D-to-3D transition, the generation quality diminished when designers provided flat and orthographic 2D schemes without any perspective distortion (Figure 19Ⓑ).

In addition to these edge cases caused by user operations, we also critically discussed FusionProtor's common limitations. On the one hand, FusionProtor lacks the capability to comprehend the structural logic and relationships between components. For example, in Figure 19Ⓒ1, FusionProtor generated 3D components that lack structural coherence and assembly relationships. Similarly, in Figure 19Ⓒ2, FusionProtor generated two connected parts, yet the structural dimensions of the joints differed significantly. On the other hand, designers pointed out that although FusionProtor achieved remarkable 3D quality, these schemes still fall short of being directly applicable to engineering production. This includes the presence of imperfect and inaccurate geometry (Figure 19Ⓓ1) and textures (Figure 19Ⓓ2).

Building upon above understandings, we proposed FusionProtor's best practices. First, we recommend using FusionProtor primarily for conceptual design prototyping and simulation, as it is well-suited for high-fidelity presentations and effective stakeholder communication. It is not appropriate for engineering production. We particularly recommend FusionProtor for prototyping and simulating products with multiple independent components and rotational relationships, such as articulated products. In practical application, we advise designers to input extensive textual descriptions and select diverse reference styles to generate candidate designs that align with their intentions. Additionally, designers should avoid capturing physical prototypes from orthogonal angles to enhance the 3D generation quality.

We proposed FusionProtor's optimization space based on critical discussion. Specifically, we are considering optimizing the scale and size of generated 3D models. 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 [95] might allow for the creation of components with more proportionate relative sizes. Additionally, we hope to enable FusionProtor to have structural knowledge. As the multi-modal vision language models, like GPT-4V [53], Gemini 1.5 [52], and Claude 3 [19], have validated their potential in structural and engineering reasoning, FusionProtor's structural understanding ability might be improved by the integration of these large multi-modal models.

7.4 Limitations

We discuss the limitations in this study. First, the method for inputting physical design representations into the GAI is basic. FusionProtor currently captures a single image, which may result in perspective distortions and be dependent on shooting quality. Future enhancements could incorporate multi-view imaging, 3D scanning, or tangible user interfaces for more accurate inputs. 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. Finally, this study was 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 AR 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.

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] Autodesk Inc. 2024. 3ds Max. Computer Software. https://www.autodesk.com/products/3ds-max

[5] Autodesk Inc. 2024. AutoCAD 2025. Computer Software. https://www.autodesk.com/products/autocad/overview

[6] Autodesk Inc. 2024. Fusion 360. Computer Software. https://www.autodesk.com/products/fusion-360/overview

[7] Autodesk Inc. 2024. Maya. Computer Software. https://www.autodesk.com/products/maya

[8] 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.

[9] 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.

[10] 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.

[11] W Beitz, G Pahl, and K Grote. 1996. Engineering design: a systematic approach. *Mrs Bulletin* 71 (1996), 30.

[12] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] Comfyanonymous. 2023. *ComfyUI*. https://github.com/comfyanonymous/ComfyUI

[19] Anthropic Company. 2023. Gemini 1.5. https://claude.ai.

[20] Midjourney Company. 2022. Midjourney. https://www.midjourney.com.

[21] J Corbett and JR Crookall. 1986. Design for economic manufacture. *CIRP Annals* 35, 1 (1986), 93–97.

[22] Dassault Systèmes. 2024. SolidWorks 2025. Computer Software. https://www.solidworks.com

[23] 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.

[24] 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.

[25] 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.

[26] Blender Foundation. 2024. Blender. Computer Software. https://www.blender.org

[27] 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.

[28] 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.

[29] Chaos Group. 2022. V-Ray. Computer Software. https://www.chaos.com/vray

[30] 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.

[31] 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).

[32] 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.

[33] Rong Huang, Haichuan Lin, Chuanzhang Chen, Kang Zhang, and Wei Zeng. 2024. PlantoGraphy: Incorporating iterative design process into generative artificial intelligence for landscape rendering. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–19.

[34] 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.

[35] 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).

[36] 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.

[37] 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).

[38] Lee Kent, Chris Snider, James Gopsill, and Ben Hicks. 2021. Mixed reality in design prototyping: A systematic review. *Design Studies* 77 (2021), 101046.

[39] 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.

[40] 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).

[41] 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.

[42] 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.

[43] 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.

[44] 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.

[45] 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.

[46] 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.

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

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

[49] David Mathias, Chris Snider, Ben Hicks, and Charlie Ranscombe. 2019. Accelerating product prototyping through hybrid methods: Coupling 3D printing and LEGO. *Design studies* 62 (2019), 68–99.

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

[51] 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.

[52] OpenAI. 2023. Gemini 1.5. https://deepmind.google/technologies/gemini.

[53] OpenAI. 2023. GPT-4V(ision) system card. https://chat.openai.com/.

[54] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems* 32 (2019).

[55] 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.

[56] 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).

[57] 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
[58] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*. PMLR, 8748–8763.
[59] 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.
[60] Will Reese. 2008. Nginx: the high-performance web server and reverse proxy. *Linux Journal* 2008, 173 (2008), 2.
[61] 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.
[62] Robert McNeel Associates. 2021. Rhinoceros 3D (Rhino) 8. Computer Software. https://www.rhino3d.com
[63] Robert McNeel Associates. 2022. Grasshopper for Rhino. Graphical Algorithm Editor, Computer Software. https://www.grasshopper3d.com
[64] 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.
[65] Armin Ronacher. 2023. *Flask*. https://flask.palletsprojects.com/en/2.2.x/. Accessed Sep 12, 2024.
[66] Mishal Shah. 2023. *Swift*. https://www.swift.org/ Accessed Sep 12, 2024.
[67] 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.
[68] 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.
[69] 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.
[70] 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).
[71] 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.
[72] Jiaming Song, Chenlin Meng, and Stefano Ermon. 2020. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502* (2020).
[73] 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.
[74] 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.
[75] 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.
[76] 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).
[77] Jon Swain. 2018. A hybrid approach to thematic analysis in qualitative research: Using a practical example. *Sage research methods* (2018).
[78] Kolors Team. 2024. Kolors: Effective Training of Diffusion Model for Photorealistic Text-to-Image Synthesis. *arXiv preprint* (2024).
[79] Deemos Tech. 2024. Rodin Gen-1. https://hyperhuman.deemos.com/rodin.
[80] 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).
[81] 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.
[82] Patrick von Platen, Suraj Patil, Anton Lozhkov, Pedro Cuenca, Nathan Lambert, Kashif Rasul, Mishig Davaadorj, Dhruv Nair, Sayak Paul, William Berman, Yiyi Xu, Steven Liu, and Thomas Wolf. 2022. Diffusers: State-of-the-art diffusion models. https://github.com/huggingface/diffusers.
[83] Sitong Wang, Savvas Petridis, Taeahn Kwon, Xiaojuan Ma, and Lydia B Chilton. 2023. PopBlends: Strategies for conceptual blending with large language models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.
[84] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* 13, 4 (2004), 600–612.
[85] Justin D Weisz, Michael Muller, Jessica He, and Stephanie Houde. 2023. Toward general design principles for generative AI applications. *arXiv preprint arXiv:2301.05578* (2023).
[86] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2020. Transformers: State-of-the-Art Natural Language Processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Association for Computational Linguistics, Online, 38–45. https://www.aclweb.org/anthology/2020.emnlp-demos.6
[87] 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.
[88] 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.
[89] 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).
[90] 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.
[91] 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.
[92] 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).
[93] Lemeng Wu Xiaoyu Xiang Fanyi Xiao Chenchen Zhu Xiaoliang Dai Dilin Wang Fei Sun Forrest Iandola Raghuraman Krishnamoorthi Vikas Chandra Yunyang Xiong, Bala Varadarajan. 2023. EfficientSAM: Leveraged Masked Image Pretraining for Efficient Segment Anything. *arXiv:2312.00863* (2023).
[94] 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.
[95] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2023. Adding Conditional Control to Text-to-Image Diffusion Models. https://github.com/lllyasviel/ControlNet-v1-1nightly
[96] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 586–595.
[97] 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).