SKETCH2PROTOTYPE: RAPID CONCEPTUAL DESIGN EXPLORATION AND PROTOTYPING WITH GENERATIVE AI

Kristen M. Edwards*
Massachusetts Institute of Technology
Dept. of Mechanical Engineering
kme@mit.edu

Brandon Man*
Massachusetts Institute of Technology
Dept. of Mechanical Engineering
bm557@mit.edu

Faez Ahmed
Massachusetts Institute of Technology
Dept. of Mechanical Engineering
faez@mit.edu

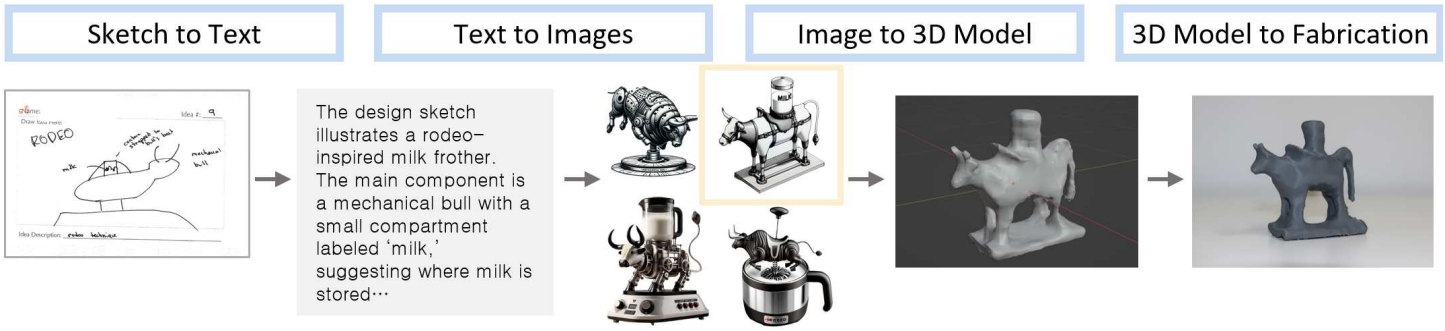

FIGURE 1: The Sketch2Prototype framework takes in a conceptual design sketch as input, and produces multiple inspired images, a 3D model, and finally a fabricated prototype.

ABSTRACT

Sketch2Prototype is an AI-based framework that transforms a hand-drawn sketch into a diverse set of 2D images and 3D prototypes through sketch-to-text, text-to-image, and image-to-3D stages. This framework, shown across various sketches, rapidly generates text, image, and 3D modalities for enhanced early-stage design exploration. We show that using text as an intermediate modality outperforms direct sketch-to-3D baselines for generating diverse and manufacturable 3D models. We find limitations in current image-to-3D techniques, while noting the value of the text modality for user-feedback and iterative design augmentation.

INTRODUCTION

During product design and development, a design concept moves through many modalities. It may be represented as a sketch, a textual description, a looks-like or works-like prototype, and finally materialize as a finished product [1]. Early in the engineering design process, sketches and prototypes are pivotal for conveying ideas, investigating various design possibilities, and exploring the design space [2]. Developing a looks-like prototype is an essential step in the engineering design process, offering a tangible, visual representation of a product idea, and communicating the design concept to stakeholders. However, prototyping can be time-consuming and resource-intensive, involving multiple iterations and manual adjustments to achieve the desired outcome [3].

In the phase of conceptual design, it's typical to progress from sketching to prototyping in a linear fashion. Yet, research indicates that tackling these activities concurrently can offer significant advantages [2]. Given that conceptual design determines up to 70-80% of a product's lifetime cost [4, 5] exploring the design space with proper breadth and depth is indeed valuable. Despite this, sketching and prototyping are often done in sequence because sketching is quicker and has lower overhead than prototyping [1].

Recent breakthroughs in generative AI have enabled people to generate novel, unseen images by learning underlying patterns in training data. Through generative models, the looks-like prototyping process can be streamlined, allowing for rapid generation and iteration of design options, thus significantly reducing time and costs associated with manual methods. This enables design space exploration, motivating designers with a diverse set of examples. Moreover, incorporating machine learning in the prototype development process opens the door for enhanced user interaction and feedback through easy iterations, as shown later in Figure 7. Furthermore, communication between the development team can be bolstered by having a physical manifestation of their design vision. Ultimately, the integration of machine

[1]These authors contributed equally to this work.

learning in creating looks-like prototypes facilitates a more efficient, cost-effective, and user-centered approach to product development, aligning technological innovation with aesthetic and practical design needs.

In this work, we propose Sketch2Prototype, a framework for understanding sketches, generating new conceptual images inspired by those sketches, converting the images to 3D models, and finally fabricating a looks-like prototype from these 3D models. There are various existing models that can perform each of these subtasks, we demonstrate several state-of-the-art methods. We found that GPT-4V(ision), a vision language model by OpenAI, 2023, is able to interpret and explain hand-drawn sketches [6]. Therefore, we use GPT-4V, to convert sketches into textual prompt descriptions, then DALL-E 3, which is a generative text-to-image model, generates a set of more descriptive images from the text. We then use those images to generate a 3D model which, after postprocessing, we fabricated via additive manufacturing. We demonstrate Sketch2Prototype in a series of case studies with real hand drawn sketches of milk frothers, phone stands, pen and coin holders, and mugs. Our method enhances design space exploration by three means: 1) inherent design expansion caused from automatically generating multiple 2D images inspired by one sketch 2) increased breadth and depth of exploration made possible by working with sketches and prototypes in parallel [2], and 3) allowing for user-centered feedback via the text modality. The contributions of this work are as follows:

1. We introduced a generative AI-based framework to rapidly create a prototype from a sketch, enabling design exploration as the design moves through sketch, text, image, and 3D modalities.
2. We compared our framework to direct sketch-to-3D and ControlNet-generated image-to-3D frameworks. Our model generates more diverse images and manufacturable designs.
3. We demonstrated examples of the successful Sketch2Prototype, moving from a hand drawn sketch to a fabricated 3D looks-like prototype for four design categories and six designs.
4. We built an open-source dataset of 1,087 milk frother sketches each with four paired images inspired by the sketch and generated by our framework. This results in 4,348 images.

1 RELATED WORK

In the following sections we discuss related works regarding sketching and prototyping in engineering design, recent advancements in vision language models, image-to-3D models, and 3D representations.

1.1 Sketching and Prototyping in Engineering Design

Sketching is documented as a valuable skill in engineering design, and researchers have studied ways to encourage and understand sketching in engineering education [7, 8]. Sketching provides a rapid external representation that comes at very little cognitive cost [9]. Researchers also explored creativity and decision making with sketches [10], using sketching for finite element analysis [11], and using machine learning to predict creativity-ratings from sketches and text [12, 13].

In product design, prototypes can be defined as "an approximation of the product along one or more dimensions of interest" [1]. Prototypes, as well as the process of building and testing them, offer invaluable information to designers [8]. As such, many works have surveyed and explored different prototyping strategies [14,15]. Research has explored how combining sketching and prototyping during conceptual design impacts design space exploration. On average, only sketching leads to a broader design space and generated more novel designs, only prototyping leads to more aesthetically pleasing designs with better functionally, both sketching and prototyping explored and generated final ideas that were perceived as more creative [2]. Sketching can lead to a higher quantity of designs [16], but prototyping can be used to both explore and refine designs [17]. Furthermore, this work suggests benefits of using both sketch and prototype modalities during conceptual design to explore the design space with both breadth and depth and ultimately generate creative designs.

One challenge that prevents designers from prototyping in parallel with sketching is that prototypes are often slower to create and have higher associated costs than sketching [3], and designers are often reluctant to spend money and time on things when uncertainty is high, like early in the conceptual stage [1]. This is where we believe the integration of machine learning to efficiently create looks-like prototypes presents a transformative opportunity.

1.2 Large Language Models and Vision Language Models

Large language models (LLMs) are billion-parameter transformers that are pre-trained on significant amounts of data, enabling them to perform a wide variety of natural language processing tasks such as translation, summarization and recognition. LLMs such as LLaMA [18] learn to generate text aligned with human preferences through Reinforcement Learning with Human Feedback.

Vision language models (VLMs) have also become quite popular. To create cohesive understanding between image and text, CLIP (Contrastive Language-Image Pre-training) [19] is a model that creates a joint embedding space between vision and language. CLIP is an efficient metric for measuring similarity between an image-text or image-image pair [20]. Hence, VLMs

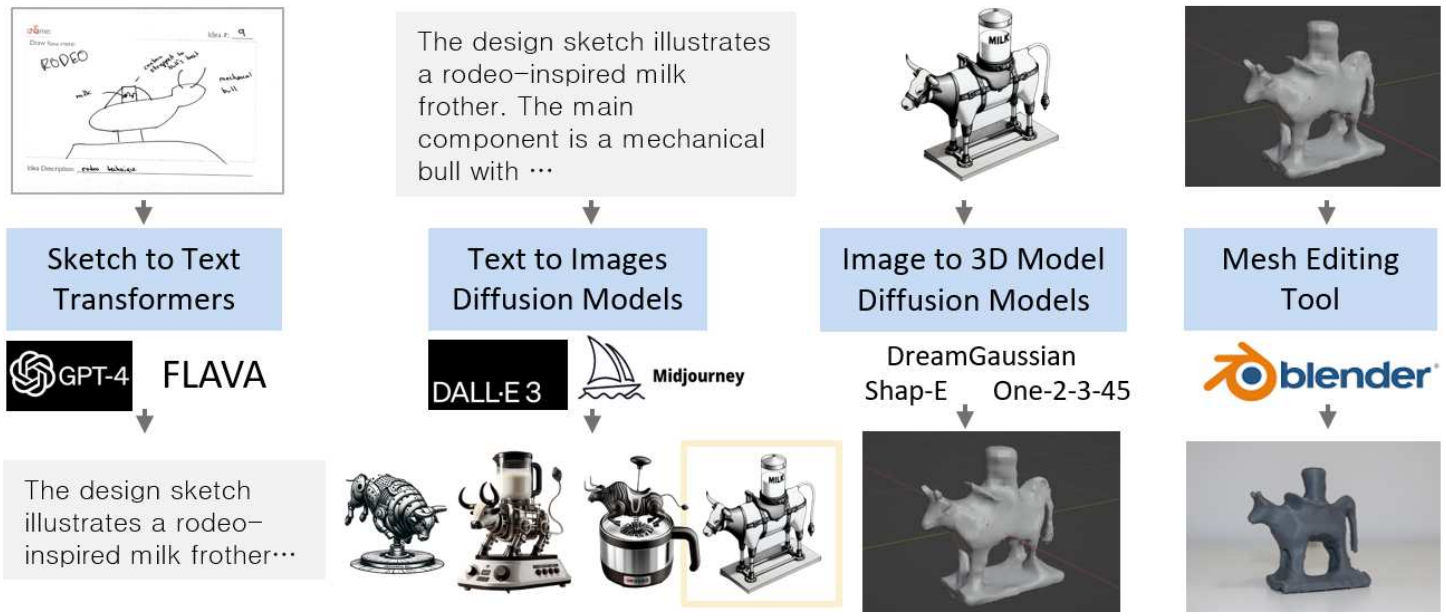

FIGURE 2: The Sketch2Prototype framework uses transformer-based models for the sketch-to- text and text-to-image steps, as well as an encoder and conditional diffusion model for image- to-3D model. Post-processing of the 3D model is performed in Blender.

such as GPT-4V and FLAVA [21] are similar to LLMs except that they train on multimodal datasets and often employ CLIP or similar methodologies learn a joint space between language and vision. VLMs need to understand complex relationships between text and image, thus requiring multimodal data for training. Although there is a certain degree of overlap in their applications, VLMs are distinctively advantageous for tasks necessitating visual comprehension, in contrast to LLMs, which excel in purely text-based endeavors. Given the inherently multimodal nature of early-stage design, VLMs are particularly well-suited for such applications.

Text-to-image synthesis has been explored via models like Imagen [22], DALL-E 3 [23]. Many researchers have used diffusion models for text-to-image tasks as diffusion models tend to be faster. Many text-to-image models also enable users control over their picture by prompting the model to change specific regions of an image via text, known as inpainting. However, inpainting for these models is often limited to text.

To enable users to control image generation with their own sketches, models such as ControlNet and T2IAdapter have emerged. These models freeze the text-to-image model and use the users' sketches to guide the text-to-image generation process. However, these models give too much control to the users, and often results in images too similar to the users' sketch, meaning less exploration of the design space, as we show later, these models lead to fixation on the original concept. By leveraging VLMs' ability to provide text descriptions, which can then be expanded by text-to-image models, we generate a highly diverse set of prototypes.

1.3 3D Representations and Image-to-3D Models

3D representations enable users to visualize the dimensions and proportions of objects. Furthermore, 3D representations enable manufacturing decisions, by displaying different geometric constraints. Lastly, 3D prototypes enable useful feedback from designers, customers, and stakeholders regarding the user-experience with a design [8]. Neural Radiance Fields (NeRFs) [24] have become a popular method for 3D scene representations. Although NeRFs are used in 3D reconstruction [25] and generation [26] , optimizing NeRFs are time consuming to train and memory intensive. 3D Gaussian splatting [27] is a recent alternative to NeRFs and has demonstrated promising results in both speed and quality in 3D reconstruction. Recent work has tried applying Gaussian Splatting to generation tasks [28] that outperform methods that use NeRFs for 3D representations.

Image-to-3D models try to generate 3D assets from a single image, which can also be reformulated as a single-view 3D reconstruction tasks, but often produce blurry results [29]. Using image captioning models, text-to-3D methods can be adapted for image-to-3D generation [30]. Dream Gaussian [28] is a recent model that uses companioned mesh extraction and texture refinement in UV space. Even though the results of Dream Gaussian are promising, it fails to produce high quality models of unseen models. Shap-E [31] is another recent image-to-3D and text-to-3D model that utilizes a conditional diffusion model to output

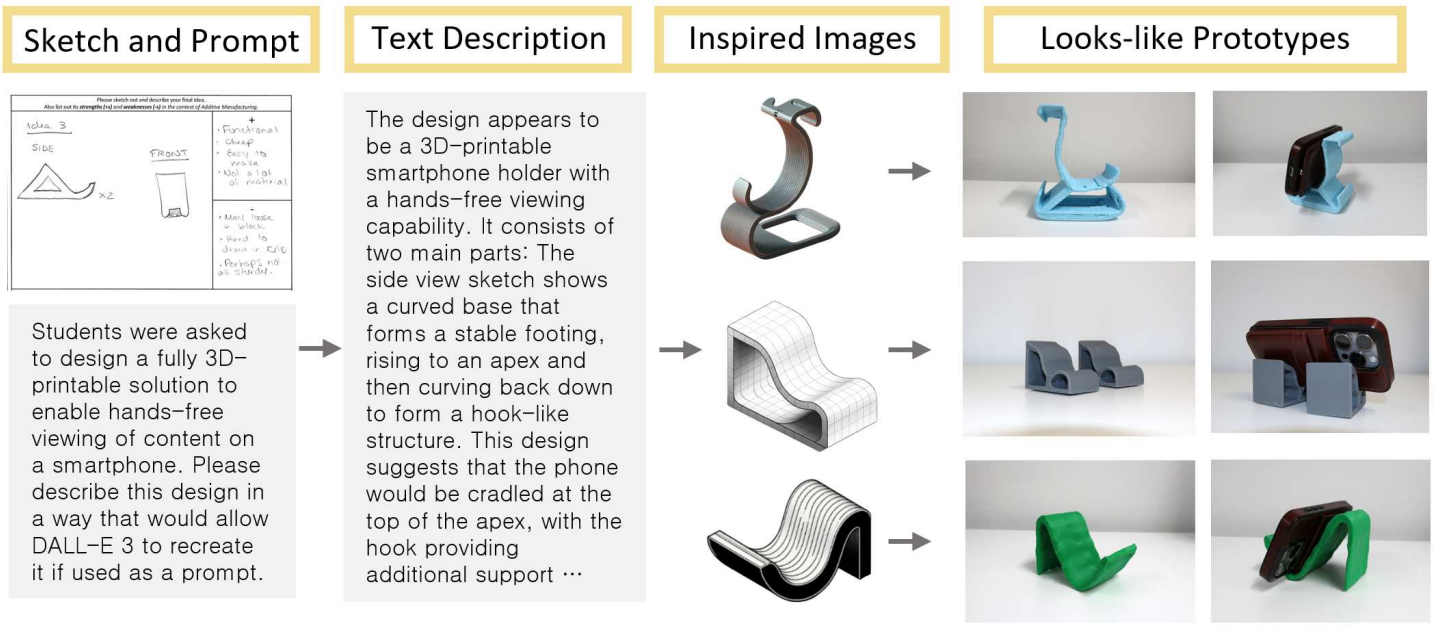

FIGURE 3: Our framework enables exploration of the design space by automatically generating multiple diverse images inspired by one sketch. Here a single sketch results in three fabricated looks-like prototypes.

high fidelity 3D objects.

In engineering design, it is often time consuming for designers to create high-quality images that can be used for image-to-3D generation tasks. Sketches are often abstract, lack detail, and are often unfit for image generation. Past work has tried to predict 3D functionality from a 2D image; however, 3D information performs best [32]. Our work shows an end-to-end system that generates multiple high-quality images from the original sketch, which can then generate a printable 3D prototype.

2 METHODOLOGY

In the following sections, we expound on the multi-stage process where a sketch is transformed into text, then to images. We further discuss the post-processing and fabrication stages, where 3D models are refined in Blender to meet fabrication standards and subsequently 3D printed to materialize the design concepts.

2.1 Framework from Sketch to Prototype

Our proposed framework treats the Sketch2Prototype problem as a sequence of tasks that move between design modalities: from sketch-to-text, then from text-to-image(s), and finally from image-to- 3D model, as shown in figure 2. For sketch-to-text, we fed our sketch as input along with a verbal description of the sketch into GPT-4V and prompted it to give it a description of the image. We also asked GPT-4V to describe it such that it will be passed as a prompt into DALL-E 3. DALL-E 3 performs text-to-image by converting the text description of the original sketch into a text embedding, then feeding it into a diffusion prior to generate an image embedding, which finally gets decoded into an image. We chose not to extract the words found on the sketches as they may give semantic meaning to specific areas of the image. To generate a variety of novel designs, we ask DALL-E 3 to generate 4 images from the original prompt. DALL-E 3 also attempts to generate more diverse images by rephrasing the input prompt.

The resulting image generated from DALL-E 3 may contain text, which negatively affects the generation quality when converting from image to 3D. To prevent this, we manually select a set of images that do not include any text and feed it into our image-to-3D model. Current state-of-the-art models such as Shap-E [31] and DreamGaussian have varied performance depending on the provided image. As a result, we feed our images into three state-of-the-art models (One-2-3-45, DreamGaussian and Shap-E) and pick the mesh that is most similar to the original image while also being the most manufacturable.

2.2 Post-processing and Fabrication

While models like Shap-E excel at generating 3D models, they may not adhere to fabrication requirements. Hence, we perform post-processing of the 3D model in Blender 3.6.5. Shap-E outputs a 3D model in the Polygon File Format (PLY) family. This can be directly imported to Blender 3.6.5, post-processed as

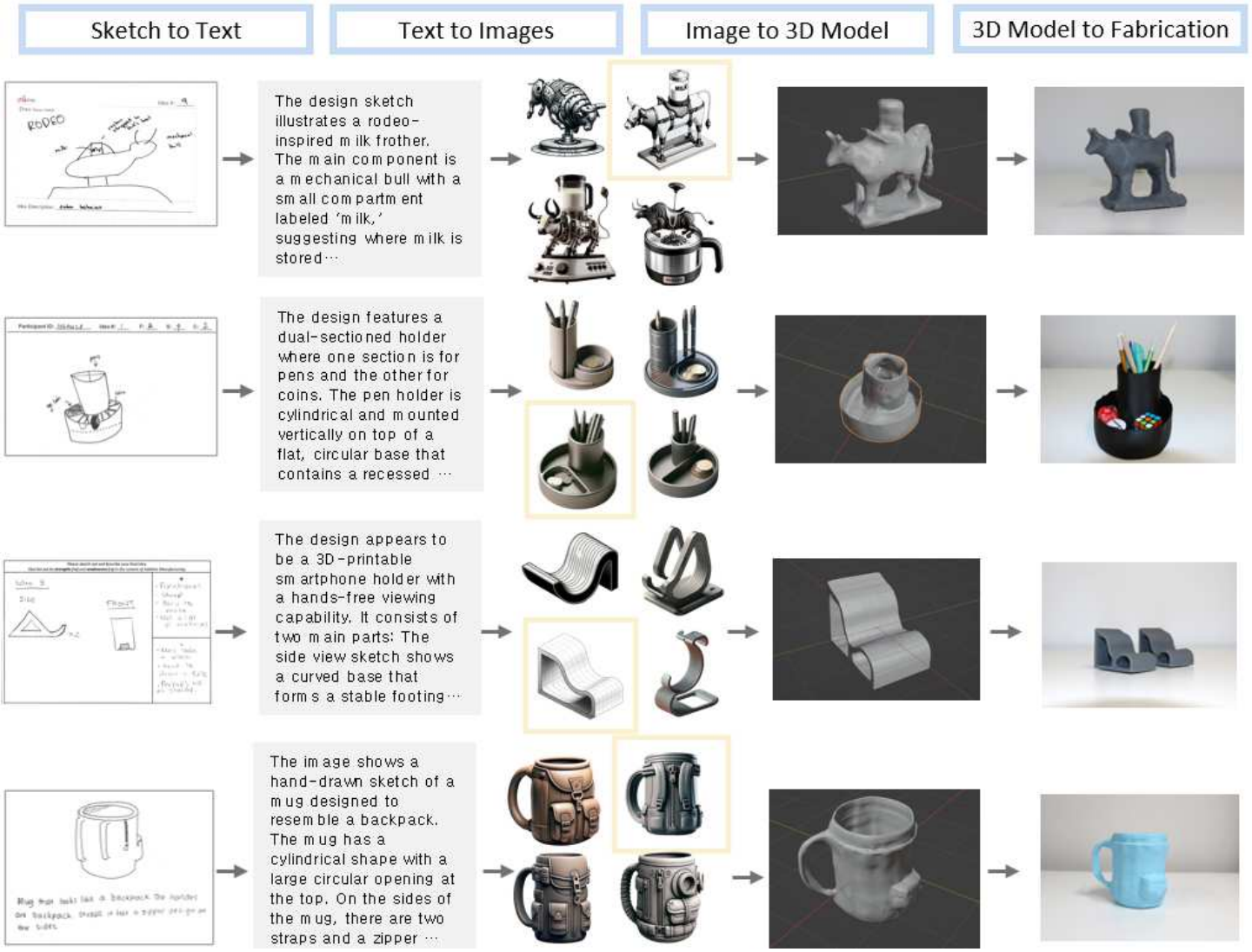

FIGURE 4: Examples of the full Sketch2Prototype framework for four different design types: rodeo-inspired milk frother, pen-and-coin holder, phone stand, and backpack-inspired mug.

needed, and exported as an STL file, which can be 3D printed. We print the models with either the Bambu Lab X1-Carbon Combo 3D Printer, which is a fused deposition modelling printer, or Formlabs Form 3 Printer, which is a stereolithography printer.

3 RESULTS

In this section, we display the results of our framework via a number of examples. Figure 3 demonstrates how the Sketch2Prototype framework enables exploration of the design space. From one design sketch of a phone stand, our framework leads to three diverse prototypes. The images and 3D printed models show a diverse set of functional phone stands. Figure 4 showcases the full Sketch2Prototype framework for four different design types: a rodeo-inspired milk frother, a pen-and- coin holder, a phone stand, and a backpack-inspired mug. For each of these, we demonstrate an automatic exploration of the design space in the text-to-image step. Here, using a text description created from a sketch via generative-AI, a designer is automatically presented with any number of detailed design images inspired by the sketch. We chose to display four images for each sketch, however there is no imposed limit on this. A benefit of design exploration in this stage is that it mitigates design fixation, and can thus be used as an assistant for designers. Furthermore, showing a designer multiple diverse examples can aid in creative ideation. To assess the diversity and feasibility of our designs,

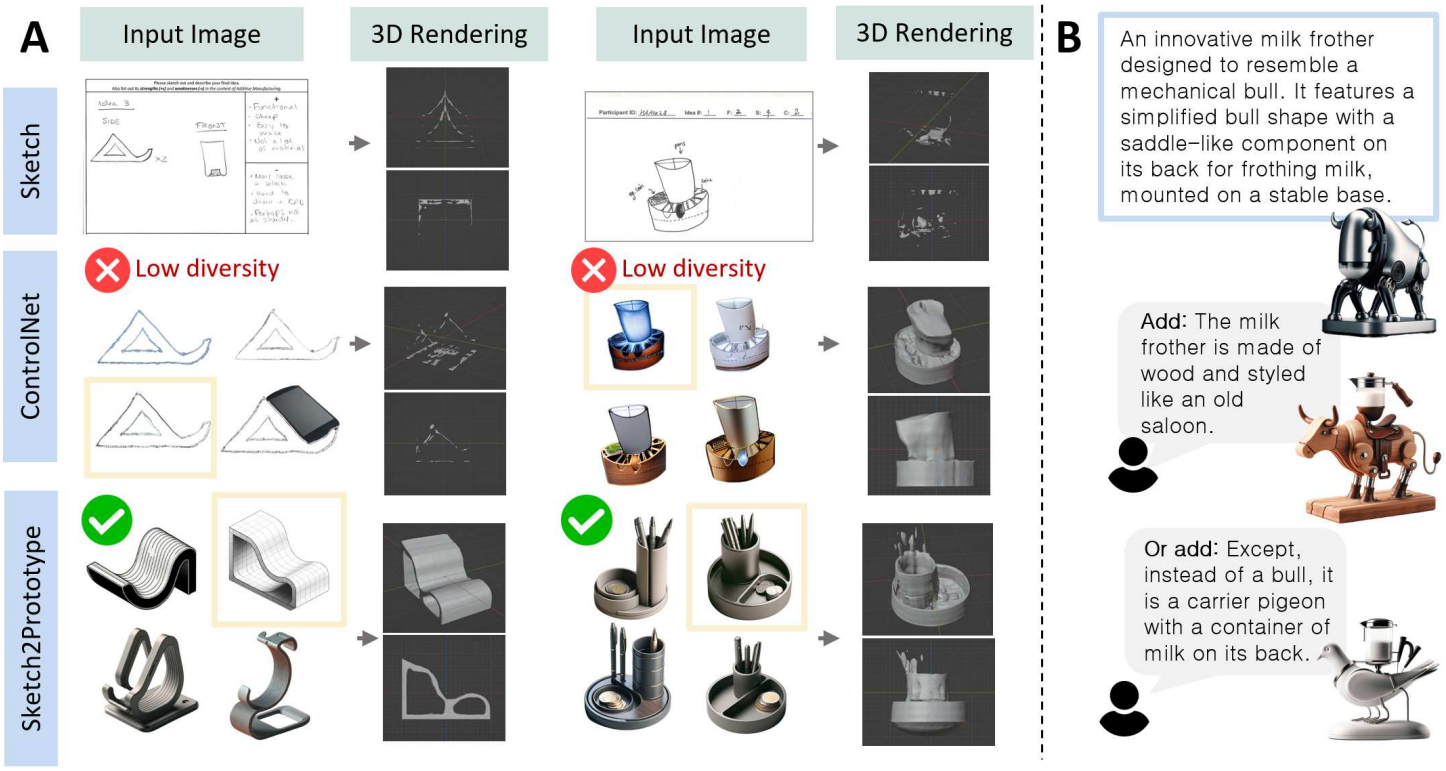

FIGURE 5: **A:** 3D models generated from varying input images: Sketch2Prototype generates more diverse and manufacturable designs. **B:** The text modality allows for user control. We append text to the original prompt to generate different designs.

we compare the diversity and manufacturability of our model's designs to those made from a sketch alone or using ControlNet, which adheres to the sketch geometry (Figure 5A).

3.1 Enhanced diversity and manufacturability over baselines

We perform a qualitative evaluation of Sketch2Prototype by generating 3D models with two baseline approaches. The first approach is directly passing an unprocessed sketch into Shap-E to generate the resulting mesh. The second approach is passing our sketch into ControlNet to generate 4 candidate images. We then pass each image into Shap-E and, for standardization, select the first generated mesh. For Sketch2Prototype, we also generate 4 candidate images and perform the same mesh selection process. We test our method on a phone stand design and a pen-and-coin holder. Sketch-to-text via GPT-4V, text-to-image via DALL-E 3 or ControlNet, and image-to-3D via Shap-E each take a matter of seconds, so the time difference between these three approaches is negligible.

Results are shown in Figure 5A. We can see that the meshes generated from unprocessed sketches are sparse and unmanufacturable. For ControlNet, the generated images lack diversity. In the case of the phone stand, due to the simplicity of the sketch, this process also produces unmanufacturable designs. The ControlNet generated images lack diversity due to directly matching the sketch geometry. Finally, our model generates diverse variations of each sketch. Our model also generated functional and cohesive meshes for prototyping.

3.2 Generation and exploration via human-in-the-loop feedback

We also evaluate the controllability of Sketch2Prototype via the text modality. In Figure 5B, we add sentences to the DALL-E 3 prompt to alter the output image according to designer feedback. For example, when the original prompt is appended with The milk frother is made of wood and styled like and old saloon," the output image changes form accordingly. The intermediate text modality thus enables users to add iterative feedback, allowing for extra user control. Hence, there is immense value in using text as an intermediary modality to help with exploration and addition of new requirements, which are difficult to update in sketch directly.

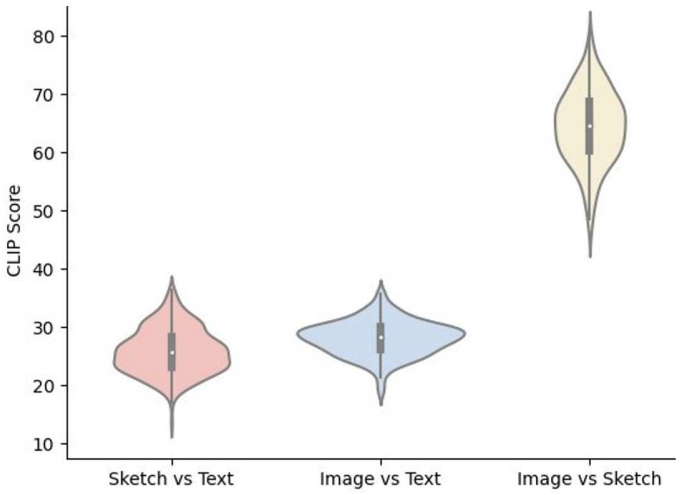

FIGURE 6: CLIP scores between Sketch and Text, Image and Text, and Image and Sketch modalities for sets of the sketch, text, and images corresponding to the same design. A high CLIP score indicates a high level of similarity and alignment between the respective pairs.

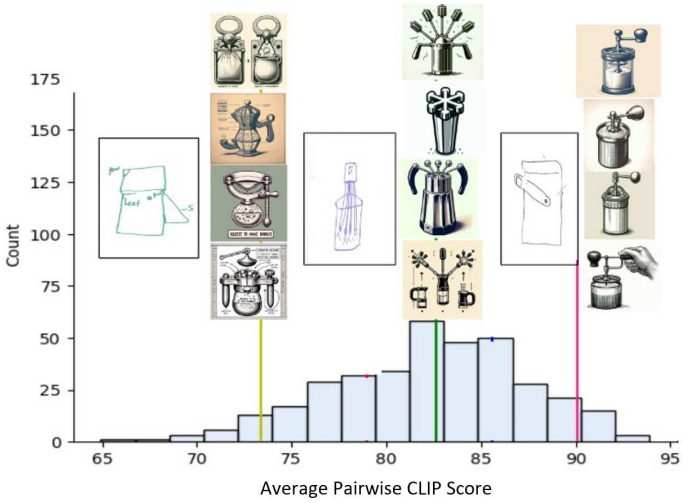

FIGURE 7: Average pairwise CLIP score between the four generated images. A lower CLIP score indicates a more diverse set of images. Yellow, red, green, blue and pink lines correspond to 5th, 50th, and 95th percentiles respectively. The original sketches are shown to the left of the set of images generated from them.

3.3 Dataset alignment and diversity

We show quantitatively that Sketch2Prototype generates images that match the original description while being diverse. We generated a synthetic dataset from a set of milk frother designs and evaluated its alignment with the original dataset and its diversity. This dataset of milk frothers contains 359 hand-drawn sketches of milk frothers drawn by unique individuals. Each milk frother design often has a brief text description of the design concept in addition to text annotations on the sketch. To generate the synthetic dataset, we provided the image to GPT-4V and incorporated the text description into prompt that asks GPT-4V to generate a DALL-E 3 prompt. For each sketch, we generate 4 images, giving us a synthetic dataset of 1,436 total images.

For our generated dataset, we show 1) that the generated images align with their sketches, and 2) that the generated images represent a diverse set, thus exploring the design space. To show our synthetic dataset aligns with the original dataset, we compute the CLIP score between each sketch in the original dataset and the provided text description. We also take the average CLIP score between the 4 images in the augmented dataset with the text description. Finally, we take the CLIP score between the sketches and the images. The results are listed in the plot on Figure 6.

The resulting average CLIP score for the sketch vs text, image vs text and sketch vs image sets are 25.8, 28.1 and 64.4 respectively. The higher average between the image vs text and sketch vs text is expected, since GPT-4V gave a more detailed text representation for generation compared to the original sketch. Likewise, CLIP scores between two images are generally higher than CLIP scores between image and text, so it is expected that the image-sketch CLIP score is higher than the remaining two.

To measure diversity, we compute the average pairwise CLIP score for each of the 4 images in each sample. Higher CLIP scores indicate less diverse datasets. Figure 7 illustrates the distribution of these pairwise CLIP scores. At the 5th, 50th, and 95th percentiles, we show an example of a sketch and four generated images to give better context on how these scores correspond to the diversity images. As percentile increases, we observe a decrease in geometric variation and fewer components. At the 95th percentile, the milk frother designs shown have almost identical geometries - every design are variations of a cylindrical container attached to a handle. However, designs at the 5th percentile have varied shapes and multiple components, such as handles of different curvatures, and different containers.

4 DISCUSSION

4.1 Integrating AI into the Design Thinking Process

In this work, we proposed a framework, Sketch2Prototype, that utilizes generative AI to rapidly generate a textual description, a diverse set of 2D images, and 3D models from a hand drawn sketch. We compare our framework to two baseline models, using sketch-alone and using ControlNet-generated images. We showed that our framework generates more diverse designs and manufacturable models than the others (Figure 5A). Sketch2Prototype also increases the breadth and depth of explo-

ration by allowing designers to work with sketches and prototypes in parallel. We exhibit the entire framework, resulting in six fabricated prototypes from four hand-drawn sketches.

We also demonstrate how user feedback can be incorporated via the text modality. Our framework represents a design as a sketch, text, image, and 3D model. The intermediate text modality can be easily edited by a designer to add more requirements that are not present in original text. This allows iterative refinement and improvement, shown in Figure 5B. The two results shown in Figure 5 demonstrate the balance between user-control and automatic design expansion. While ControlNet allows for strict geometric adherence to an input image, this eliminates most diversity and may not be desirable when dealing with an imprecise hand-drawn sketch. On the other hand, DALL-E 3 generates very diverse designs, however iterative feedback via text may be necessary to generate a desired image.

Tools such as Sketch2Prototype can be incorporated into traditional design thinking processes, enhancing ideation, and prototype development. Sketch2Prototype enables engineers to rapidly explore the design space by expanding simple, abstract sketches into diverse images and 3D printable looks-like prototypes. Our framework allows for human-in-the-loop feedback in the text-to-image phase, and facilitates prototype development by converting these images into 3D models.

4.2 Limitations within the existing framework

We must emphasize that this is a preliminary exploration of how emerging AI tools may benefit designers via accelerated design transformation from sketch to text, images, and 3D models. The 3D models are meant to be looks-like, not functional, prototypes that can be built using additive manufacturing techniques. Further studies are necessary to determine which AI tools perform best at each step of the framework, and, in fact, define what "best" means.

The process of combining different design transformation steps means that deviance from the desired design can be introduced at every step. In the sketch-to-text phase, the user benefits from having much control via simple text editing. However, we observe a failure case in sketch-to-text then text-to- image when GPT-4V generates a text description that is deemed an "unsafe" prompt by DALL-E 3. This indicates that GPT-4V's text generation is not grounded on DALL-E 3's safety mechanisms.

Another limitation is the lack of repeatability in the text-to-image stage. Using off-the-shelf image generation tools such as DALL-E 3 means that the same prompt will not generate the same image when repeated. Furthermore, changes to the design are often global rather than local, even if the textual prompt only requests a local change.

The greatest limitation that we observe is in the image-to-3D stage. Though we utilize state of the art image-to-3D models, design details are often lost at this stage, and 3D models generated using image-to-3D are often non-smooth, fragmented, and sometimes non-manufacturable. As a result, postprocessing is required either to smooth surfaces, fill holes, or remove unmanufacturable parts. Even with postprocessing, image-to-3D models tend to create uneven surfaces or holes, which indicate that NeRFs, while strong for 3D visualization purposes, may not be ideal for 3D printing. Due to the postprocessing step, the time taken for Sketch2Prototype increases.

Finally, aside from manufacturability problems, there may be a lack of control over the final mesh. For earlier steps in the framework, users can edit intermediate representations to better control the product, such as editing descriptions during sketch-to-text or inpainting during text-to-image. The only control users have over a mesh is the postprocessing stage, which is limited to removal or minor edits to surfaces. Enabling users to "inpaint" 3D meshes would give significantly more flexibility over their final product.

5 FUTURE WORK

Future research could explore the application of this framework in more complex design scenarios, such as multi-component systems or intricate structures. These tasks are challenging since models would need to identify distinct parts, understand part-interfaces, and ensure compatibility. This challenge becomes even harder when dealing with sketches that have internal components, as internal components need to have correct proportions with respect to their container in order to fit, and we note that the image-to-3D models we use did not capture internal components well.

Current research on image-to-3D models is often concerned with synthesizing 3D images from objects for visualization purposes, but making these meshes functional is much harder. This work reveals that NeRFs may not be an ideal candidate for representing manufacturable prototypes. An area for exploration may be to create a new representation of meshes specifically for 3D printing purposes.

This work aims to demonstrate how existing AI tools enable transformation from sketch to text, image, and 3D modalities, which can enhance design space exploration. Sketch2Prototype should assist designers in efficiently ideating with different modalities. To this end, future work may include incorporating user-centered design principles in the Sketch2Prototype framework, focusing on intuitive interfaces and user feedback.

6 CONCLUSION

We demonstrate a framework to convert sketches into fabricated prototypes via intermediate steps: sketch-to-text, text-to-image, and image-to-3D. We show that our framework enhances design space exploration by generating a set of diverse 2D images and 3D models from a single sketch, and by enabling de-

signers to work with sketches and prototypes in parallel. We find that using text as an intermediate modality allows for iterative user feedback and enhanced user control. Furthermore, text- to-image-to-3D generates more diverse and manufacturable 3D models than sketch-to-3D baselines. However, manufacturability is still a limitation of current image-to-3D models. The Sketch2Prototype framework gains potential as each step is actively worked on by the machine learning community.

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under Grant No. 2231254 and the NSF Graduate Research Fellowship.

REFERENCES

[1] Ulrich, K. T., Eppinger, S. D., and Yang, M. C., 2020. *Product Design and Development*. McGraw-Hill Education, New York, NY.

[2] Bao, Q., Faas, D., and Yang, M., 2018. "Interplay of sketching & prototyping in early stage product design". *International Journal of Design Creativity and Innovation,* ***6***(3-4), pp. 146–168.

[3] Lauff, C., Menold, J., and Wood, K., 2019. "Prototyping canvas: Design tool for planning purposeful prototypes". In Proceedings of the Design Society: International Conference on Engineering Design, Vol. 1, pp. 1563–1572.

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

[5] Pahl, G., Beitz, W., Feldhusen, J., and Grote, K.-H., 2007. *Engineering Design: A Systematic Approach*. Springer London.

[6] Picard, C., Edwards, K., Doris, A., Man, B., Giannone, G., Alam, M., and Ahmed, F., 2023. "From concept to manufacturing: Evaluating vision-language models for engineering design". *arXiv preprint arXiv:2311.12668*.

[7] Schmidt, L., Hernandez, N., and Ruocco, A., 2012. "Research on encouraging sketching in engineering design". *AI EDAM (Artificial Intelligence for Engineering Design, Analysis and Manufacturing),* ***26***(3), pp. 303–315.

[8] Das, M., and Yang, M., 2022. "Assessing early stage design sketches and reflections on prototyping". *ASME Journal of Mechanical Design,* ***144***(4), p. 041403.

[9] Goldschmidt, G., 2014. "Modeling the role of sketching in design idea generation". In *Anthology of Theory and Models of Design*, A. Chakrabarti and L. Blessing, eds. Springer, London.

[10] Toh, C., and Miller, S., 2016. "Choosing creativity: the role of individual risk and ambiguity aversion on creative concept selection in engineering design". *Research in Engineering Design,* ***27***, pp. 195–219.

[11] Murugappan, S., Piya, C., Yang, M., and Ramani, K., 2017. "Feasy: A sketch-based tool for finite element analysis". *ASME Journal of Computing and Information Science in Engineering,* ***17***(3), p. 031009.

[12] Edwards, K., Peng, A., Miller, S., and Ahmed, F., 2022. "If a picture is worth 1000 words, is a word worth 1000 features for design metric estimation?". *ASME Journal of Mechanical Design,* ***144***(4), p. 041402.

[13] Song, B., Miller, S., and Ahmed, F., 2023. "Attention-enhanced multimodal learning for conceptual design evaluations". *ASME Journal of Mechanical Design,* ***145***(4), p. 041410.

[14] Camburn, B., Viswanathan, V., Linsey, J., Anderson, D., Jensen, D., Crawford, R., Otto, K., and Wood, K., 2017. "Design prototyping methods: State of the art in strategies, techniques, and guidelines". *Design Science,* ***3***, p. E13.

[15] Hansen, C., and Özkil, A., 2020. "From idea to production: A retrospective and longitudinal case study of prototypes and prototyping strategies". *ASME Journal of Mechanical Design,* ***142***(3), p. 031115.

[16] Neeley, L., Lim, K., Zhu, A., and Yang, M., 2013. "Building fast to think faster: Exploiting rapid prototyping to accelerate ideation during early stage design". In ASME International Design Engineering Technical Conferences.

[17] Elverum, C., Welo, T., and Tronvoll, S., 2016. "Prototyping in new product development: Strategy considerations". *Procedia CIRP,* ***50***, pp. 117–122.

[18] Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., and Lample, G., 2023. Llama: Open and efficient foundation language models.

[19] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., and Sutskever, I., 2021. Learning transferable visual models from natural language supervision.

[20] Hessel, J., Holtzman, A., Forbes, M., Bras, R. L., and Choi, Y., 2022. Clipscore: A reference-free evaluation metric for image captioning.

[21] Singh, A., Hu, R., Goswami, V., Couairon, G., Galuba, W., Rohrbach, M., and Kiela, D., 2022. Flava: A foundational language and vision alignment model.

[22] Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E., Ghasemipour, S. K. S., Ayan, B. K., Mahdavi, S. S., Lopes, R. G., Salimans, T., Ho, J., Fleet, D. J., and Norouzi, M., 2022. Photorealistic text-to-image diffusion models with deep language understanding.

[23] Betker, J., Goh, G., Jing, L., Brooks, T., Wang, J., Li, L., Ouyang, L., Zhuang, J., Lee, J., Guo, Y., Manassra, W., Dhariwal, P., Chu, C., Jiao, Y., and Ramesh, A., 2022. Improving image generation with better captions.

[24] Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T.,

Ramamoorthi, R., and Ng, R., 2020. Nerf: Representing scenes as neural radiance fields for view synthesis.

[25] Barron, J. T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., and Srinivasan, P. P., 2021. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields.

[26] Poole, B., Jain, A., Barron, J. T., and Mildenhall, B., 2022. Dreamfusion: Text-to-3d using 2d diffusion.

[27] Kerbl, B., Kopanas, G., Leimkühler, T., and Drettakis, G., 2023. "3d gaussian splatting for real-time radiance field rendering". *ACM Transactions on Graphics,* ***42***(4), July.

[28] Tang, J., Ren, J., Zhou, H., Liu, Z., and Zeng, G., 2023. "Dreamgaussian: Generative gaussian splatting for efficient 3d content creation". *arXiv preprint arXiv:2309.16653*.

[29] Yu, A., Ye, V., Tancik, M., and Kanazawa, A., 2021. pixelnerf: Neural radiance fields from one or few images.

[30] Melas-Kyriazi, L., Rupprecht, C., Laina, I., and Vedaldi, A., 2023. Realfusion: 360deg reconstruction of any object from a single image.

[31] Jun, H., and Nichol, A., 2023. Shap-e: Generating conditional 3d implicit functions.

[32] Edwards, K., Addala, V., and Ahmed, F., 2021. "Design form and function prediction from a single image". In ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference.