Ergonomics In Design and Kansai Engineering, Vol. 170, 2025, 55–64
https://doi.org/10.54941/ahfe1006167

MIX: An Image Generation System Using Image Prompts for Industrial Design

Longfei Zhou[1], Wenxi Sun[1], Yuxin Ding[1], Xinda Chen[1], and Luohe Ni[2]

[1]China Academy of Art, 218 Nanshan Road, Shangcheng District, Hangzhou, Zhejiang Province, China
[2]Central Saint Martins, Granary Building, 1 Granary Square, King's Cross, London N1C 4AA, United Kingdom

ABSTRACT

Generative AI is increasingly being applied in the industrial design field. However, the existing AI design tools are still challenging due to complex operations and parameter settings. To solve these problems, we conducted comprehensive interviews with industry experts and proposed the MIX system. The system emphasizes image prompts and is designed to support designers for more efficient task completion. MIX incorporated both the IP-Adapter and the ControlNet control model with multimodal language models, making more natural interaction possible. It enables designers to quickly produce high-quality design renderings from target reference images. Through qualitative and quantitative user experiments, we found that the MIX system performs well in both efficiency and practicality in design development. Finally, we discuss about when it would be more appropriate to use text prompts versus image prompts in these different design scenarios.

Keywords: Generative artificial intelligence, Image prompts, Product rendering

INTRODUCTION

Industrial design heavily relies on the creativity of designers and their ability to communicate visually. At the initial stage of a design project, designers are tasked with transforming a large number of creative ideas into visual presentations. This process validates the design concept's feasibility and facilitates feasibility and satisfaction assessments with various project stakeholders. The industrial design process primarily consists of four significant phases: (a) concept design, (b) sketch development, (c) 3D modelling, and (d) material rendering. In the traditional design process, the majority of time is consumed during the stages of sketching ideas, modelling structures and appearances, arranging lighting, applying textures, and rendering final effects. The process of visually presenting and verifying design concepts is often time-consuming and resource-intensive.

However, this approach is accompanied by the inherent limitations and complexities associated with expressing visual elements through text.

Furthermore, mainstream AI tools, such as Stable Diffusion, despite having an open-source web UI, require a high level of intricacy in parameter settings. Similarly, the node-based interaction method employed by ComfyUI presents a steep learning curve, making it challenging for designers to rapidly acquire proficiency. While simplified interface versions are available on the market, such as DUIYOU, they still prioritize parameter adjustments, resulting in a less intuitive operation.

To address the aforementioned challenges, the MIX system was developed. MIX is an AI design tool that aims to assist industrial designers in improving efficiency during the product concept phase. Firstly, MIX offers a significant increase in output efficiency during the rendering process. Secondly, compared to mainstream AI tools such as Stable Diffusion and Midjourney, it simplifies the workflow and user interface for generating images through prompts. MIX employs image input prompts to level the playing field with low-barrier-to-entry efficiency and accuracy, delivering predictable, high-quality product renderings, as illustrated in Figure 2.

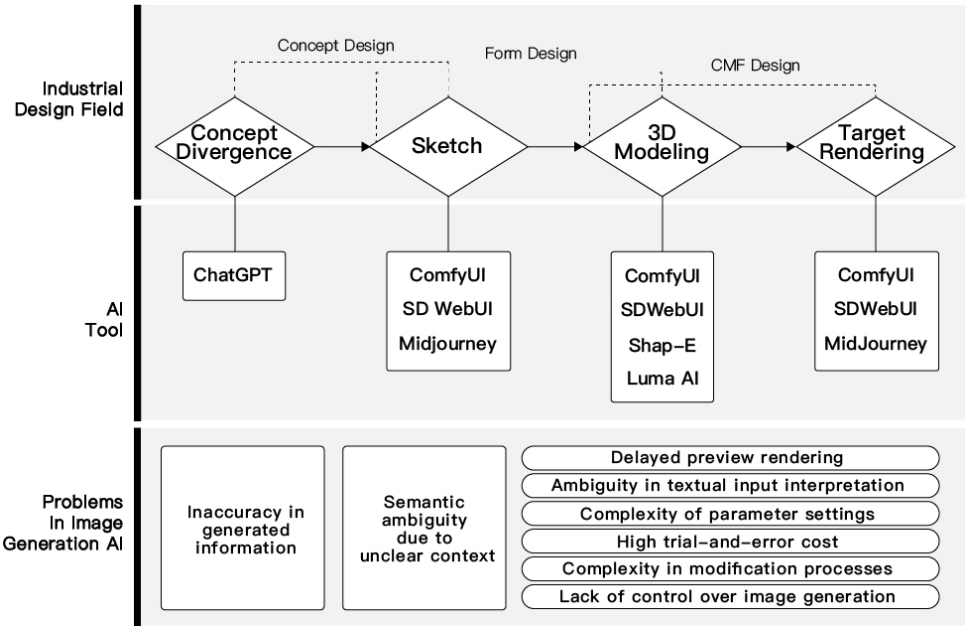

Figure 1: Corresponding AI tools and problems in the industrial design process.

The key innovations of this paper can be summarized as follows:

(1) The practical application of image-prompt-based generation in industrial design. Image prompts provide designers with clear visual references, reducing semantic ambiguity and thereby making image generation more straightforward, resulting in accurate and efficient output.

(2) The natural interaction interface design of MIX. MIX's interactive design reduces its operational complexity and enhances the ease of use of the image generation tool through real-time image feedback, a simplified

image generation process, and optimized workflows, enabling designers to significantly increase the efficiency of their creative visual expression.

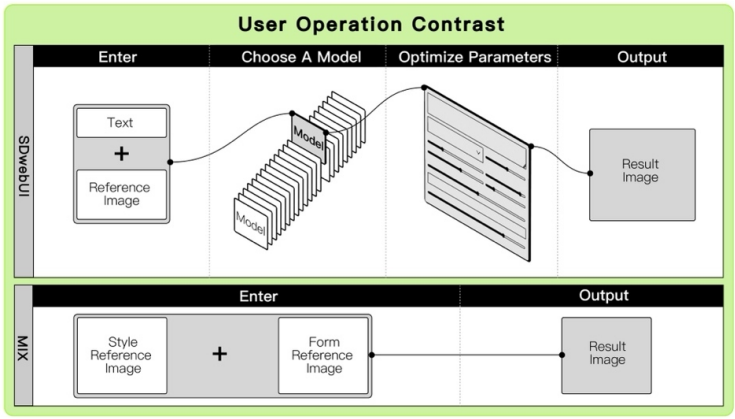

Figure 2: Comparison of webUI workflow and MIX workflow, the advantages of MIX user experience compared to webUI are: image-as-prompts, action-as-response, and WYSIWYG.

RELATED WORK

Since the information revolution began, computer and automation technologies have been widely applied, which has also accelerated the digitalization and automation of design processes, enhancing the efficiency of designers in their workflows. Entering the 21st century, industries have adopted intelligent manufacturing and Internet of Things (IoT) technologies as core components, propelling the intelligent transformation of industrial design processes. This phase not only emphasizes the integration of design, manufacturing, supply chain, and user experience, but also expands the design process to cover all stages of the product lifecycle. Intelligent design leverages big data, AI, machine learning, and IoT technologies to optimize design decisions. For example, Autodesk's Fusion 360 generative design module explores solutions through algorithms, enhancing design performance and manufacturability using machine learning and finite element analysis. Additionally, VR and AR technologies support product testing and demonstration in virtual environments, improving problem-solving efficiency during the design phase. In the context of industrial digitalization, the design process is highly collaborative and real-time, promoting crowd-sourced design and global collaborative design, making design more flexible, rapid, and personalized. With the explosion of artificial intelligence, industrial design is evolving towards intelligence. Overall, industrial design is transitioning from an era driven mainly by manual craftsmanship and experience to one dominated by AI-driven standardization, automation, and digitalization.

Industrial designers utilize image generation models to rapidly create renderings for validation and presentation of their design ideas. These models can assist designers in exploring a wider range of design possibilities. However, their controllability has been significantly restrained due to the limitations of text-prompt-driven frameworks, which have practically rendered them ineffective as design tools, lacking the meaningful and interpretable controls required by users. Most text-based tools generate images that do not align with the designers' intentions and cannot be directly applied; consequently, designers have to invest a considerable amount of time fine-tuning prompts. As text expression is open-ended, the potential outputs are virtually limitless. In workflows dominated by text-prompt AI tools, each iteration of an image necessitates new prompts, a process that often feels arbitrary and lacks systematic control for designers.

DESIGN OBJECTIVE

The goal is to use AI to assist in working together with a designer in an efficient and accurate manner, thereby boosting the work efficiency of the designer in all aspects across the industrial design process. Industrial design generally involves a process through which it gradually changes from conceptual to refined solutions, including constant communication and revisions with its stakeholders. The primary objective of AI-aided design is to minimize human effort while maximizing output, thereby enhancing the creativity and expressiveness of designers, keeping the focus on the designer rather than replacing them. To achieve this objective, the system design needs to focus on the following four key points:

(1) Considering that designers tend to prefer using images rather than text to express their ideas, the system design should focus on visualizing prompts where text descriptions fail to accurately convey design concepts. Additionally, during the design process, designers often spend a significant amount of time and effort in iterative communication with other stakeholders after visualizing their design concepts, making constant revisions. By visualizing prompts, it is possible to reduce misunderstandings from text descriptions and present ideas more intuitively to stakeholders, enabling quicker decision-making.

(2) To meet designers' higher productivity demands, the system design should offer fewer parameters, simplifying the design process and lowering the operational threshold, allowing designers to interact with the system in a more natural, effortless manner without the need for additional learning.

(3) In the design process that moves from concept exploration to product rendering, the product renderings should closely match and reflect the design ideas. Therefore, the system's image generation needs to be more controllable, more precise, and better aligned with the designer's expectations.

SYSTEM DESIGN

To achieve the above design objectives, the MIX system was designed, which is an AI industrial design rendering tool centred on the Image Prompt (IP-Adapter). It innovatively replaces the traditional textual prompts with image prompts to rapidly and accurately render the target image.

MIX is an AI design tool designed to enhance the efficiency of designers. It significantly boosts rendering efficiency in the conceptual image rendering phase of industrial design. Additionally, compared to other AI tools like Stable Diffusion WebUI and Midjourney, it optimizes interaction processes to deliver highly efficient, precise, and predictable rendered images. To achieve this, two key areas were focused on: interaction design and AI workflow.

(1) Interaction Design: MIX's interface integrates four key visual elements in industrial design—colour, material, process, and form—into a single, highly concentrated screen. The interface follows principles such as exposing fewer parameters, using image prompts for input, and offering rapid preview outputs. These principles make the tool intuitive and easy to modify. MIX's primary interaction method is drag-and-drop, resulting in a clean and simple overall interface. The left side of the interface houses the style reference image gallery, while the right side contains the form gallery, which includes product sketch line drawings or white models. The central panel displays the rendered image of the target product, which merges the selected style and form reference images.

(2) AI Workflow: For the AI workflow, a dual-image input system was adopted for style and form, corresponding to two essential industrial design elements: shape and material. Once the two channels are input, components such as the LLM prompt generator, IP-Adapter, and ControlNet work together to synthesize the latent information required for the diffusion model, generating the desired images. After post-processing, the final image is output. The workflow is built around the following principles: a) Simplified and intuitive user input; b) Separation of input parameters into distinct style and form information; c) Intelligent dynamic parameter filling.

MIX IMPLEMENTATION AND EFFECT RENDERING

To ensure reliable image generation based on image prompts,we referred to LLM-grounded diffusion,a multimodal language model-based image descriptor is utilized in combination with the flow-based node structure of ComfyUI to create the image generation functionality. The Product Design (minimalism-eddiemauro) model serves as the foundational base model. As illustrated in Figure 3, MIX's primary structure involves extracting form and style reference information from the input images, which then proceed through the image generation and post-processing workflow.

Image Information Descriptor. Although MIX is an image generation tool based on image prompts, text prompts are still utilized during the intermediate process. These intermediate text prompts do not directly interact with the user; instead, they are automatically generated with the assistance of a multimodal large language model. This step is primarily to ensure that the image model can more accurately comprehend the style and form. In Appendix14, a set of comparisons was conducted between the prompt parser results from style and form reference images, images generated entirely by IP-Adapter's image prompts, those generated solely by text prompts, and images generated by combining image and text prompts. The findings revealed that the combination of both resulted in a superior understanding

of style and form. The image descriptor is composed of three parsers, each employing the GPT-4o model: the style parser, the form parser, and the combine parser. The style and form reference images are processed by the style parser and form parser, respectively, and their outputs are amalgamated in the combine parser to generate the final text prompt.

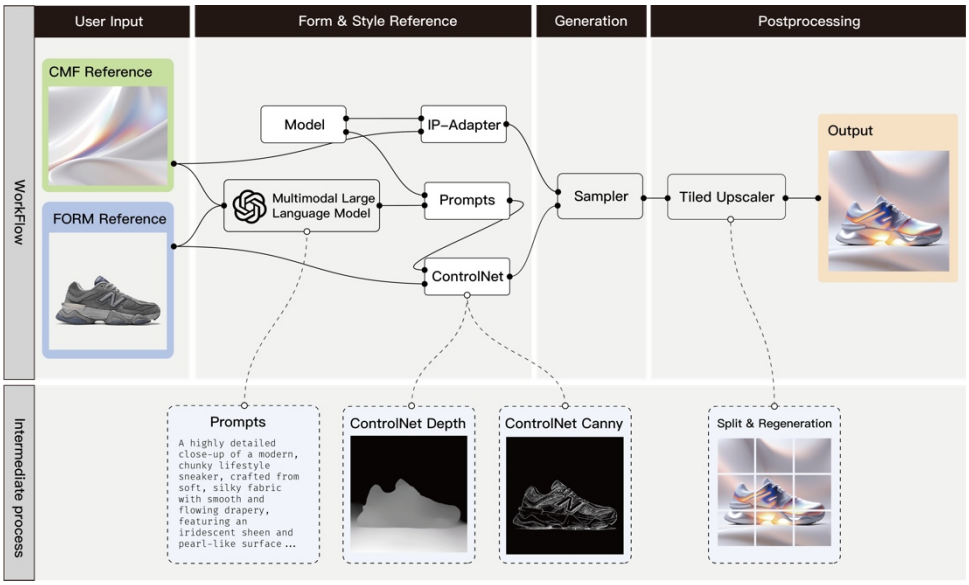

Figure 3: The technical composition of MIX.

ComfyUI Workflow. The workflow of MIX's ComfyUI, as depicted in Figure 4, is primarily composed of four types of nodes: model resource loading nodes, image generation nodes, ControlNet control nodes, and image postprocessing nodes.

(1) During the initialization process, model loading is necessitated. A set of model loaders were incorporated, including: Checkpoint Loader Simple for loading the specified image generation model, VAE Loader for loading the image encoder and decoder, and CLIP Vision Loader for loading the CLIP Vision model, enabling the encoding process between images and text. Additionally, IP-Adapter Model Loader is employed to load the IP-Adapter model, which is crucial for implementing the image prompt.

(2) To achieve a better correlation between the input image and its form, ControlNet control nodes were utilized, including MiDaS-Depth Map Pre-processor and Line Art Pre-processor. The MiDaS-Depth Map Pre-processor pre-processes the input to predict the depth information in the input image, ensuring consistency in the 3D structure of the output image. On the other hand, the Line Art Pre-processor pre-processes the input for the prediction of the line information in the input image, guaranteeing that the shape and contours of the output image match those of the input. Control-Net Apply: It applies ControlNet so that during image formation, it accommodates the pre-processing of the depth map and the line art.

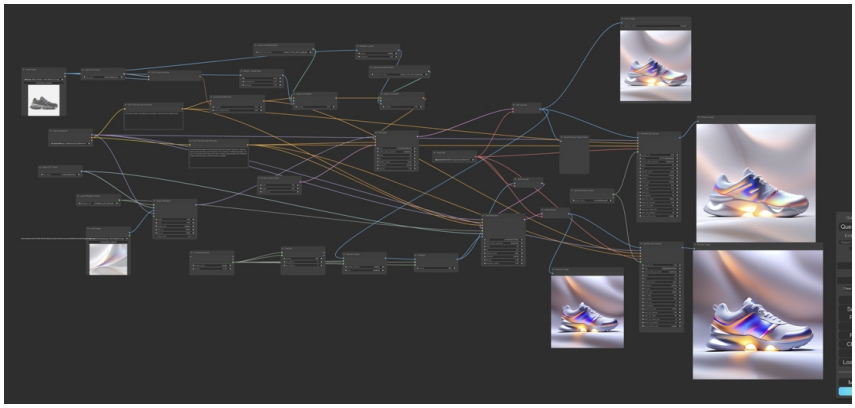

Figure 4: The AI workflow of MIX.

(3) To generate images with a style that matches the CMF (Colour, Material, Finish) output, the image generation nodes include: Empty Latent Image for creating an empty latent image as the basis for further processing, CLIP Text Encode for encoding text prompts and negative prompts to guide image generation, VAE Encode and VAE Decode for encoding and decoding latent images, converting the latent representations into visual images, and K-Sampler, the core image generation node that employs the specified sampler (e.g., Euler), the diffusion model, and both forward and backward text prompts to generate the latent image. Additionally, IP-Adapter Apply is utilized for applying the IP-Adapter to adjust and optimize the generated image, ensuring that the output image closely matches the reference.

(4) The image post-processing nodes serve to enrich the output image's details, including Image-Scale and Tile-Split for enlarging the image and splitting it into tiles, preparing it for large image generation and further processing, as well as Upscale Model Loader and Ultimate-SD Upscale for using an upscaling model to increase the image's resolution, thereby improving overall image quality.

Demonstration of the Effects. Figure 5 shows the effect of using the MIX system for product concept image generation.

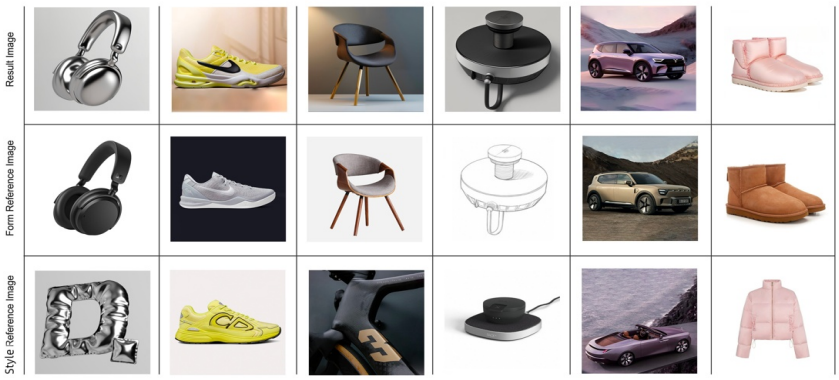

Figure 5: The results of MIX.

CONCLUSION

Simultaneously, this study has some limitations. During the experiment, it was discovered that the guidance within MIX's interface was insufficient, resulting in a lack of user-friendliness for first-time users. In terms of participant selection, efforts were made to ensure that familiarity with both traditional and AI tools was as consistent as possible, but this may have affected the objectivity of the results. Moreover, with the current technical approach, MIX can only generate images as a whole, and it does not yet allow for fine-tuned control or adjustments to specific areas. The quality of the image prompts is also crucial—using material texture image prompts often fails to generate high-quality images.

In future endeavours, the user interface will be further optimized to address the aforementioned limitations and issues. This optimization will encompass enhanced guidance for users and an improved selection process for image prompts. Additionally, aesthetic knowledge will be flexibly integrated into machine learning algorithms to provide interventions and recommendations for uploaded reference images. A model akin to Segment Anything (SAM) is being considered for implementation to layer the form reference images, thereby enabling precise control over the design process.

The contemporary design landscape is markedly distinct from its historical counterpart, primarily due to the emergence of intelligent agent characteristics in modern tools. The human-machine relationship has transcended the simplistic paradigm of user and tool, evolving into a symbiotic collaboration between human wisdom and machine intelligence in this transformative era. As this relationship enters a new phase, the field of design itself is undergoing a profound metamorphosis. Ezio Manzini, the founder of the DESIS (Design for Social Innovation and Sustainability) network, has postulated a future of design characterized by universal participation. This prediction gains credibility as the integration of design and artificial intelligence deepens. The democratization of design empowers individuals not only to become designers but also to exert control over technology. The trajectory of human-machine interaction has progressed from the ergonomic adaptations of humans to machines to the natural language interactions exemplified by ChatGPT. The future of design is replete with expansive horizons and boundless possibilities for advancement. Technological advancements are equipping designers with increasingly diverse creative avenues. While human cognition continues to rely on machine intelligence for certain tasks, it is noteworthy that machines are becoming progressively dependent on human wisdom for guidance and context.

REFERENCES

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).

Alice Cai, Steven R Rick, Jennifer L Heyman, Yanxia Zhang, Alexandre Filipowicz, Matthew Hong, Matt Klenk, and Thomas Malone. 2023. DesignAID: Using generative AI and semantic diversity for design inspiration. In Proceedings of The ACM Collective Intelligence Conference. 1–11.

Amir Hertz, Ron Mokady, Jay Tenenbaum, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. 2022. Prompt-to-prompt image editing with cross attention control. arXiv preprint arXiv:2208.01626 (2022).

Autodesk, Inc. 2024. Generative Design. Autodesk. (2024). https://www.autodesk.com.cn/solutions/generative-design. Retrieved August 29, 2024.

Behance. 2024. Behance. https://www.behance.net. Accessed: 2024-09-01.

Celina Burlin. 2023. Explainability to enhance creativity: A human-centered approach to prompt engineering and task allocation in text-to-image models for design purposes. (2023).

Comfyanonymous. [n. d.]. ComfyUI: A powerful and modular stable diffusion GUI with a graph/nodes interface. https://github.com/comfyanonymous/ComfyUI. Accessed: 2024-09-01.

Donald D Chamberlin. 2000. An adaptation of dataflow methods for WYSIWYG document processing. In Proceedings of the ACM conference on Document processing systems. 101–109.

Eddie Mauro. 2024. Product Design Minimalism - Eddie Mauro. https://civitai.com/models/23893/product-design-minimalism-eddiemauro. Accessed: 2023-09-01.

Henning Kagermann, Wolfgang Wahlster, and Johannes Helbig. 2013. Recommendations for Implementing the Strategic Initiative Industrie 4.0. (2013), 1–82.

Israel Cohen, Yiteng Huang, Jingdong Chen, Jacob Benesty, Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. 2009. Pearson correlation coefficient. Noise reduction in speech processing (2009), 1–4.

J Corney, C Hayes, V Sundararajan, and P Wright. 2005. The CAD/CAM interface: A 25-year retrospective. (2005).

Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising Diffusion Probabilistic Models. arXiv.org (06 2020). https://doi.org/10.48550/arXiv.2006.11239v1

Joo Young Choi, Jaesung R Park, Inkyu Park, Jaewoong Cho, Albert No, and Ernest K Ryu. 2024. Simple Drop-in LoRA Conditioning on Attention Layers Will Improve Your Diffusion Model. arXiv preprint arXiv:2405.03958 (2024).

Long Lian, Boyi Li, Adam Yala, and Trevor Darrell. 2023. LLM-grounded Diffusion: Enhancing Prompt Understanding of Text-to-Image Diffusion Models with Large Language Models. Trans. Mach. Learn. Res. 2024 (2023). https://api.semanticscholar.org/CorpusID:258841035

Marvin Minsky. 1988. Society of Mind. Simon and Schuster.

Marzia Mortati. 2022. New design knowledge and the fifth order of design. Design issues 38, 4 (2022), 21–34.

Merriam-Webster. 2023. Definition of Generative AI. https://www.merriam-webster.com/dictionary/generative%20AI.

Michael Muller, Lydia B Chilton, Anna Kantosalo, Charles Patrick Martin, and Greg Walsh. 2022. GenAICHI: Generative AI and HCI. In CHI Conference on Human Factors in Computing Systems Extended Abstracts. 1–7.

Minsuk Chang, Stefania Druga, Alexander J Fiannaca, Pedro Vergani, Chinmay Kulkarni, Carrie J Cai, and Michael Terry. 2023. The prompt artists. In Proceedings of the 15th Conference on Creativity and Cognition. 75–87.

Reiner Birkl, Diana Wofk, and Matthias Müller. 2023. MiDaS v3.1 – A Model Zoo for Robust Monocular Relative Depth Estimation. arXiv preprint arXiv:2307.14460 (2023).

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 (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 168, 19 pages. https://doi.org/10.1145/3613904.3642824

Szymon Machała, Norbert Chamier-Gliszczyński, and Tomasz Królikowski. 2022. Application of AR/VR Technology in Industry 4.0. Procedia Computer Science 207 (2022), 2990–2998. https://doi.org/10.1016/j.procs.2022.09.357

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 (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 5, 17 pages. https://doi.org/10.1145/3491102.3501931

Tomas Lawton, Kazjon Grace, and Francisco J Ibarrola. 2023. When is a Tool a Tool? User Perceptions of System Agency in Human–AI Co-Creative Drawing. In Proceedings of the 2023 ACM Designing Interactive Systems Conference (Pittsburgh, PA, USA) (DIS '23). Association for Computing Machinery, New York, NY, USA, 1978–1996. https://doi.org/10.1145/3563657.3595977

Vivian Liu and Lydia B Chilton. 2022. Design guidelines for prompt engineering text-to-image generative models. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 1–23.