# Inkspire: Supporting Designers to Prototype Product Designs through Sketching

DAVID CHUAN-EN LIN, Carnegie Mellon University, USA
HYEONSU B. KANG, Carnegie Mellon University, USA
NIKOLAS MARTELARO, Carnegie Mellon University, USA
ANIKET KITTUR, Carnegie Mellon University, USA
YAN-YING CHEN, Toyota Research Institute, USA
MATTHEW K. HONG, Toyota Research Institute, USA

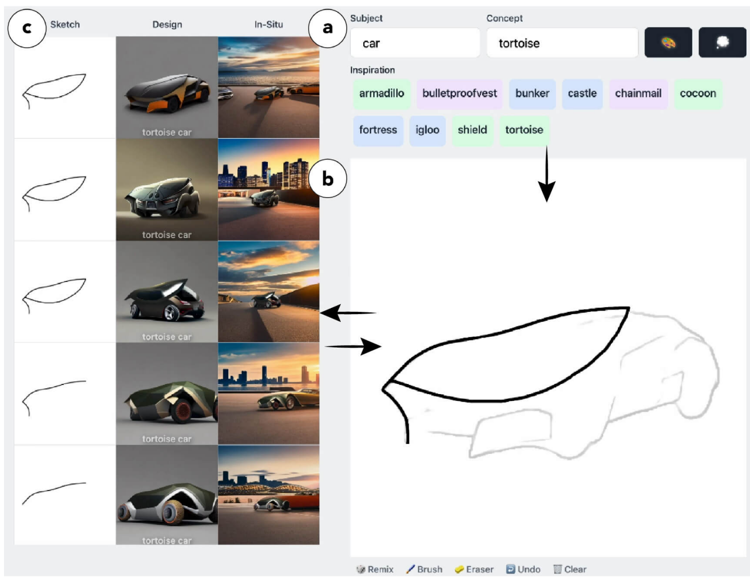


Fig. 1. Inkspire supports designers to prototype product designs through sketching. Using Inkspire, designers can ideate analogical design concepts (a), generate designs through sketching and transform them back into sketches (b), and visualize designs in realistic settings. Designers can view how the designs have evolved over time (c).

CCS Concepts: • **Human-centered computing → Human computer interaction (HCI)**.

Additional Key Words and Phrases: generative AI, sketching, iterative design

## 1 INTRODUCTION

We have seen significant progress in the capabilities of text-to-image (T2I) models, many of which are now able to synthesize ultra-realistic images using text [2, 9, 12]. These models not only accelerate the process of converting thoughts into visuals but also create serendipitous moments for users to be pleasantly surprised. Recent research has also opened up new possibilities for translating one image representation into another, such as transforming a sketched drawing into detailed designs or transforming skeletal poses into full-body portraits, enabling an additional channel of

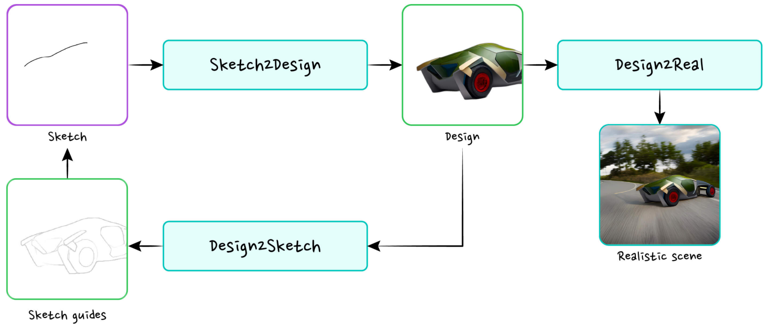


Fig. 2. Inkspire system overview, including Sketch2Design, Design2Sketch, and Design2Real. Inkspire completes the iterative sketch-to-design-to-sketch feedback loop and visualizes designs situated in realistic scenes.

control for image generation beyond text [15]. Consequently, many designers have begun embracing the use of T2I models to support and enhance their creative work.

However, despite the capabilities T2I models unlock, their integration into the designer's creative process is not without its challenges. To understand these challenges, we conducted a comprehensive day-long exchange session with automotive designers, discussing how T2I models can be incorporated into their work. They outlined their design process, from conceptualizing a car design based on a specific theme to producing realistic renderings of the vehicle in various scenes for presentation to stakeholders, highlighting both the benefits and limitations of T2I models. Overall, we identified three primary challenges:

- C1 Designers often begin a design project from an abstract core theme, such as a car that embodies a sense of "protectiveness." However, they found that T2I models struggle to interpret such abstract concepts, leading to unsatisfactory results.
- C2 Designers emphasize that iteration is crucial to the design process. Yet, they feel that the experience of using T2I models is a one-way process. They continually generate new designs, akin to a slot machine, but find it challenging to build upon previously generated ones.
- C3 Designers often end a design project by situating their design in a realistic scene, which helps stakeholders visualize the design more clearly. However, we observed that designers created car renderings in a modeling engine and then manually constructed a surrounding scene in a laborious process by creating and placing individual elements such as roads, trees, and pedestrians.

To address these challenges, we created Inkspire. Using Inkspire, designers can brainstorm analogical design ideas from abstract themes (C1), generate designs with T2I models through sketching and transform previously generated designs back into sketches to continue building on top of them (C2), and visualize designs in realistic settings (C3).

## 2 INKSPIRE

The Inkspire system consists of three main components: Sketch2Design, Design2Sketch, and Design2Real (Figure 2).

### 2.1 Sketch2Design

The *Sketch2Design* component helps users brainstorm design concepts and generate product designs through sketching (Figure 3). First, the user specifies the subject that they are designing for (e.g., car) and an initial abstract concept (e.g.,

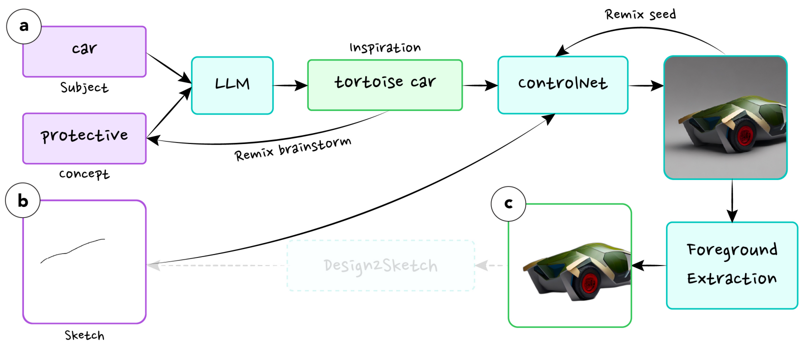


Fig. 3. Sketch2Design pipeline, including (a) concept generation, (b) sketch-guided design generation, and (c) foreground extraction.

protective). To brainstorm more concrete design ideas for the abstract concept, we leverage Large Language Models (LLMs) [3] to generate analogical inspirations (Figure3a). Prior work has found that breaking down a problem into smaller steps can improve the correctness of the LLM's responses [13]. We thus take a two-step approach. We first prompt the LLM to detail the design principles for the given subject: Describe the key design principles in <subject> design in one short paragraph. This also serves as context [7] for the LLM to generate more well-thought-out inspirations in the Step 2. Given the design principles, we then prompt the LLM to generate inspirations from the domains of nature, architecture, and fashion: You are a <subject> designer. The design principles in <subject> design are as follows: <design principles from Step 1>. Brainstorm analogical inspirations for <subject> design to convey a sense of <concept> from one of the following domains: nature, architecture, or fashion. We empirically found prompting specifically for the domains of nature, architecture, and fashion to lead to particularly interesting inspirations (Figure 1a). The user may select a recommended inspiration and continue branching out to explore further inspirations.

After selecting a design inspiration, the user may create product designs by sketching on a canvas. The user may start off with as little as a single stroke (Figure 3b). Using ControlNet [15] to guide Stable Diffusion [12], we generate a product design guided by the stroke. The user may continue adding additional strokes. Each time a stroke is drawn, we generate a new design, making the creation process iterative and implicitly encouraging users to focus on sketching instead of on engineering text prompts (i.e., the current paradigm of working with T2I models). We maintain the same initial seed between generations to maintain consistency and for speedy generations. The user may click the "remix" button to change to a different seed and generate varied designs. Finally, we remove unnecessary backgrounds from the generated design using a foreground extraction method [10] (Figure 3c).

### 2.2 Design2Sketch

The *Design2Sketch* component helps users build on top of previously generated designs by converting them into sketches (Figure 4). Given the converted sketch, the user may visualize it as an underlay on their canvas, similar to tracing paper [1]. This enables the user to draw inspiration from aspects of the previously generated designs and also helps them overcome the challenge of starting with a blank canvas [5], especially during the early stages of sketching. The user can continue iterating through sketching, completing the sketch-to-design-to-sketch feedback loop. We introduce a novel approach for converting designs to sketches. First, we perform semantic segmentation on the design [6]. Given the segmentation map, we then draw the boundaries (Figure 4a). Second, we extract soft edges from the design [14]

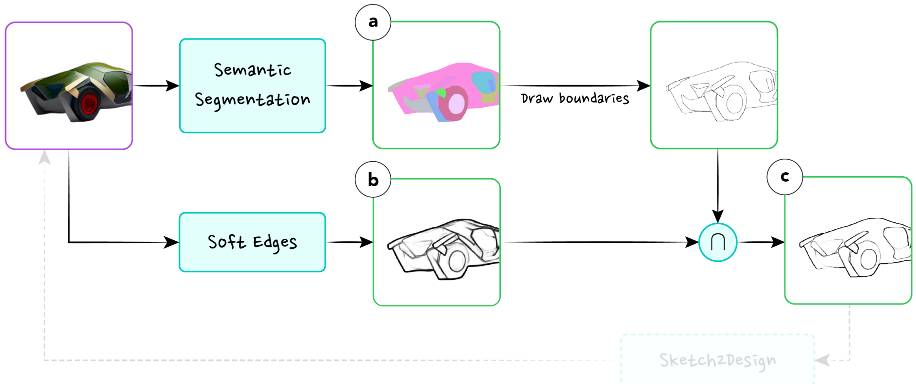


Fig. 4. Design2Sketch pipeline, including (a) semantic segmentation, (b) soft edge extraction, and (c) finding their intersection.

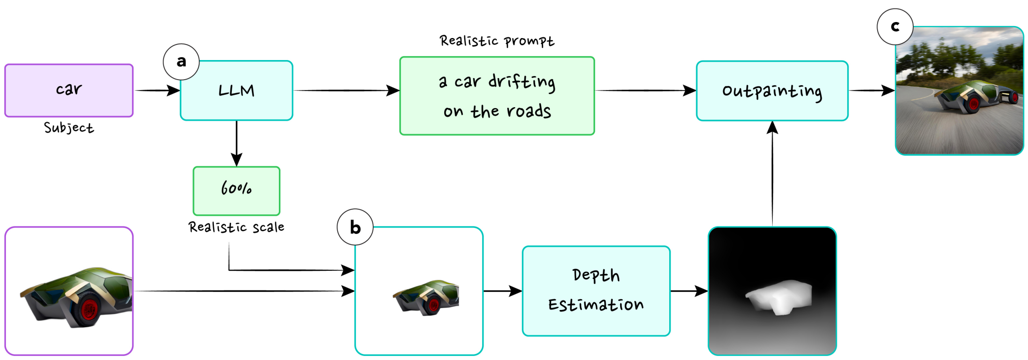


Fig. 5. Design2Real pipeline, including (a) realistic prompt and scale generation, (b) depth estimation, and (c) outpainting.

(Figure 4b). Finally, we take the intersection between the segmentation map boundaries and the extracted soft edges as the sketch (Figure 4c). This approach combines the best of both worlds from segmentation and edge extraction: It focuses on the design's main structural lines and eliminates the redundant lines caused by texture, which is a limitation of edge extraction methods [4, 14], while creating varying thickness and opacity of lines to achieve a sketch-style look.

### 2.3 Design2Real

The *Design2Real* component helps users visualize how a generated design would look in a realistic environment (Figure 5). The component includes a pipeline of LLM prompting, rescaling, depth estimation, and outpainting. First, we prompt the LLM to describe the subject in a realistic scenario (Figure 5a): `Describe <subject> in its natural setting in one short sentence.` We also prompt the LLM to reason an appropriate scale for the main subject: `How much of the picture would <subject> typically cover in a photograph? Give a percentage.` We rescale the design accordingly (Figure 5b). Second, we generate a depth map of the design with a depth estimation method [11]. This serves as an approximation of the design's 3D structure. Finally, given the depth map and the prompt describing the subject in a realistic scenario, we outpaint the image (Figure 5c). Given the 3D approximation from the depth map, we are able to plant the design into a scene with realistic lighting and shadows.

## 3 CONCLUSION AND FUTURE WORK

In this short paper, we introduce Inkspire, a system for designers to prototype product designs through sketching. Our next step is to evaluate our system with designers. Figure 1 shows our user interface. For future work, it may be interesting to explore extending 2D designs into 3D models [8] and enabling sketch-based control of 3D models. We hope Inkspire can support designers to more effectively co-create with T2I models.

## REFERENCES


[1] 2024. *How to Learn to Draw by Tracing.* Retrieved February 16, 2024 from https://monikazagrobelna.com/2020/08/16/how-to-learn-to-draw-by-tracing/

[2] James Betker, Gabriel Goh, Li Jing, Tim Brooks, Jianfeng Wang, Linjie Li, Long Ouyang, Juntang Zhuang, Joyce Lee, Yufei Guo, et al. 2023. Improving image generation with better captions. *Computer Science. https://cdn. openai. com/papers/dall-e-3. pdf* 2, 3 (2023), 8.

[3] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.

[4] John Canny. 1986. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence* 6 (1986), 679–698.

[5] Caneel K Joyce. 2009. *The blank page: Effects of constraint on creativity.* University of California, Berkeley.

[6] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. 2023. Segment anything. *arXiv preprint arXiv:2304.02643* (2023).

[7] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems* 33 (2020), 9459–9474.

[8] Ruoshi Liu, Rundi Wu, Basile Van Hoorick, Pavel Tokmakov, Sergey Zakharov, and Carl Vondrick. 2023. Zero-1-to-3: Zero-shot one image to 3d object. In *Proceedings of the IEEE/CVF International Conference on Computer Vision.* 9298–9309.

[9] 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).

[10] Xuebin Qin, Zichen Zhang, Chenyang Huang, Masood Dehghan, Osmar R Zaiane, and Martin Jagersand. 2020. U2-Net: Going deeper with nested U-structure for salient object detection. *Pattern recognition* 106 (2020), 107404.

[11] René Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, and Vladlen Koltun. 2020. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *IEEE transactions on pattern analysis and machine intelligence* 44, 3 (2020), 1623–1637.

[12] 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.

[13] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems* 35 (2022), 24824–24837.

[14] Saining Xie and Zhuowen Tu. 2015. Holistically-nested edge detection. In *Proceedings of the IEEE international conference on computer vision.* 1395–1403.

[15] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2023. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision.* 3836–3847.