GenQuery: Supporting Expressive Visual Search with Generative Models

Kihoon Son
kihoon.son@kaist.ac.kr
School of Computing, KAIST
Daejeon, Republic of Korea

DaEun Choi
daeun.choi@kaist.ac.kr
School of Computing, KAIST
Daejeon, Republic of Korea

Tae Soo Kim
taesoo.kim@kaist.ac.kr
School of Computing, KAIST
Daejeon, Republic of Korea

Young-Ho Kim
yghokim@younghokim.net
NAVER AI Lab
Seongnam, Republic of Korea

Juho Kim
juhokim@kaist.ac.kr
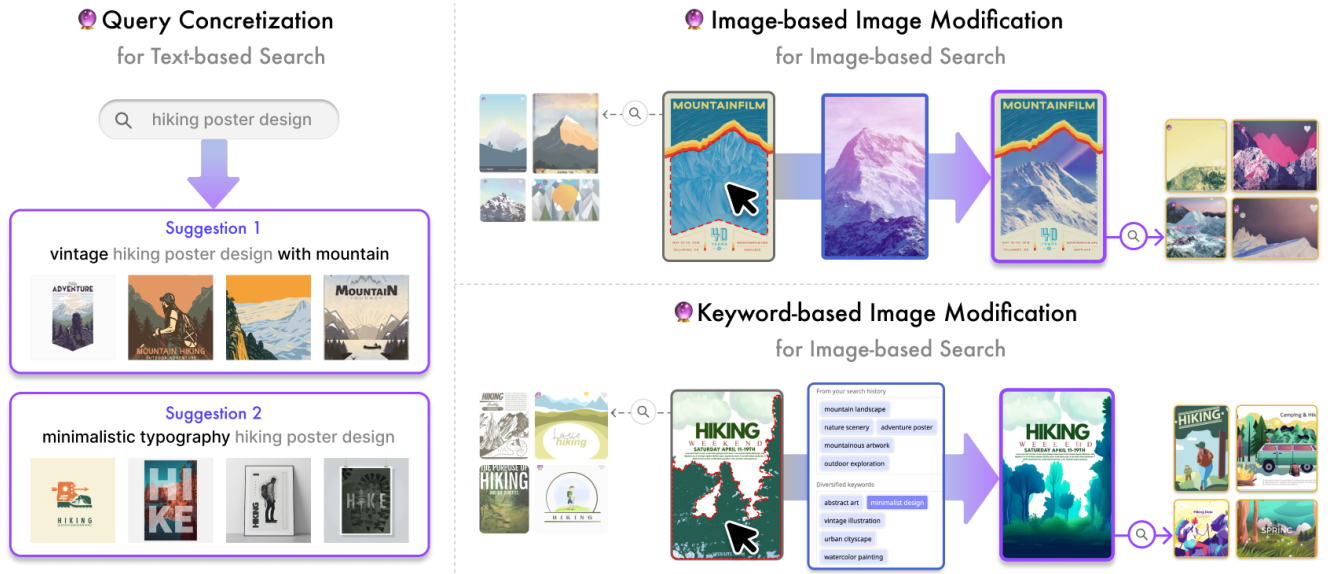School of Computing, KAIST
Daejeon, Republic of Korea

Figure 1: Three main features of *GenQuery*: (1) Query concretization, (2) Image-based image modification, and (3) Keyword-based image modification feature. Query concretization concretizes the user's vague text query for the text-based search. Image-based image modification allows the user to select the area in an image (red dot line above) and to change the area of the image based on the reference image (purple mountain image). Keyword-based modification allows the user to change the selected area of an image (red dot line below) based on the keywords suggested from the user's search history. Both modification features' output could be utilized as an image-based search input and the results are changed due to the image modifications.

ABSTRACT

Designers rely on visual search to explore and develop ideas in early design stages. However, designers can struggle to identify suitable text queries to initiate a search or to discover images for similarity-based search that can adequately express their intent. We propose *GenQuery*, a novel system that integrates generative models into the visual search process. *GenQuery* can automatically elaborate on users' queries and surface concrete search directions when users only have abstract ideas. To support precise expression of search intents, the system enables users to generatively modify images and use these in similarity-based search. In a comparative user study (N=16), designers felt that they could more accurately express their intents and find more satisfactory outcomes with *Gen-Query* compared to a tool without generative features. Furthermore, the unpredictability of generations allowed participants to uncover more diverse outcomes. By supporting both convergence and divergence, *GenQuery* led to a more creative experience.

CCS CONCEPTS

• **Human-centered computing → Interactive systems and tools**.

KEYWORDS

Visual Search; Visual Exploration; Generative Model; Search Intent Expression; Creativity Support; Generative Search

1 INTRODUCTION

Visual search—searching and archiving diverse visual ideas—is an essential activity in the early ideation stage of the design process [49]. In this process, designers find images close to their current ideas by entering a search query, assessing the search results, and setting new search directions for further exploration based on this assessment [19]. Beyond searching for images related to their current ideas (i.e., convergent thinking), designers also aim to uncover diverse and creative ideas through this visual search process (i.e., divergent thinking) [16]. This type of divergent exploration can help designers avoid fixating on specific ideas [19, 22] and drive the generation of more creative ideas [40]. Various tools (e.g., Pinterest[1], Behance[2], or Dribbble[3]) support visual search with two methods: text-based [54] (i.e., inputting a text query to search for relevant images) and image-based search [25, 61] (i.e., inputting or clicking on an image to search for similar images).

However, designers can struggle to sufficiently express their intents (e.g., the type of designs they want to find or explore) through these search methods. For text-based search, it is challenging for designers to concretize their abstract thoughts into concrete search keywords [30, 52]. On the other hand, image-based search partly addresses this challenge by allowing designers to search for designs by using other designs as queries [4, 21, 43]. As these queried images can represent their search intent, the designer can search without having to put their thoughts into words. However, this type of search is typically limited to only retrieving designs that are *similar* according to the tool, meaning that designers can neither explore divergently by searching for dissimilar designs nor designate specific aspects of designs for similarity search [41]. Thus, designers can fail to effectively explore the design space as they do not know how to convey what they want or the tool does not allow them to do.

Through a formative study with eight designers, we probed deeper into designers' intents during visual search, their challenges in expressing these, and how they wish to express them. By asking participants to think-aloud while conducting a visual search, we observed that participants started with abstract intents and concretized these by repeatedly testing various search queries—which could be tedious and time-consuming. After concretizing a search

query, participants explored around the design space by performing image-based search with designs that were most similar to their intents in the search results. However, although participants wanted to explore designs that were similar in terms of fine-grained aspects (e.g., mood, color, shape, or layout), the search tool used only considered the overall similarity of designs. For example, when converging on their search, participants wanted to search for designs that combined aspects from designs they already found and, when diverging, they wanted to search for designs that were similar to a found design in certain aspects but different in other aspects.

To address these problems, this work investigates how generative models—e.g., Large Language Models (LLMs) or Text-to-Image (T2I) models—could support intent expression during visual search. Specifically, we look at the ability of these models to take a rough idea and develop these into more concrete sketches. For example, LLMs can effortlessly draft stories based on a starter sentence [9, 53], and T2I models can produce paintings and illustrations from a set of keywords [8, 35]. In this sense, the generative models can be clear solutions for the aforementioned challenges in visual search: LLMs can expand on and concretize the user's vague intents, and T2I models can query images for the user when none of the search results match their intents. While substantial work has highlighted the limitations of generative models in accurately executing users' intents [27, 29, 59, 64], we see an opportunity in integrating generative models as an intermediate layer in visual search tools: they can expand on users' intents and, as the outputs are leveraged as search inputs, the quality and accuracy of the generations are less consequential. Furthermore, the unpredictability of these generative models can benefit the visual search process it could encourage spontaneous divergent explorations that can prevent fixation [22].

Therefore, we propose *GenQuery* allows the users to concretize the abstract text query, express search intent through visuals, and diversify their search intent based on the search history. First, the user can concretize their vague search query through auto-complete suggestions with more specific search directions in text-based search (Figure 1 Query concretization). Second, the user can edit an image to generate an intent-aligned image as a search input through image-based image modification (Figure 1 Image-based modification). Third, the user can diversify their search intent through keyword-based image modification with the keywords suggested from the search history (e.g., saved visuals or inputted text queries). The modified image can also be used as a search input (Figure 1 Keyword-based modification).

To evaluate *GenQuery*, we conducted a within-subjects study where 16 designers were asked to search for design ideas for two design tasks: ideation for a hiking club recruitment poster and an architecture exhibition poster. They were asked to save a minimum of five designs from the design ideation process using either *GenQuery* or a baseline system similar to Pinterest, which allows only text- and image-based search without generation-based features.

Findings from our study demonstrated that *GenQuery* allowed users to express their visual search intent easily and accurately compared to the baseline. Designers also felt the design ideas they discovered through *GenQuery* were more diverse and creative significantly. Regarding the visual search pattern, the designer reduced the amount of text-based searches by *GenQuery* by 71.2% compared to the baseline. *GenQuery* derived 34.4% of *search by generation*

among the entire image-based search, and designers saved 35.8% of the designs out of all the designs they saved through this new visual search pattern. As a result, designers were able to further develop their ideas in *GenQuery* by roughly generating concepts and exploring similar results to the generated outputs. Based on these results, we discuss the degree of controllability over the generative outputs depending on how the user's search intern is concretized. We also discuss the various types of information for tracking the user's search intent more accurately in the visual search process.

This paper's contributions are three-fold:

- Findings from a formative study: According to the convergent and divergent search stages, the study revealed how designers would express the search intent with what type of modalities and their related challenges.
- *GenQuery*: Design and implementation of a visual search tool that incorporates generative models to enable designers to concretize and express their visual search intents effectively.
- Findings from a user study: The findings demonstrated that generative models can enhance user's accuracy of expression and flexibility in navigation during visual search, leading to more creative and diverse ideas.

2 RELATED WORK

In our work, we aim to support designers' visual search process by aiding them in concretizing and expressing their search intents through generative outputs. To this end, we review related work in (1) the visual search process, (2) prior approaches to support search intent expression, and (3)generative models as an intermediate layer for intent expression.

2.1 Visual Search for Early Design Ideation

Visual search is an essential activity in designers' initial ideation process. During visual search, designers look for and explore existing designs relevant to their ideas and archive identified designs, typically by organizing them as mood boards [11]. This process enables designers to (1) generate, (2) develop, and (3) verify their ideas [2, 11]. In terms of generation, exploring diverse ideas can prevent design fixation [22], where a designer may excessively focus only on a single idea. For development, as designers archive designs during visual search, they can then leverage these resources to develop their own creative concepts by taking inspiration from these ideas [5, 17, 24] or to discuss ideas with other designers [37]. Finally, after developing a design direction, designers can quickly evaluate whether they want to further pursue this direction by observing actual design cases [19]. Thus, visual search is a fundamental part of the design process, and assisting designers to search and explore references more effectively can enhance how they design. Thus, due to the importance of the visual search process in the ideation and development of design ideas, our work aims to support by integrating generative models into the visual search process to further aid designers in concretizing, expressing, and diversifying their search intents.

2.2 Supporting More Expressive Visual Search

Due to the significance of visual search, there has been a wide array of tools that have been proposed to facilitate the process.

Typically, these tools support two main search methods: text-based search [54], where a user inputs a text query to find relevant visuals, and image-based search [61] (i.e., exemplar search [25]), where a user inputs or selects an image to search for similar images. However, there are three main problems associated with these approaches. First, in text-based search, it can be challenging for designers to think of appropriate keywords that will return envisioned designs [4, 12, 19, 21, 32, 52, 63]. Second, while previous work has proposed more intuitive methods for users to input visuals instead of text queries (e.g., sketches [4, 21, 45, 52], wireframes [4] or example designs [12, 43, 63]), these approaches depend on the user having a concrete idea of what type of design they are looking for. Third, as these search tools [21, 30, 32] focus on returning results that are similar to the input data but do not allow the user to control the degree of "similarity", users can struggle to look for results that align with their intents or to explore diverse ideas.

Several methods have been proposed that aim to enhance how designers can express their search intents by handling the limitations of text and example visuals as search inputs. The Composing Text and Image to Image Retrieval (CTI-IR) method [63] and Stylette [26] allows users to express their intents through both visuals *and* natural language, and combines these modalities to retrieve more relevant images or suggestions. Leveraging a designer's previous actions, BIGexplore [50] suggests images by interpreting the user's mouse actions in an exploratory search tool, and Kovacs et al. [31] propose a method that recommends images by interpreting designers' preferences from their previously used styles. Furthermore, if the search results are unsatisfactory, WhittleSearch [32] modifies the search results through the user's natural language commands. As a limitation of existing search tools is that they focus on returning similar results, prior work has also proposed approaches that instead suggest diverse designs. One thread of work focused on extracting the semantic meaning of images and then recommending images that are associated with that semantic meaning [25, 30]. GANSpiration [41] uses a generative model to generate images of GUIs based on an input image randomly, and then searches for diverse GUIs based on their similarities with the generated images.

The aforementioned research has focused on supporting modalities that can help users to accurately express their search intents, methods that help users express their intent more easily, and approaches that allow users to explore diverse designs. However, no method has been proposed so far that can comprehensively address the problems in the visual search process to support the essential goal of the visual search process.

2.3 Generative Models as a Channel for Intent Expression

With the advancements in generative AI models (e.g., LLM or T2I), these models have been increasingly utilized across diverse tasks (e.g., writing [42, 53, 62], education [33], and prototyping [23]).

Specifically, various researchers have investigated how to leverage the generative capabilities of these models to stimulate idea generation and facilitate creative thinking. For example, TaleBrush [9] uses an LLM to quickly draft out a story based on a writer's sketch of a character's fortune. PopBlends [57] helps designers to ideate

blends between products and pop culture references by using an LLM to identify relevant ideas and their combinations. For screenplay writing, Dramatron [38] generates characters, descriptions, and actions based on a writer's outline to help them ideate on how to further progress and develop their screenplay. These work have demonstrated that LLMs can expand and elaborate on users' initial rough ideas, which can stimulate convergent and divergent thinking in creative tasks.

Researchers have actively conducted research on T2I models to provide visual outputs that are aligned with users' intent. As these models can produce high-quality images from a single prompt, various researchers have applied these to various domains, such as art [7], graphic illustrations in news [35], 3D modeling [36], and fictional world-building [10]. However, controlling the generated outputs solely through text prompts can be challenging, which has led to the development of prompting guidelines [34] and systems, like Promptify [3] and RePrompt [58], that support users in composing and selecting prompts for these models. Researchers have also explored novel techniques that allow for inputting different modalities into T2I models to enhance the controllability of their outputs further. These modalities include: reference images [60], rough sketches [56], canny edge [64], or object-segmented images [15]. Beyond modalities, researchers have also proposed techniques that allow users to use these models for purposes beyond simple image generation. For example, there are techniques for stitching multiple objects or images [46, 51], additional modification of generated output based on editing the inputted prompt [14, 20, 39], inpainting or infilling areas inside an image [44, 48], or extending an image by generating outside of its original boundaries [44]. These threads of work have enabled the use of various modalities and interactions that can support more effective control of T2I models. In this work, we investigate how we can integrate these generative techniques into a visual search tool to help users more effectively and efficiently express their search intents.

3 FORMATIVE STUDY

To investigate the limitation of existing visual search tools (Section 2.2) within the actual search process and the user's search intent to the challenges in detail, we conducted a formative study with eight designers using Pinterest[4]. Specifically, this study aims to observe the visual search process for ideation or brainstorming during early design stages. The main goal of this study is to understand the designers' overall patterns during visual search:

- When does the mode change between text-based search and image-based search?
- What is the search type among image- or text-based when they want to find a desired image or explore diverse images?

Next, we also aim to observe the user's visual search intents when they faced the obstacles work Section 2.2 through a think-aloud design and probe them on possible modalities that could support the expression of their intent.

3.1 Participants

We recruited 8 designers (5 female and 3 male; D1-D8) in various design domains: architecture (D1 and D3), branding (D6 and D8), editorial design (D5 and D7), and web/mobile design (D2 and D4). All designers reported to have actively used visual search tools (Pinterest, Behance[5], or Dribbbble[6]) during the early ideation stages of their design process.

3.2 Procedure

We asked participants first to perform a visual search task using a given tool (40 minutes), and then we conducted a semi-structured interview about their experience (20 minutes). We chose Pinterest as the search tool for the study as it is one of the most commonly used exemplar-based visual search tools [25]. It allows users to input a text search query, select images to view similar examples, and filter search results based on high-level keywords.

For the visual search task, we provided three topics: startup landing page design, brochure design with an oriental painting aesthetic, and architectural poster design. Participants conducted the task with a think-aloud explaining their visual search intent (e.g., why they inputted a specific query or clicked on a visual). When participants struggled to fully express the intent verbally, we also asked them to explain themselves by referencing the images shown in the tool. For each search action that they performed, we also asked participants to explain what types of search results they were expecting or wanted to see before they actually looked at the results. After seeing the results, we then asked them to explain whether the results were satisfactory or not, and why this was the case. If the participant gave a non-specific or abstract explanation during the think-aloud, the authors prompted them to describe their intent more concretely through additional questions.

After the task, we asked participants about how they conducted visual search in their actual design work, how well they felt they could express their search intents using the given tool, and whether the system could adequately understand their intent. We also asked participants about possible modalities that could help them express their intent better with visual search tools and about other additional supports that they wanted during the process.

3.3 Visual Search Pattern, Problems, and Findings

The visual search process in an exemplar-based tool consists of a text-based search phase, where the designer inputs keywords to search for images, and an image-based search phase, where the designer clicks on images to explore other similar images. We observed that designers' visual search process follows a general structure (Figure 2).

3.3.1 Iterative Search Query Editing Caused by Vague Initial Intent. Designers initiate the visual search through the text-based search. All participants (D1-D8) initiated their search for the given task by first inputting vague text queries (e.g., landing page design or architecture poster). They explained that they search in this way because they do not know how to formulate their intent into search
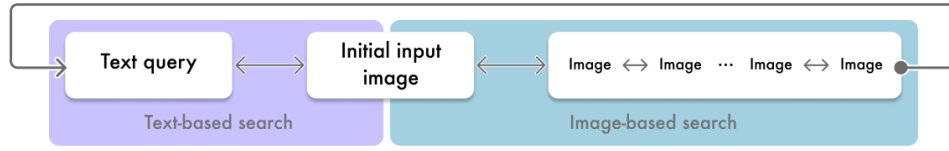
Figure 2: The overall visual search process in an exemplar-based search tool: Designers initiate their visual search with text-based search to find an image that they want to explore. When the designers find an image they would look for, they perform an image-based search by clicking the image. After seeing sufficient similar results with the initial input image (in the middle of an image-based search), they try to explore diverse images through the image-based search by clicking images that are different from the ones seen so far. In this process, when the designers feel they are continuously searching for similar images, they return to the text-based search to explore other images.

terms, and initially, their intent is not clearly defined. Participants said they usually start searching with a vague query and gradually refine it by adding one keyword at a time through repeated text-based searches. In this process, D3 mentioned that finding an appropriate search query can be challenging, as it takes a lot of time and there are instances where they cannot find the adequate query. Although the goal of the text-based search is to find an adequate image to start the image-based search, many participants (D1-D4 and D6-D7) face difficulties doing this with only the text-based search. Participants (D1-2 and D5) mentioned if they could get the concretized query and see the search results before inputting the query, the repetitive process of concretizing their initial search intent would be reduced.

3.3.2 Difficulty in Finding an Intent-Aligned Visual in Image-based Search. Participants initiated an image-based search from a visual found through the text-based search. Among the search results, participants chose the image most closely aligned with their search intent (i.e., most closely resembled what they were looking for).

However, participants (D1, D3-D5, and D8) mentioned how it could be challenging to find an image that adequately expressed their intent. In many cases, most search results did not appeal to the participants (D1-D8). As participants had no other alternatives, they resorted to conducting multiple consecutive image-based searches where they clicked on the image that was closest to their intent until they were able to find images that represented their intent. While participants could not find one image that expressed their intent, we observed that several participants (D2, D4, and D6-D8) explained how multiple images could be combined to express their intent more accurately. D2 and D8 mentioned how, in their actual design work, they edited images in external tools and then used these edited images as references for image-based search.

As a reason for seeking the desired image as precisely as possible, participants (D1, D5, and D8) stated that confirming their envisioned visual elements and examining related search results help determine the next search direction. Ultimately, if they could not find the desired image during this process, they reverted to the text-based search, modifying the text query to initiate a new round of visual exploration.

3.3.3 Diversifying Search Intent for Idea Exploration. When the participants found sufficient images that resembled the visual aligned with their intent, they (D1-D8) started to search for alternatives

that were different in terms of design elements (e.g., color, shape, composition, or layout). This reflects divergent thinking [13], an essential aspect of the creative ideation process [16].

However, even when participants wanted to explore more diverse images, the current visual search tool predominantly only surfaced images with similar styles or overall color moods. When participants felt that the system did not provide diverse results, they would return to the first step of visual search (e.g., text-based search) and restart their exploration.

Interestingly, when they started to look for diverse images, participants (D2-D4, D6, and D8) frequently employed a combination of abstract verbal terms and images to explain what kind of diverse search results they wanted (e.g., "I would like to make this part more modern" by D6). Regarding this, D4 mentioned how it was natural to use verbal explanations to express the types of diverse results due to the abstract nature of natural language. They believed that this ambiguity allowed them to explore a more diversified output.

3.4 Design Goals

Based on the findings, we defined three design goals for visual search interactions:

- DG1: Provide recommendations for concretized search queries and each query's corresponding search results based on the initial query of the user.
- DG2: Support image-based image modification to clearly express the search intent of the user.
- DG3: Support keyword-based image modification to diversify the search intent of the user.

4 *GENQUERY*

We introduce *GenQuery*, a system that leverages generative models to enhance the expression of visual search intent, thereby assisting users in their visual search process. *GenQuery* provide text- and image-based search functionalities.

In contrast to existing visual search systems, *GenQuery* guides users in the early stages of the visual search process to help concretize their intent, particularly during the initial text-based search, where users might have vague text queries. *GenQuery* provides concretized directions by adding related keywords for these vague queries, aiding the concrete formulation of the visual search intent. Next, during image-based searches, when users have a specific intent they wish to explore, *GenQuery* generates visuals representing
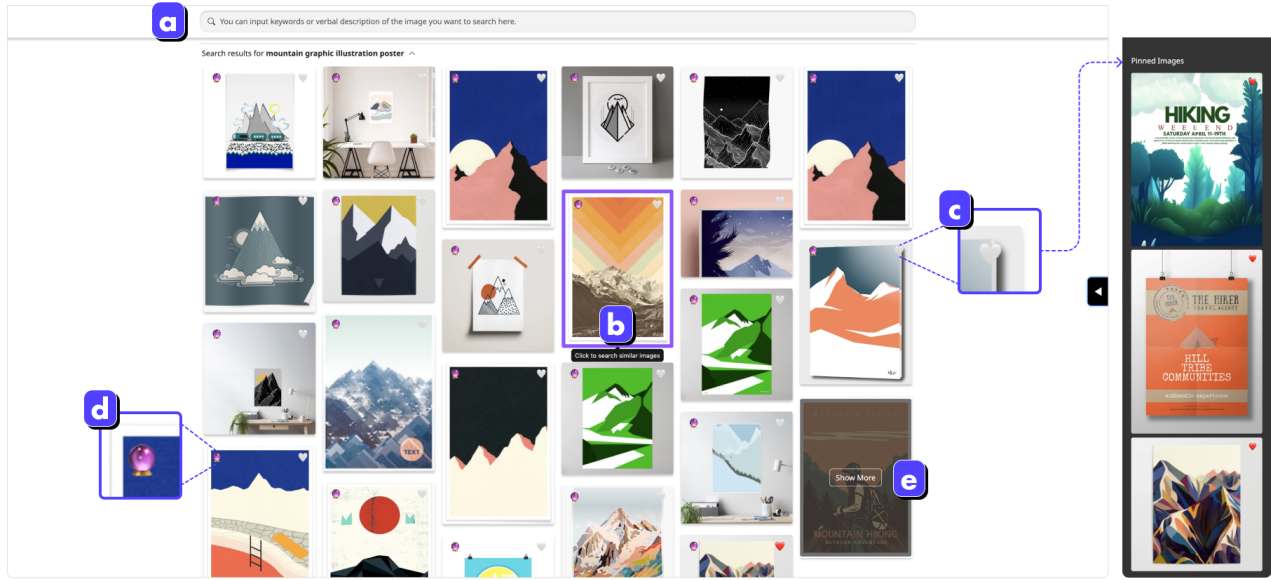
Figure 3: Interface of *GenQuery*. *GenQuery* shows the image search results as a gallery form. (a) Text prompt input box for text-based search: User can input a text description for the desired image here; (b) Clickable image for image-based search: An image in the gallery is clickable to provoke the image-based search. When the image is clicked, *GenQuery* shows similar images to the clicked one at the bottom of the gallery; (c) Like button: The user can click the like button to save the design into the side panel; (d) Generation button: To edit one of the searched images for generating a new input, the user can click the marble emoji left top of image card. When the user clicks this button, the generation panel pops out below; (e) Show more button: This button is clicked when the user wants to see more search results

users' search intent by blending various images on a regional basis. Users can then base their searches on these mixed visuals to pinpoint their desired outcomes. Once the desired visual is found through image-based search, users can further explore diverse visual search results using keywords (e.g., "A more minimalist style"). The system also considers users' search history to comprehend their intent and offers more diversified keywords. This enables users to modify their desired images in a broader spectrum, thus allowing for a richer visual search experience.

Users can utilize *GenQuery* to concretize better (DG1), accurately express (DG2), and diversify (DG3) their search intent using generation-based interactions. Our system allows users to engage in an effective and inspiring visual search process, aligning with their design objectives.

4.1 User Scenario

This section describes how *GenQuery* can assist visual search users in various situations through the process of *Lily* using our system. *Lily* is a student majoring in industrial design and she intends to design a poster to promote her hiking club to which she belongs. Before starting her design work, she decided to use *GenQuery* to explore possible poster designs and to discover creative ideas.

4.1.1 Concretize the Vague Search Query. Firstly, *Lily* begins her search with the keyword *"hiking poster design"* to explore various hiking poster designs. After entering the prompt in the top text search bar, *GenQuery* provides *Lily* with five suggestions on how

to concretize her prompt (Figure 4-a), along with expected search results for each suggestion (Figure 4-b) in Figure 3-a. *Lily* probes the concretized prompts and corresponding search results by pressing the up or down arrow keys (Figure 4-c) to decide what sort of text prompt to input.

4.1.2 Express Visual Search Intent Accurately through Image-based Image Modification. While searching for images and navigating through an image gallery after inputting a text prompt, *Lily* discovers a poster image she likes. She is happy with the overall composition and layout of the poster but doesn't quite like the mountain located in the middle of it. She wants posters depicting mountains of light purple color and finds another design representing the mountain design. Then, she wants to mix these two images to see if there are designs similar to the poster design she's envisioning. To this end, she clicks the generation button (Figure 3-d), and then the generation panel is shown (Figure 5). She then selects the mountain area from the initially discovered poster image (Figure 5-a) and replaces this area with the mountain design above search results or in the side panel (Figure 5-b). Lastly, she clicks the generation button in Figure 5-b to see the generation output. She loves the generation result and proceeds with the image-based search using this generated image to find other similar designs (Figure 5-c). She finds the other images (Figure 5-c1) aligned with her intent.

4.1.3 Diversify Visual Search Intent through Keyword-based Image Modification. After a certain amount of image searches about mountains, *Lily* begins to worry that she has been too focused on
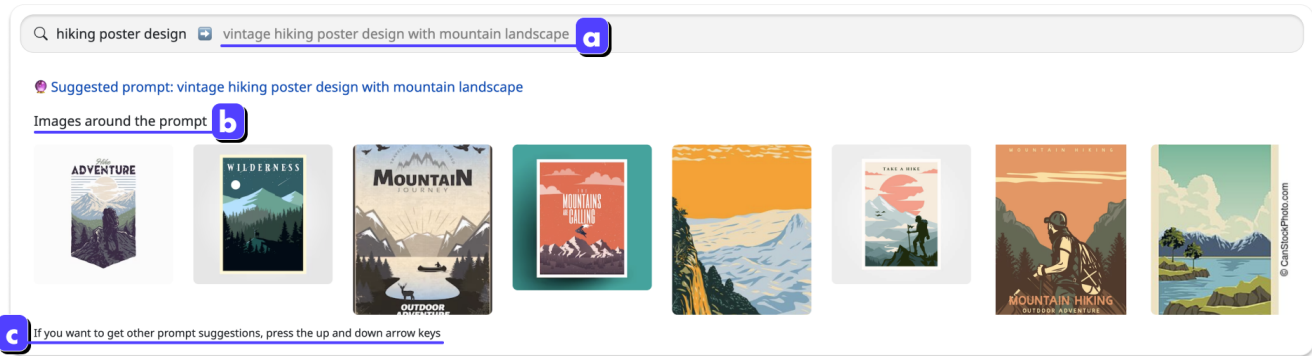
Figure 4: Query concretization, which allows the user to concretize the user's initial abstract search query through LLM zero shot prompting: (a) Suggested search query. User can swap their query by pressing the tab key; (b) Images searched by the suggested query are shown below in the search bar. Each image is clickable to process an image-based search; (c) *GenQuery* provides five suggestions at a time, and the user can explore other suggestions by pressing the up and down arrow keys.
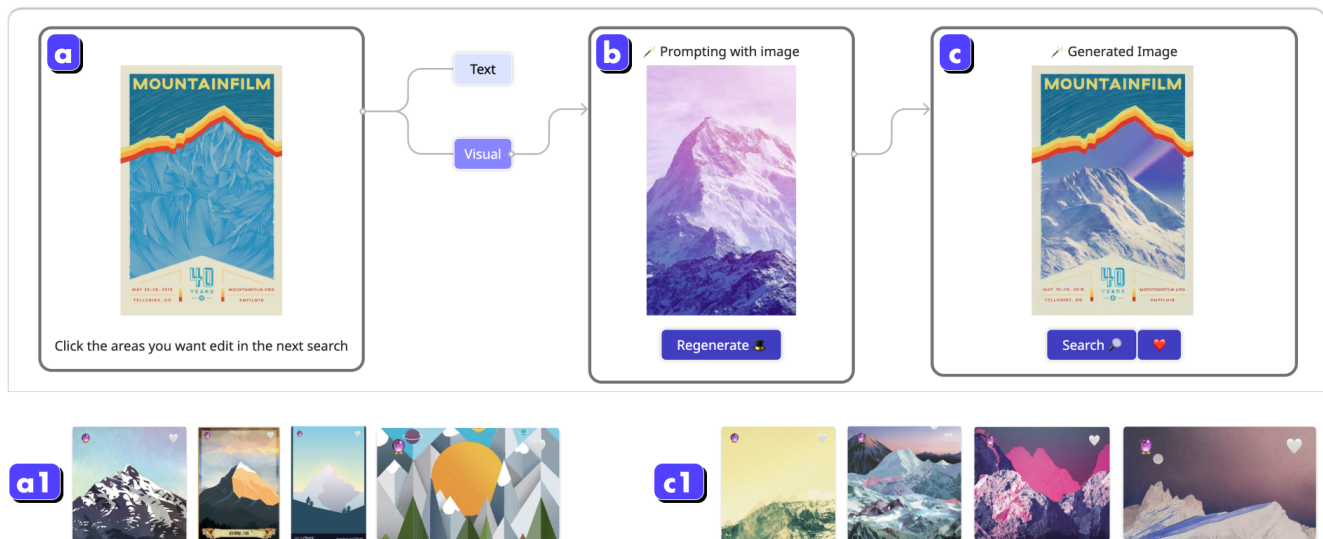

Figure 5: Image-based image modification, which allows the user to express a clear search intent through reference-based editing and search: (a) An image the user wants to modify. In this panel, the user can click or drag the areas in the image that he/she wants to modify. Currently, the blue ice mountain illustration has been selected; (b) An image the user wants to refer to in the editing process. The user can select a reference image from the search results or the saved design list; (c) The generation output. The user can do a regenerate or an image-based search with this generated result. The (a1) and (c1) show the difference in search results searched by the original image (a) and generated image (c).

the mountain landscape for the poster design. She believes that the view a person would see while hiking would also be appropriate for a club poster image. So, she starts to look for poster images that feature forest scenes, and she finds a poster. (Figure 6-a). However, the forest in this poster is only expressed in a single green color, so she decides to modify it.

However, since she is unsure how to modify it, she is considering editing the image with text keywords to see the possible modification directions of this design. So she clicks the text button in the generation panel. Looking at the keyword suggestions, she can find

some keywords related to her previous search history (Figure 6-b1). Also, she discovers that she hasn't looked at many designs related to *"minimalism"* yet (Figure 6-b2). Thus, she selects the tree area (Figure 6-a) and inputs the keyword *"blue and green color forest illustration"* along with the suggested keyword *"minimalist"* (Figure 6-b3). After seeing the result, she adores this design. Then, she starts a new image-based search in this direction using this newly generated image and finds search results related to the new direction (Figure 6-c1).
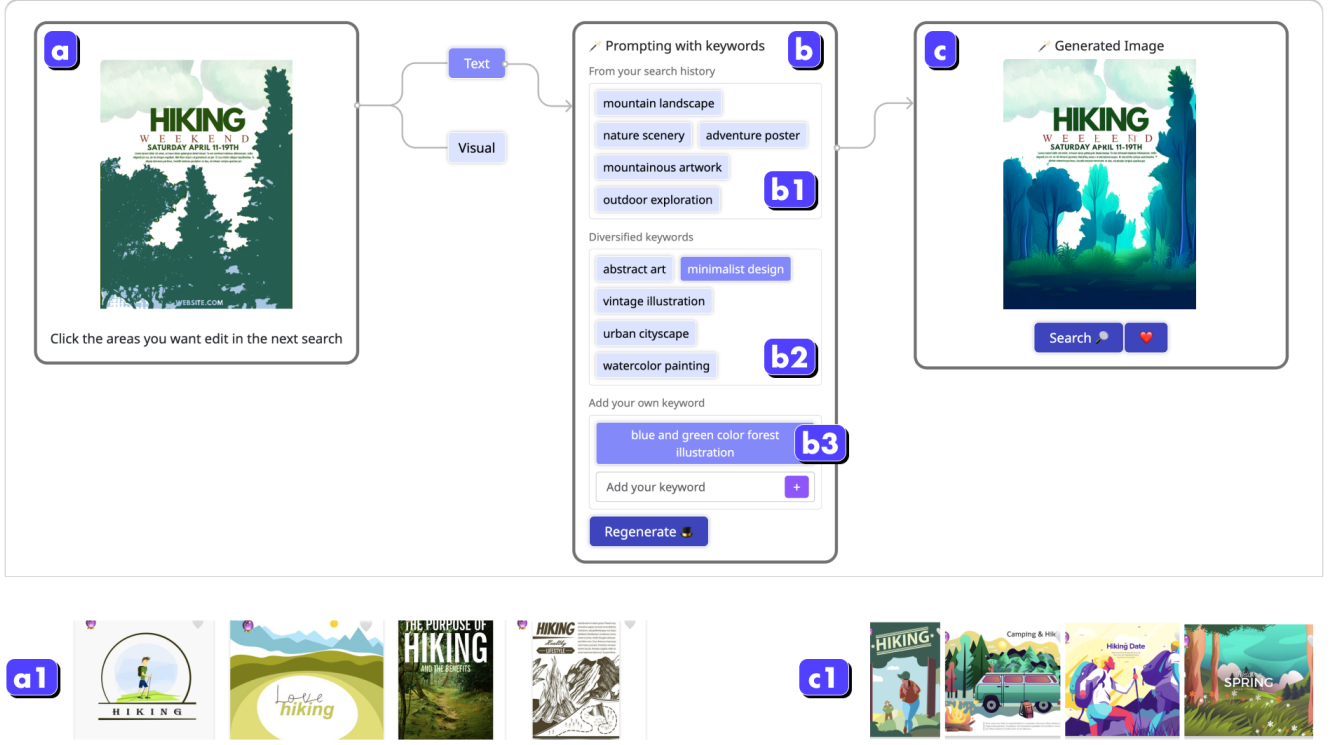
Figure 6: Keyword-based image modification, which allows the user to diversify their search intent through keywords-based editing and search: (a) An image the user wants to modify. Currently, the green forest illustration part has been selected; (b) Keywords suggestion panel based on the search history (e.g., search queries and saved image descriptions). *GenQuery* suggest the keywords similar(b1) and new (b2) to previous search history. Also, the user can make their own keyword in b3; (c) The generation output. The user can do a regenerate or an image-based search with this generated result. The (a1) and (c1) show the difference in search results searched by the original image (a) and generated image (c).

4.1.4 Basic Visual Search Functions in GenQuery. Fundamentally, *Gen-Query* offers functionalities similar to existing visual search tools (e.g., Pinterest), such as searching for visuals with a text prompt (Figure 3-a), the operation to click an image to find similar images (Figure 3-b), and the feature to save desired images and archive them in the right-side panel (Figure 3-c). By clicking the **Show more** button, more results can be viewed from the current search results (Figure 3-e). In the system, the search results are all provided to the user in a format stacked under the most recent search results.

4.2 Technical Details

In this section, we introduce more technical details of *GenQuery*. We specifically focus on the three key functionalities and implementation details.

4.2.1 Query Concretization. *GenQuery* provides five concretized queries based on the text query entered by the user in Figure 3-(a), about one second later. To ensure a smooth user experience with *GenQuery*, it was critical to deliver suggestions at near-real-time speed; hence, we used the gpt-3.5-turbo-16k-0613 LLM model. To deliver concretized queries as output, we applied zero-shot prompting with the [Current Search Query] entered by the user, following the prompting instruction in the Appendix A. For more

accurate results, we prompted the LLM to explain the reasoning behind the concretized queries. Using these concretized queries, *Gen-Query* performs image-based searches and presents the top eight results as shown in Figure 4-(b).

4.2.2 Image-based Image Modification. Upon the user clicking on Figure 3-(d), the panel in Figure 5 is displayed. In this process, the server, leveraging SAM [28], segments the clicked image into individual objects and sends this to the front-end, enabling users to make selections within the image by area. When the user selects an image to refer to and presses the Generation button in Figure 5-(b), the server constructs a mask corresponding to the area selected in Figure 5-(a). Subsequently, utilizing the mask image, the original image, and the reference image, a new image output is generated through the PaintByExample model [60]. We decided to use this model considering the type of modality utilized by the model and its fast generation time (3-4 sec). The resulting creation is displayed in Figure 5-(c).

4.2.3 Keyword-based Image Modification. In the process outlined in §4.2.2, after selecting the area of the image, the user can also press the Text button instead of the Visual to see Figure 6-(b). To aid the user in modifying the image based on keywords, we performed

zero-shot prompting using the same LLM model with the query suggestions. To elaborate, we used the description of the image the user wants to modify ([Description of Current Image]), the last five search queries ([Search Query History]), and descriptions of the last five saved images ([Descriptions of Saved Images]) in our prompting. Through this, we have made the LLM generate keywords that are similar to those that have been searched so far, as well as different keywords. To achieve better suggestion output, we instruct LLM to predict the user's search intent and suggest keywords in the prompt (See Appendix B). Also, we prompt the model to include the reasons for the predicted search intent in its answer. *GenQuery* displays these generated keywords in Figure 6-(b). Once the user has selected the keywords to modify the image and clicked the Generation button, the Kandinsky2.2 [48] diffusion model is used to generate a new image based on the original image, mask image, and selected keywords. As in §4.2.2, we chose to use this model in Figure 6 considering the modality used by the diffusion model and the generation speed (2-3 seconds).

4.2.4 Implementation Details. *GenQuery* is a web-based system, which is composed of a ReactJS front-end and a Python Flask server as the back-end. For the image dataset and basic search functions (e.g., text-based search and image-based search), we used the `clip-retrieval` library [1] on the hosted API provided by the library. Given an input query or image, the library processes the input into an embedding and then queries similar images from the LAION-5B dataset [47] through the API. We used a machine with an AMD Ryzen 9 5900X 12-Core Processor and NVIDIA GeForce RTX 3090 to implement *GenQuery*.

5 EVALUATION

To investigate the impact of the Query concretization (Figure 4), Image-based (Figure 5), and Keyword-based image modification (Figure 6) of *GenQuery*, we conducted a within-subjects study with 16 designers. This study aims to comprehensively investigate how the visual search experience and designers' search patterns change when the generative support is incorporated. The three features proposed for *GenQuery* are built upon traditional visual search features like standard text- and image-based search. For comparison, we establish a baseline that retains these traditional features and excludes generation support, distinguishing it from other approaches like generation-focused interfaces such as Opal [35] or multi-dimensional search-based methods such as MetaMap [25]. We believe that comparing with this standard-like baseline will more effectively showcase the generation's impact on the visual search process. From the study's main objective, the baseline followed the design of *GenQuery*: it supported the same image-based and text-based search features and allowed for saving designs. The baseline supports the same main features of famous visual search tools like Pinterest or Dribbble.

In the study, participants performed two different visual search tasks using both *GenQuery* and the baseline, and saved designs of interest during their explorations. Through this user study, we focused on answering the following research question:

- RQ1: Can *GenQuery* support the user's satisfactory visual search process?

- RQ2: How do the methods of intent expression of *GenQuery* change the user's visual search patterns compared to the visual search process of baseline?
- RQ3: What are the positive and negative effects of the generation process in the visual search process?

5.1 Participants

For the study, we recruited a total of sixteen participants (Age Mean = 27.19, Age SD = 4.09, Female = 10, and Male = 6). We recruited participants who met the following conditions: 1) majoring in design-related majors or currently working in design, and 2) have extensive experience using exemplar-based tools to conduct visual search during the initial stages of the design process. In order to recruit participants from more diverse design domains, we recruited people through online advertisements and word-of-mouth. As a result, we recruited participants in industrial design (six participants), architecture and interior design (four participants), and graphic design (five participants).

5.2 Tasks and Procedure

The study was conducted both in-person and online. For participants who could not participate in person, we conducted the study online using Zoom [7]. The tasks given in the study involved performing a visual search for poster design ideas for 20 minutes. Here, we instructed the participants that the goal of visual search is design ideation, similar to the early stage of the design process. Thus, we explained to the participants that the final goal of this study is to save at least five visuals as design ideas during the process. This task is based on a visual search process of the early design phase for ideation, which was identified from the formative study. Although there are various ways to design ideation, we specifically oriented the participants to focus on conducting the ideation process only through the visual search with the provided search tools. Additionally, we selected the poster design task. The tasks given to participants were 1) a poster design for a hiking club advertisement and 2) a poster design for an architecture exhibition.

The total study time was 1 hour and 30 minutes, which started with an introduction to the study and asking participants for their informed consent. Then, for each task, participants were provided with a 5-minute tutorial of the tool they would be using and proceeded to perform the task for 20 minutes. After each task, participants responded to a survey for 5 minutes. To avoid ordering effects, both the conditions used in the tasks and the tasks themselves were counterbalanced. Participants were provided with a 5-minute break between tasks. After participants completed both tasks, we conducted semi-structured interviews for 20 minutes to investigate the differences in participants' experiences between the two conditions. As compensation, we provided participants with 45,000 KRW (approximately 34 USD).

5.3 Measures

For measures, the surveys conducted after each task included questions that asked participants to rate on a 7-point Likert scale their satisfaction with the saved designs, the quality of these designs [25],

---

[7]https://zoom.us/

their satisfaction with the search process, and Behavior Intention [55]. The survey also included questions that asked participants to rate how well they could express their search intent, along with questions related to the Creativity Support Index (CSI) [6]. Additionally, questions from the NASA-TLX questionnaire [18] were also included in the survey to investigate the potential additional workload that may arise from the generative features. In the case of standardized measures like CSI and NASA-TLX, we conducted a paired t-test. Except for these measures, we conducted a Shapiro-Wilk test to assess the normality of the data. Subsequently, we used a paired t-test for parametric data, while a Wilcoxon signed-rank test was applied for non-parametric data.

In the in-depth interviews, the interview questions focused on what differences participants perceived between their search processes when they had or did not have the generative features. Also, we asked participants on whether there were differences in how they expressed their intents and, if so, what impact it had on their search processes. We also inquired about participants' search patterns by asking them about when they switched between types of search (e.g., image-based vs text-based), when they continued to use the same type of search, and what pattern did they perform more frequently for each condition and why. Furthermore, we also asked participants whether they performed any new patterns when they had the generative features of *GenQuery*.

Additionally, we analyzed participants' interaction logs to measure the number of designs that they saved. To investigate how the overall search processes differed between conditions, we measured the number of text- and image-based search actions that were performed. Furthermore, we also measured how often different search patterns occurred: text-based search to another text-based search (**T-T**), text-based to image-based (**T-I**), image-based to image-based (**I-I**), and image-based to text-based (**I-T**). Lastly, we analyzed patterns related to the generative feature, such as the proportion of cases where the generation results were saved, the proportion of searches performed with the generated results, and the proportion of results saved after a generation-based search.

6 RESULTS

In the study, we observed that participants used the generative features of *GenQuery* to more accurately express their visual search intent and, as a result, obtain more satisfactory search results. Participants expressed to continue using *GenQuery* than baseline significantly. Additionally, we found that, due to the generative features, participants could find more diverse and creative design ideas while performing fewer text- and image-based searches. Also, we found a new visual search pattern through *GenQuery*, *Search by Generation*, which means an image-based search through the image- or keyword-based modified image output.

6.1 RQ1: High Satisfaction with Visual Search Process Using the Generative Features

Participants rated their satisfaction with the search results to be significantly higher when using *GenQuery* (Figure 7; Q1, Mean_diff=0.94, p<0.05). In particular, participants felt significantly higher satisfaction in terms of the diversity (Figure 7; Q3, Mean_diff=1.50, p<0.01)

and creativity (Figure 7; Q3, Mean_diff=1.69, p<0.001) of the discovered ideas. Additionally, participants willingness to use each tool in the future was significantly higher when they used *GenQuery* (Figure 7; Q14, Mean_diff=1.75, p<0.01). P4 mentioned that the process of exploring designs became more enjoyable due to the integration of generative steps. Most participants stated in interviews that they preferred the designs found through *GenQuery* (13/16). P5 said that this was because "*the generative feature seemed to consistently suggest new directions during the visual search process where one can continue exploring unconsciously*". Particularly, when using the baseline, participants (P4, P8, and P12) described their process to mostly involve repeatedly choosing an image from the search results that are closest to what they want, not exploring various design ideas. On the other hand, *GenQuery* generation allowed them to search for ideas diversely; participants showed a strong desire to use *GenQuery* in their design process (Figure 7; Q12, Mean_diff=1.81, p<0.01).

Another reason for participants' higher satisfaction with the search results in *GenQuery* was that the ideas found through the system included more advanced design ideas than their initial ideas. Participants attributed this to the different ways in which the generative features supported the search process. First, participants (P1, P7, P11, and P13) described how the query concretization Figure 4 allowed them to dive into more quickly concrete design ideas compared to when they started with abstract search queries. Second, participants (P2, P4-P6, P8, P12, and P14) mentioned the image- and keyword-based image modification (Figure 5 and Figure 6) helped them to think about the next design search direction by actually observing search results similar to the modification output. In particular, P4 said, "*In the initial design process, it is essential to save the ideas and merge those again to review whether the idea is good or not. However, in this tool, you can simply and quickly try the design you want, so the design process supported by GenQuery seemed to cover both the initial and middle stages of design.*" This was a view echoed by many participants (P2, P4-P5, P8, P12). In summary, *GenQuery* assisted designers in more quickly and easily navigating the developed design ideas during the visual search process, leading to higher satisfaction levels.

On the other hand, there was no significant difference in the quantity of designs found between *GenQuery* and the baseline (In Figure 7 Q2, Mean_diff=0.56, p>0.05; Baseline_mean = 12.44, Baseline_SD = 5.67; The average number of saved designs *GenQuery*_mean = 9.75, *GenQuery*_SD = 4.23, p>0.05). Regarding this, P6 said that there were cases where she had to repeatedly modify the prompt or image used in the generative features because the generated results did not satisfy her. Furthermore, P6 mentioned that the generation process led her to repeatedly modify an image unknowingly until the system created a result that she wanted. In the visual search process using *GenQuery*, the images generated by the modification process account for 45.8% (SD=28.6%) of the total saved ideas. P10 said "*In the conventional search process, even if undesired results appeared, I could easily ignore them because there is no opportunity to modify the image. But the generation process of GenQuery made me immersed in the generation process itself.*" As there were quite a few instances where the quality of the generated results was low or the generation went in an unpredictable direction, participants showed the opinion that there was not much
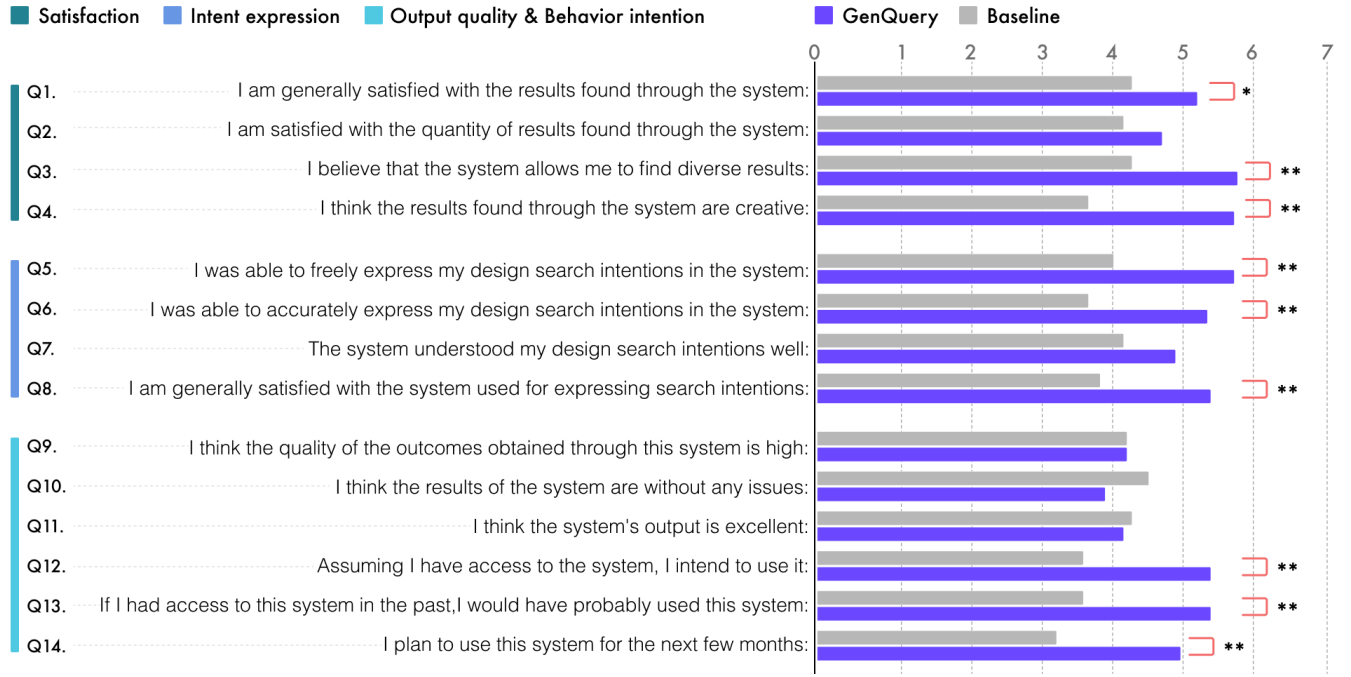
Figure 7: Survey results in the user study. The survey asked participants to rate their Satisfaction [25], Intent Expression, and Output Quality & Behavior intention [55] on a 7-point Likert scale. ∗: p<0.05 and ∗∗: p<0.01.

difference in the quality of the entire results obtained from the system (Figure 7; Q9, Mean_diff=0.00, p>0.05). Specifically, P5 stated that there were cases when he was not very satisfied just by looking at the results generated by the system, and about 30-40% of the generated results were not satisfactory. All participants who expressed dissatisfaction during the creation process had specific images they wanted to generate, but the generative model failed to meet their needs.

In summary of the findings for RQ 1, *GenQuery* provided great satisfaction overall, especially regarding the diversity and creativity of the final identified design products. However, it did not provide great satisfaction in terms of the quality of the generated product itself and the overall quantity of the identified products.

6.2 RQ2: Findings of New Visual Search Patterns with A More Efficiency

Participants were able to express their visual search intent more accurately through *GenQuery*, and as a result, they conducted a more efficient visual search process. As shown in Figure 7, participants could express their visual search intent more freely (Figure 7; Q5, Mean_diff=1.69, p<0.01) and accurately (Figure 7; Q6, Mean_diff=1.69, p<0.01) in *GenQuery*. Overall, they were highly satisfied with this (Figure 7; Q8, Mean_diff=1.56, p<0.01). However, as mentioned in Section 6.1 about the low controllability of the generation process, there was no significant difference in the survey asking whether the system understood their search intent well (Figure 7; Q7, Mean_diff=0.75, p>0.05). P3 explained that he could elaborate their search intent through *GenQuery* compared to the baseline, and though it was not perfect, it was somewhat possible

to express it. Particularly, P3 stated that the overall visual search process seemed to be more efficient due to the generation process.

In fact, analyzing the logs of the participants in both systems (Figure 8), when using the baseline, participants significantly performed more Text-based search (**T**) (*GenQuery*_mean = 3.69, Baseline_mean = 12.81, p<0.001) and **Show more** (*GenQuery*_mean = 6.50, Baseline_mean = 24.44, p<0.01) actions. The **Show more** button, which is pressed to see more related search results from the search results, is equivalent to scrolling down on Pinterest. In Image-based search (**I**), although there's not a significant difference (*GenQuery*_mean = 18.38, Baseline_mean = 26.31, p>0.05), participants who used *GenQuery* definitely took fewer actions (Figure 8). Despite the difference in the number of actions, there was no significant difference in the number of final saved design ideas (Baseline_mean = 12.44, Baseline_SD = 5.67, *GenQuery*_mean = 9.75, *GenQuery*_SD = 4.23, p>0.05). According to the participants (P1-P3, P5, P8, P12, and P14), the generative feature that recommends a text query prevented unnecessary query editing processes, and the process of generating and searching for images reduced many actions performed to find the desired image within the image.

In order to analyze how the search pattern in the visual search process has changed specifically new ways of search intent expression, we also performed a pattern analysis of Text-based Search (**T**) and Image-based Search (**I**) as explained in §5.3. In the interviews session of the study, we asked about the meaning of **T-T**, **T-I**, **I-I**, and **I-T** in the visual search process, and as a result, we were able to summarize the meanings of each pattern as follows. The patterns are similar to what we found in the §3.3.
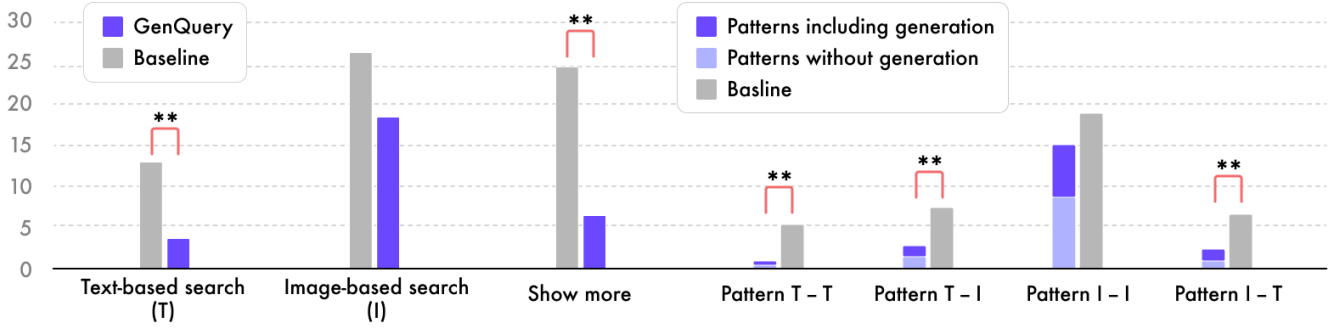
Figure 8: Visual search pattern analysis results. The left side graph shows the difference between *GenQuery* and baseline in terms of text-based search (T), image-based search (I), and Show more button-clicking action. The right side graph shows the difference between *GenQuery* and baseline in terms of search pattern of T-T, T-I, I-I, and I-T.

- **T-T**: Pattern performed when unable to decide on the desired design area to start a visual search
- **T-I**: Pattern performed when at least some of the desired visuals are included in the search results
- **I-I**: Pattern performed when the desired image is currently visible or when wanting a result that is somewhat slightly different than the results found so far
- **I-T**: Pattern performed when wanting to refresh the search process because it seems only to find similar images

As seen in Figure 8, apart from **I-I**, participants were able to drastically reduce the count of the other three types of patterns when using *GenQuery* (T-T_diff = -4.94; T-I_diff = -6.00; I-T_diff = -5.63; All p<0.01). Interview results revealed that many participants (P3, P6-P7, and P10-P16) felt they particularly performed **T-T** and **I-T** patterns often in the baseline. For this reason, P12 mentioned, "*I feel like I've performed these two patterns quite often as I frequently face difficulties to find the appropriate search term or when the visuals I'm currently looking for feel not so novel.*" In contrast, when using *GenQuery*, they reported to have often performed the **I-I** pattern (P3, P5, P9-P12, and P14-P15). P14 elaborated, "*Finding the desired image was relatively easy, and once I found this image, I wanted to look for a slightly different image. The process of modifying the image for a generation was very suitable for this type of pattern.*" Notably, P3 said that "*It was possible to generate by referencing other images in the two features of modifying the image, and also, the keywords showed what kind of things I have mainly searched for so far (Figure 6), that helped me understand in what direction I should explore at that point.*" In other words, *GenQuery* had a pronounced impact on the overall visual search pattern, and the **I-I** pattern, which became the core of the visual search, had a significant role.

Furthermore, we observed that due to the generative feature, participants could generate a design expressing the direction they wanted to explore, search with it, and consequently save the images they wanted. We named this search pattern as *Search by Generation*. As can be seen in Figure 8, the four patterns on the right are represented in *GenQuery* divided into cases with and without image generation actions included between them. When using *GenQuery*, image generation actions are included between the actions in **T-T**, **T-I**, **I-I**, and **I-T** at rates of 66.7% (SD=42.4%), 46.7% (SD=37.5%),

47.7% (SD=22.5%), and 62.9% (SD=44.6%) respectively. Specifically, the pattern of *Search by Generation* is 34.4% (SD=20.1%, Min=3.8%, Max=87.5%) among the image-based search. P6, who performed the lowest amount of *Search by Generation*, mentioned, "As I got more immersed in the process of generating the final result, I couldn't think much about using it for search." In contrast, P15, who performed the most generation searches, stated, "The feature to modify and search really allowed me to express my ideas more accurately, so I found myself conducting a lot of those searches. It was a feature that I had always felt was necessary while using platforms like Pinterest." Also, the Figure 10 illustrates the examples of *Search by Generation* we observed in the study. While there may be variations among participants, we did observe this search pattern frequently results at various points throughout the visual search process Figure 9. When using *GenQuery*, the ratio of searching with a newly generated image and saving a design within the results was also calculated. This value was 35.8% (SD=21.3%) of the total saved images. In other words, Participants saved around 36% images through the *Search by Generation* among the total amount of saved designs. Search through generation

Based on the findings, we have demonstrated how the visual search process has changed with the generative features of *GenQuery*. The generative features made the search process more efficient by helping to express the precise visual search intent, making it possible to find a similar quantity of results more satisfactorily with fewer search actions.

6.3 RQ3: Strength and Weakness of Unexpectedness of Generative Model and Creativity

The unexpectedness provided by the generative model, especially the T2I model generating images, offered both positive and negative aspects to participants during the visual search process.

Firstly, on the negative side, participants thought the quality of the results generated by the model was relatively poorer than the actual designs. However, what made the search experience worse was the low controllability to generate desired images, as confirmed in the interview response from P10. P10 said using *GenQuery* was

Figure 9: All participants' action log visualization in *GenQuery* including *Search by Generation*, Image-based search, Tex-based search, Image modification (Figure 5 and Figure 6), Query concretization (Figure 4), Saved, and Unsaved actions: Generation search means an image-based search by the generated output. The two red lines indicate the unsatisfactory generation cases because the participants were deeply engrossed in the generation process without searching. The actual images generated from these processes are illustrated in Figure 11 as well.
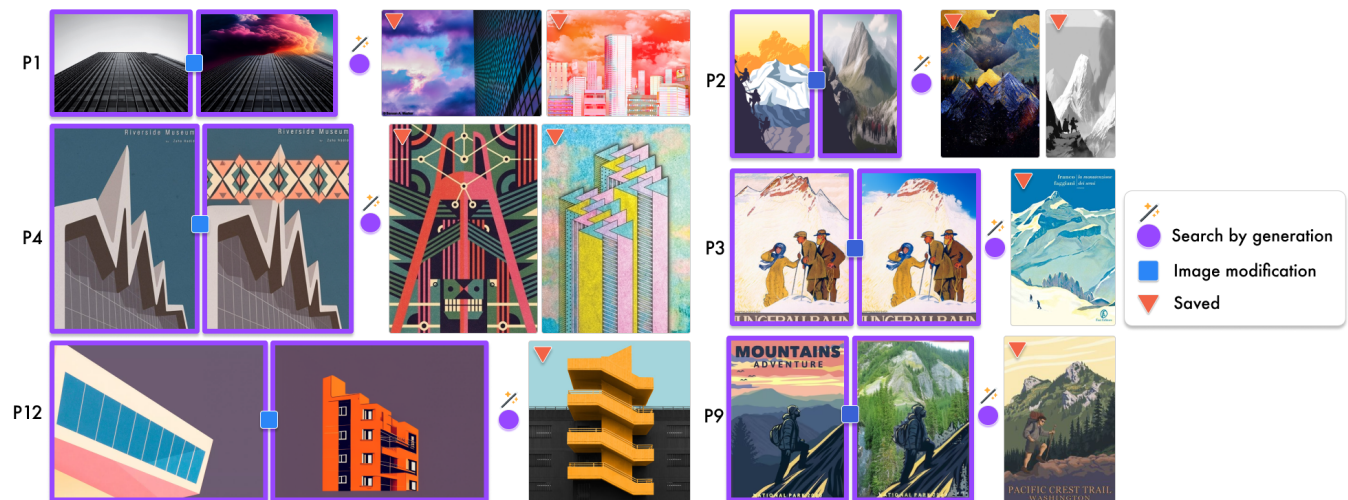

Figure 10: The actual examples of saving design ideas through *Search by Generation* in *GenQuery*. The left side is the case of task 1, and the right side is the example of task 2.

		GenQuery		**Baseline**		Statistics	
		mean	std	mean	std	p-value	Sig.
NASA-TLX [18]	Mental	2.56	1.36	2.69	1.82	0.86	-
	Physical	1.25	0.77	2.06	2.02	0.06	+
	Temporal	1.63	0.81	2.38	2.06	0.19	-
	Effort	3.69	1.70	3.31	1.58	0.53	-
	Performance	5.00	1.15	4.88	0.89	0.74	-
	Frustration	2.00	1.59	2.69	2.06	0.17	-
Creativity Support Index (CSI) [6]	Enjoyment	8.19	1.88	4.19	0.93	0.00	***
	Exploration	8.28	1.13	3.53	1.22	0.00	***
	Expressivness	8.25	1.46	3.75	1.34	0.00	***
	Immersion	7.03	2.93	4.63	1.34	0.01	**
	Results Worth Effort	7.41	1.75	4.88	1.07	0.00	***
	Collaboration	—	—	—	—	—	—

Table 1: The results of NATA-TLX ([18]; 7-Likert Scale) and Creativity Support Index ([6]; 10-Likert Scale) survey: Since *GenQuery* dosen't support the collaboration with other designers, we excluded the Collaboration related questions (we assigned the weight value as zero) questions in the final calculation of CSI score. -: p > .100, +: .050 < p < .100, *: p < .050, **: p < .010, ***: p < .001

very challenging, citing the difficulty of controlling the text-to-image model as the reason. He explained that, especially when he had images he wanted to create in mind, he found it more difficult to accept the generated result and seemed to get deeply immersed in the generation process due to repetitive prompting processes (e.g., changing the reference image or keywords). Due to this low controllability of the model, three participants were unsatisfied with *GenQuery*.

However, as seen in Table 1, this generation process did not cause an additional burden. On the other hand, many participants (P1-P2, P4-P5, P7, P11-P13, and P16) reported being led to contemplate unexpected exploration paths due to the model's unforeseen image generations. In interviews, several participants (P1, P4, P12, and P16) stated they were able to plan their next search process by witnessing the generation in an unexpected direction. Related to this, P16 commented, "*Even though the generation output was strange at times, I became more curious about the search outcome through it. And when I tried to generate, I got a result that I couldn't imagine in my head, so I actually searched with it.*" P16 mentioned that this process was interesting as it felt like evaluating the results generated by the AI.

Furthermore, the unexpectedness resulting from low controllability served as a starting point for users to engage in various attempts. P2 remarked, "*I was initially puzzled seeing the generation results, but shortly after, a new idea popped up looking at it,*" and further added, "*Through this generation process, it was possible to consider up to ten ideas in the search process that usually involves contemplating only three.*" Even, P3 described experiencing a sense that the part they had to think about diminished when using *Gen-Query*. *GenQuery* consistently proposed new directions within the range set by the user, and as a result, participants using *GenQuery* responded in the survey that they had overwhelmingly creative experiences compared to the baseline (Table 1, CSI).

We discovered the strengths that can be exhibited in terms of design creativity when leveraging the incompleteness of T2I models in the visual search process. Although some participants felt discomfort regarding this aspect, ultimately, our study results show

that such incompleteness could be sufficiently utilized as an interaction in the visual search process, and more broadly, in the visual search process that requires diverse stimulation. This turned the perceived flaw of the T2I model into a feature beneficial to the creative process.

7 DISCUSSION

7.1 Generation Process in Visual Search: The Controllability on Output Depends on the Specification Level of Intent

In our study, the user's level of concreteness of the search intent influenced the experience of participants with the generation results of *GenQuery*. We observed that participants became more engrossed in the generation process when the image they wanted to see was clearly defined, as shown in Figure 11. Particularly, when the generated results contained low-quality details, participants became fixated on the generation results (Figure 11). P6 stated, "*I wanted to draw many people on the hill, and I had a specific image in my mind. However, the generated results were different from what I expected, so I kept pressing the generate button.*" P16 said, "*The legs of the person were not appearing, so I kept generating. I wanted to draw at least the legs.*" As participants tried to directly generate the desired image without *search by generation*, with unexpected outcomes resulting in a poorer user experience, and ultimately, we observed required higher controllability in this case. P6 and P16 mentioned that previewing what kind of design could be discovered through these generated results would prevent getting caught up in the generation process.

On the other hand, from participants' comments (P8 and P14), we observed that the lower controllability on generative models works well when the user's search intent was not concrete or when the fine details of output were not critical (e.g., generating backgrounds or incorporating artistic elements). With this degree of intent, P8 was willing to use the generated results for visual search, even if they differed from their expectations. In the case of generation related to abstract styles rather than specific objects, participants engaged
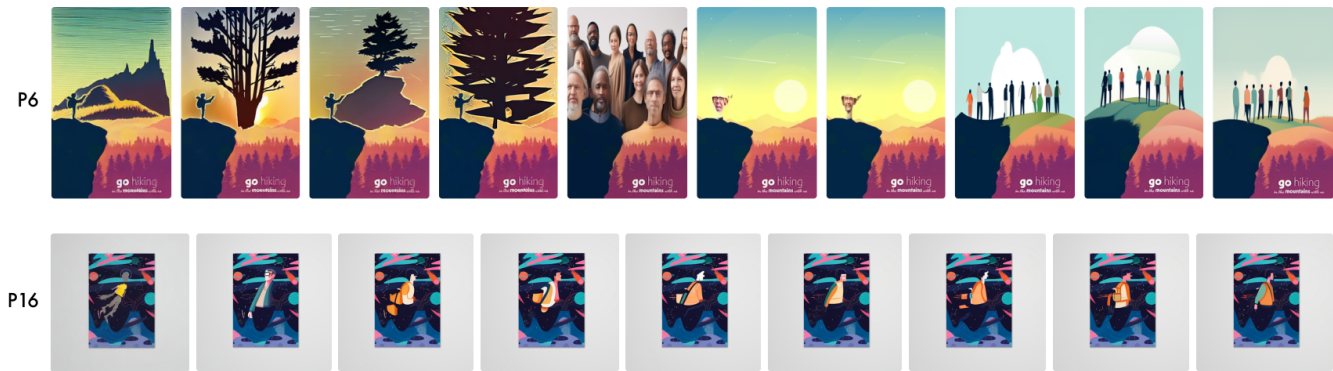
Figure 11: Examples of engrossing into the generation process of P6 and P16: P6 tried to generate the background of the image with trees and people standing on the hill. P16 tried to generate a hiking person illustration in the center of the image iteratively. Both felt unsatisfactory and explained that they had wanted to make the image that they imagined in the interview session.

more in the search process without getting deeply involved in the image modification process. Interestingly, in these cases, many participants often reconsidered their search direction based on the generated results.

In summary, we observed the generative models' output provides different user experiences based on the status of the search intent of users. A design lesson for future work is the level of controllability over the generation output would be differentiated according to the concretization level of the user's search intent.

7.2 Generation Process as Design Prototyping

GenQuery provided a design process that combined idea search and generation, and as a result, we observed a new design idea was generated through design prototyping-like actions within the search tool. For example, when P14 became familiar with the tool and thought he needed totally different ideas compared to his search history, he proceeded to search for basic design elements (e.g., basic layout) as shown in Figure 12. Then, he tried to edit a more significant part of the image and generate various new ideas. P14 was able to perform design exploration in new directions using the generated results. P14 stated, "*Even though the generated results were incomplete, the integrated generation results within the search process felt like a collaborator constantly throwing new ideas at me.*" Besides P14, some participants (P7-P8 and P14) also did additional searches to find design elements (e.g., background image) like a design prototyping process, beyond using the images of existing results. Most of the design tools for search and prototyping are clearly divided so far, as we can check from our formative study. However, by integrating the generation process and search, we found the prototyping and search processes could be interplay by complementing each other. Designers can perform prototyping actions when they need to explore new ideas after a certain level of exploration. Designers can also engage in design exploration actions to further develop raw ideas generated during the prototyping process within *GenQuery*. We envision future work on generative model-based design support tools that combine search and generation processes, allowing the designers to express their design mode freely.

7.3 Beyond Keywords Suggestion from Search History

Through the keyword suggestion in Figure 6, participants were able to define their next direction of exploration while considering the design directions they had explored so far (P1-P2, P4, P7-P8, P11, and P13-P14). P1 and P4 also stated how the user has modified images in *GenQuery* would be useful information to track the user's intent in addition to the search history information (e.g., search query and saved design description). Furthermore, to this statement, which generation outputs were saved and which ones were used for the search can also serve as valuable clues for *GenQuery* to infer the user's visual search intent. By leveraging this information, *GenQuery* could be further improved to provide search direction feedback on how to modify specific areas of an image and show the expected search results based on the modification.

7.4 Expanding the Scope: Accommodating Varied Design Intentions

Our work proposes a method that expresses visual search intent through the outputs of a generation model. Going beyond visual search intent, in future work, our work could be expanded to accommodate various forms of intent that may arise during the design process. For instance, intentions related to modifications during the design prototyping stage or intentions related to creating something new design ideas could be expressed through various modalities of the generation model (e.g., design concept keywords, design style, or design feedback). In addition, by providing the generation model with various types of information from the design process, such as sketches and command sequence patterns used in design tools, we can interpret and reveal the tacit intent of designers. By revealing the vague design intent, future design tools will support a designer by suggesting ideas related to the intent or generating several alternatives. The generative model, offering incomplete but varied modalities, will pose an effective function in future design tools, especially for designers for whom clearly expressing intent poses a challenge.
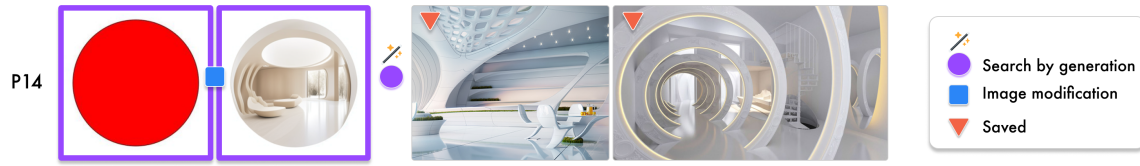
Figure 12: P14's *Search by generation* pattern within *GenQuery*: P14 searched the basic design layout (the first image on the left) and selected the red circle from the most left-side image. *GenQuery* provided the "futuristic" keyword. Then, he conducted keywords-based editing by inputting "futuristic architecture design" and generated the second image. By doing a *generation by search* with this image, P14 saved several images that were searched from the generation output.

7.5 Difference Between the Full Generation-based Exploration and Generation Search-based Exploration

Several participants (P4, P9, P13, P15-P16) explained that the *GenQuery* made them look forward to the search results though they obtained incomplete generation outputs. Specifically, P16 said, "My expectation of the generated output was lower than compared to my experience with general AI models. Rather, I enjoyed and was curious about the search results based on this incomplete generated output." Through participants' responses, the combination of specific task functions would affect the user's expectation on the generated output. Many popular text-to-image models (e.g., DALL·E [8], Firefly [9], or Midjourney [10]) support a fully generation-based exploration process. For example, Adobe's Firefly provides four outputs for one text prompt, and then the user can select and freely edit the outputs' style, concept, color tone, or composition. Also, the user can generate other similar images based on the most preferred image. However, since the idea exploration process in the early design phase requires a quick process to see various results, full generation-based exploration might be challenging to support this process. Moreover, users may want an output that perfectly fits their intent in this approach. Therefore, based on these findings, future generation-based tools can be designed by considering the level of user expectation on the generation output with the characteristics of the task.

7.6 Limitations and Possible Approaches for Generation Search

Even if participants were able to express their desired results through generation visually, there were situations where the related search results were scant. P1 commented, "*I was reasonably satisfied with the generated output, so I tried searching with them, but I obtained very few search results. It seemed like there was a lack of related content in the dataset, so I explored different design directions.*" Additionally, P11 mentioned, "*Since I couldn't see any related search results, I modified the generation results to align them more with what I wanted and then saved those results.*" When the search results were limited based on the generation output, several potential future research directions can address this issue, including 1) providing guidance for the next generation direction based on the position of the generation results within the dataset, 2) offering additional

controllability in the generation process to support the creation of desired results in detail, and 3) modifying the search algorithm. These can all be explored to address this situation effectively.

Besides, *GenQuery* has the following limitations: First, the modality for selecting the area to modify in an image should allow for more accurate selection. Specifically, the segment-based approach is useful when the object in the image is segmented clearly. However, if the segmented part was too small, the segmentation wasn't clearly divided, or there were some areas including many segments they wanted to edit, the participant wanted to select the area freely. Second, the feature of storing and reusing the modified images, modification process, and their associated search results is required. For example, when a user continuously tries to select the image's background, the system automatically provides selection suggestions based on the editing history. These two limitations were identified based on feedback from the participants in the study. The limitation should be addressed in future work to improve the system further.

8 CONCLUSION

This paper proposes *GenQuery*, a novel system that integrates generative models into the visual search process. By automatically elaborating on users' queries, *GenQuery* can surface concrete search directions when users only have abstract ideas. Also, by generatively modifying existing search results and using these to search for similar images, *GenQuery* allows users to express what they are looking for more precisely. Our study results revealed the participants felt that they could more accurately express their visual search intent. Through *GenQuery*, participants felt that they could find more diverse images with more satisfaction. Ultimately, the generation process enhanced the user's creativity in the visual search process by guiding them to new search directions beyond searching for desired images. Although the generation output is unreliable and cannot be fully controlled, *GenQuery* demonstrated the benefit of leveraging generated outputs as intermediate materials that can represent designers' intents.

---

[8]https://openai.com/dall-e-3
[9]https://firefly.adobe.com/
[10]https://www.midjourney.com

REFERENCES

[1] Romain Beaumont. 2022. Clip Retrieval: Easily compute clip embeddings and build a clip retrieval system with them. https://github.com/rom1504/clip-retrieval.

[2] Nathalie Bonnardel. 1999. Creativity in design activities: the role of analogies in a constrained cognitive environment. In *Creativity & Cognition*. ACM New York, NY, USA, New York, NY, USA, 158–165. https://doi.org/10.1145/317561.317589

[3] Stephen Brade, Bryan Wang, Mauricio Sousa, Sageev Oore, and Tovi Grossman. 2023. Promptify: Text-to-Image Generation through Interactive Prompt Exploration with Large Language Models. , 14 pages. https://doi.org/10.1145/3586183.3606725

[4] Sara Bunian, Kai Li, Chaima Jemmali, Casper Harteveld, Yun Fu, and Magy Seif Seif El-Nasr. 2021. VINS: Visual Search for Mobile User Interface Design. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 423, 14 pages. https://doi.org/10.1145/3411764.3445762

[5] Hernan Casakin and Gabriela Goldschmidt. 1999. Expertise and the use of visual analogy: implications for design education. *Design Studies* 20, 2 (1999), 153–175. https://doi.org/10.1016/S0142-694X(98)00032-5

[6] Erin Cherry and Celine Latulipe. 2014. Quantifying the Creativity Support of Digital Tools through the Creativity Support Index. *ACM Trans. Comput.-Hum. Interact.* 21, 4, Article 21 (jun 2014), 25 pages. https://doi.org/10.1145/2617588

[7] John Joon Young Chung and Eytan Adar. 2023. Artinter: AI-Powered Boundary Objects for Commissioning Visual Arts. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) *(DIS '23)*. Association for Computing Machinery, New York, NY, USA, 1997–2018. https://doi.org/10.1145/3563657.3595961

[8] John Joon Young Chung and Eytan Adar. 2023. PromptPaint: Steering Text-to-Image Generation Through Paint Medium-like Interactions. , Article 6 (2023), 17 pages. https://doi.org/10.1145/3586183.3606777

[9] John Joon Young Chung, Wooseok Kim, Kang Min Yoo, Hwaran Lee, Eytan Adar, and Minsuk Chang. 2022. TaleBrush: Sketching Stories with Generative Pretrained Language Models. 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 209, 19 pages. https://doi.org/10.1145/3491102.3501819

[10] Hai Dang, Frederik Brudy, George Fitzmaurice, and Fraser Anderson. 2023. WorldSmith: Iterative and Expressive Prompting for World Building with a Generative AI. arXiv:2308.13355 [cs.HC] https://arxiv.org/abs/2308.13355

[11] Claudia Eckert and Martin Stacey. 2000. Sources of inspiration: a language of design. *Design Studies* 21, 5 (2000), 523–538. https://doi.org/10.1016/S0142-694X(00)00022-3

[12] James Fogarty, Desney Tan, Ashish Kapoor, and Simon Winder. 2008. CueFlik: interactive concept learning in image search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. Association for Computing Machinery, New York, NY, USA, 29–38. https://doi.org/10.1145/1357054.1357061

[13] Jonas Frich, Midas Nouwens, Kim Halskov, and Peter Dalsgaard. 2021. How Digital Tools Impact Convergent and Divergent Thinking in Design Ideation. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 431, 11 pages. https://doi.org/10.1145/3411764.3445062

[14] Songwei Ge, Taesung Park, Jun-Yan Zhu, and Jia-Bin Huang. 2023. Expressive Text-to-Image Generation with Rich Text. , 7511-7522 pages. https://doi.org/10.1109/ICCV51070.2023.00694

[15] Vidit Goel, Elia Peruzzo, Yifan Jiang, Dejia Xu, Nicu Sebe, Trevor Darrell, Zhangyang Wang, and Humphrey Shi. 2023. PAIR-Diffusion: Object-Level Image Editing with Structure-and-Appearance Paired Diffusion Models. arXiv:2303.17546 [cs.CV] https://arxiv.org/abs/2303.17546

[16] Gabriela Goldschmidt. 2016. Linkographic evidence for concurrent divergent and convergent thinking in creative design. *Creativity research journal* 28, 2 (2016), 115–122. https://doi.org/10.1080/10400419.2016.1162497

[17] Gabriela Goldschmidt and Maria Smolkov. 2006. Variances in the impact of visual stimuli on design problem solving performance. *Design Studies* 27, 5 (2006), 549–569. https://doi.org/10.1016/j.destud.2006.01.002

[18] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human mental workload* 1 (1988), 139–183. https://doi.org/10.1016/s0166-4115(08)62386-9

[19] Scarlett R. Herring, Chia-Chen Chang, Jesse Krantzler, and Brian P. Bailey. 2009. Getting inspired! understanding how and why examples are used in creative design practice. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 87–96. https://doi.org/10.1145/1518701.1518717

[20] 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:2208.01626 [cs.CV] https://arxiv.org/abs/2208.01626

[21] Forrest Huang, John F. Canny, and Jeffrey Nichols. 2019. Swire: Sketch-based User Interface Retrieval. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3290605.3300334

[22] David G. Jansson and Steven M. Smith. 1991. Design fixation. *Design Studies* 12, 1 (1991), 3–11. https://doi.org/10.1016/0142-694X(91)90003-F

[23] Ellen Jiang, Kristen Olson, Edwin Toh, Alejandra Molina, Aaron Donsbach, Michael Terry, and Carrie J Cai. 2022. PromptMaker: Prompt-based Prototyping with Large Language Models. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 35, 8 pages. https://doi.org/10.1145/3491101.3503564

[24] Hyeonsu B. Kang, Gabriel Amoako, Neil Sengupta, and Steven P. Dow. 2018. Paragon: An Online Gallery for Enhancing Design Feedback with Visual Examples. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174180

[25] Youwen Kang, Zhida Sun, Sitong Wang, Zeyu Huang, Ziming Wu, and Xiaojuan Ma. 2021. MetaMap: Supporting Visual Metaphor Ideation through Multidimensional Example-based Exploration. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 427, 15 pages. https://doi.org/10.1145/3411764.3445325

[26] 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

[27] Tae Soo Kim, Yoonjoo Lee, Minsuk Chang, and Juho Kim. 2023. Cells, Generators, and Lenses: Design Framework for Object-Oriented Interaction with Large Language Models. , Article 4 (2023), 18 pages. https://doi.org/10.1145/3586183.3606833

[28] 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 B. Girshick. 2023. Segment Anything. *ArXiv* abs/2304.02643 (2023). https://arxiv.org/abs/2304.02643

[29] Hyung-Kwon Ko, Gwanmo Park, Hyeon Jeon, Jaemin Jo, Juho Kim, and Jinwook Seo. 2023. Large-Scale Text-to-Image Generation Models for Visual Artists' Creative Works. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (Sydney, NSW, Australia) *(IUI '23)*. Association for Computing Machinery, New York, NY, USA, 919–933. https://doi.org/10.1145/3581641.3584078

[30] Janin Koch, Nicolas Taffin, Andrés Lucero, and Wendy E. Mackay. 2020. SemanticCollage: Enriching Digital Mood Board Design with Semantic Labels. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands) *(DIS '20)*. Association for Computing Machinery, New York, NY, USA, 407–418. https://doi.org/10.1145/3357236.3395494

[31] Balazs Kovacs, Peter O'Donovan, Kavita Bala, and Aaron Hertzmann. 2019. Context-Aware Asset Search for Graphic Design. *IEEE Transactions on Visualization & Computer Graphics* 25, 07 (jul 2019), 2419–2429. https://doi.org/10.1109/TVCG.2018.2842734

[32] Adriana Kovashka, Devi Parikh, and Kristen Grauman. 2012. WhittleSearch: Image search with relative attribute feedback. In *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Los Alamitos, CA, USA, 2973–2980. https://doi.org/10.1109/CVPR.2012.6248026

[33] Yoonjoo Lee, Tae Soo Kim, Sungdong Kim, Yohan Yun, and Juho Kim. 2023. DAPIE: Interactive Step-by-Step Explanatory Dialogues to Answer Children's Why and How Questions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 450, 22 pages. https://doi.org/10.1145/3544548.3581369

[34] 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* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 384, 23 pages. https://doi.org/10.1145/3491102.3501825

[35] Vivian Liu, Han Qiao, and Lydia Chilton. 2022. Opal: Multimodal Image Generation for News Illustration. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 73, 17 pages. https://doi.org/10.1145/3526113.3545621

[36] Vivian Liu, Jo Vermeulen, George Fitzmaurice, and Justin Matejka. 2023. 3DALL-E: Integrating Text-to-Image AI in 3D Design Workflows. , 23 pages. https://doi.org/10.1145/3563657.3596098

[37] Andrés Lucero. 2012. Framing, aligning, paradoxing, abstracting, and directing: how design mood boards work. In *Proceedings of the Designing Interactive Systems Conference* (Newcastle Upon Tyne, United Kingdom) *(DIS '12)*. Association for Computing Machinery, New York, NY, USA, 438–447. https://doi.org/10.1145/2317956.2318021

[38] Piotr Mirowski, Kory W. Mathewson, Jaylen Pittman, and Richard Evans. 2023. Co-Writing Screenplays and Theatre Scripts with Language Models: Evaluation by Industry Professionals. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 355, 34 pages. https://doi.org/10.1145/3544548.3581225

[39] Ron Mokady, Amir Hertz, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. 2023. Null-text Inversion for Editing Real Images using Guided Diffusion Models. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Los Alamitos, CA, USA, 6038–6047. https://doi.org/10.1109/CVPR52729.2023.00585

[40] Céline Mougenot, Carole Bouchard, Améziane Aoussat, and Steeve Westerman. 2008. Inspiration, images and design : an investigation of designers' information gathering strategies. *Journal of Design Research* 7, 4 (2008), 331–351. https://doi.org/10.1504/JDR.2008.026987

[41] Mohammad Amin Mozaffari, Xinyuan Zhang, Jinghui Cheng, and Jin L.C. Guo. 2022. GANSpiration: Balancing Targeted and Serendipitous Inspiration in User Interface Design with Style-Based Generative Adversarial Network. 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 537, 15 pages. https://doi.org/10.1145/3491102.3517511

[42] Savvas Petridis, Nicholas Diakopoulos, Kevin Crowston, Mark Hansen, Keren Henderson, Stan Jastrzebski, Jeffrey V Nickerson, and Lydia B Chilton. 2023. AngleKindling: Supporting Journalistic Angle Ideation with Large Language Models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 225, 16 pages. https://doi.org/10.1145/3544548.3580907

[43] Daniel Ritchie, Ankita Arvind Kejriwal, and Scott R. Klemmer. 2011. d.tour: style-based exploration of design example galleries. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology* (Santa Barbara, California, USA) *(UIST '11)*. Association for Computing Machinery, New York, NY, USA, 165–174. https://doi.org/10.1145/2047196.2047216

[44] Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. 2022. Palette: Image-to-Image Diffusion Models. In *ACM SIGGRAPH 2022 Conference Proceedings* (Vancouver, BC, Canada) *(SIGGRAPH '22)*. Association for Computing Machinery, New York, NY, USA, Article 15, 10 pages. https://doi.org/10.1145/3528233.3530757

[45] Aneeshan Sain, Ayan Kumar Bhunia, Yongxin Yang, Tao Xiang, and Yi-Zhe Song. 2021. StyleMeUp: Towards Style-Agnostic Sketch-Based Image Retrieval. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Los Alamitos, CA, USA, 8500–8509. https://doi.org/10.1109/CVPR46437.2021.00840

[46] Vishnu Sarukkai, Linden Li, Arden Ma, Christopher Ré, and Kayvon Fatahalian. 2023. Collage Diffusion. arXiv:2303.00262 [cs.CV] https://arxiv.org/abs/2303.00262

[47] Christoph Schuhmann, Romain Beaumont, Richard Vencu, Cade Gordon, Ross Wightman, Mehdi Cherti, Theo Coombes, Aarush Katta, Clayton Mullis, Mitchell Wortsman, Patrick Schramowski, Srivatsa Kundurthy, Katherine Crowson, Ludwig Schmidt, Robert Kaczmarczyk, and Jenia Jitsev. 2022. LAION-5B: An open large-scale dataset for training next generation image-text models. arXiv:2210.08402 [cs.CV] https://arxiv.org/abs/2210.08402

[48] Arseniy Shakhmatov, Anton Razzhigaev, Aleksandr Nikolich, Vladimir Arkhipkin, Igor Pavlov, Andrey Kuznetsov, and Denis Dimitrov. 2023. kandinsky 2.2. https://github.com/ai-forever/Kandinsky-2?tab=readme-ov-file.

[49] Moushumi Sharmin, Brian P. Bailey, Cole Coats, and Kevin Hamilton. 2009. Understanding knowledge management practices for early design activity and its implications for reuse. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 2367–2376. https://doi.org/10.1145/1518701.1519064

[50] Kihoon Son, Kyungmin Kim, and Kyung Hoon Hyun. 2022. BIGexplore: Bayesian Information Gain Framework for Information Exploration. 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 37, 16 pages. https://doi.org/10.1145/3491102.3517729

[51] Yizhi Song, Zhifei Zhang, Zhe Lin, Scott Cohen, Brian Price, Jianming Zhang, Soo Ye Kim, and Daniel Aliaga. 2023. ObjectStitch: Object Compositing with Diffusion Model. , 18310-18319 pages. https://doi.org/10.1109/CVPR52729.2023.01756

[52] Christoph Csallner Soumik Mohian. 2022. PSDoodle: Fast App Screen Search via Partial Screen Doodle. In *2022 IEEE/ACM 9th International Conference on Mobile Software Engineering and Systems (MobileSoft)*. IEEE Computer Society, Los Alamitos, CA, USA, 89–99. https://doi.org/10.1145/3524613.3527816

[53] Ben Swanson, Kory Mathewson, Ben Pietrzak, Sherol Chen, and Monica Dinalescu. 2021. Story Centaur: Large Language Model Few Shot Learning as a Creative Writing Tool. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, Dimitra Gkatzia and Djamé Seddah (Eds.). Association for Computational Linguistics, Online, 244–256. https://doi.org/10.18653/v1/2021.eacl-demos.29

[54] Hideyuki Tamura and Naokazu Yokoya. 1984. Image database systems: A survey. *Pattern Recognition* 17, 1 (1984), 29–43. https://doi.org/10.1016/0031-3203(84)90033-5 Knowledge Based Image Analysis.

[55] Viswanath Venkatesh and Hillol Bala. 2008. Technology acceptance model 3 and a research agenda on interventions. *Decision Sciences* 39, 2 (2008), 273–315. https://doi.org/10.1111/j.1540-5915.2008.00192.x

[56] Andrey Voynov, Kfir Aberman, and Daniel Cohen-Or. 2023. Sketch-Guided Text-to-Image Diffusion Models. In *ACM SIGGRAPH 2023 Conference Proceedings* (Los Angeles, CA, USA) *(SIGGRAPH '23)*. Association for Computing Machinery, New York, NY, USA, Article 55, 11 pages. https://doi.org/10.1145/3588432.3591560

[57] 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* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 435, 19 pages. https://doi.org/10.1145/3544548.3580948

[58] Yunlong Wang, Shuyuan Shen, and Brian Y Lim. 2023. RePrompt: Automatic Prompt Editing to Refine AI-Generative Art Towards Precise Expressions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 22, 29 pages. https://doi.org/10.1145/3544548.3581402

[59] Tongshuang Wu, Michael Terry, and Carrie Jun Cai. 2022. AI Chains: Transparent and Controllable Human-AI Interaction by Chaining Large Language Model Prompts. 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 385, 22 pages. https://doi.org/10.1145/3491102.3517582

[60] Binxin Yang, Shuyang Gu, Bo Zhang, Ting Zhang, Xuejin Chen, Xiaoyan Sun, Dong Chen, and Fang Wen. 2023. Paint by Example: Exemplar-based Image Editing with Diffusion Models. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Los Alamitos, CA, USA, 18381–18391. https://doi.org/10.1109/CVPR52729.2023.01763

[61] Tom Yeh, Kristen Grauman, Konrad Tollmar, and Trevor Darrell. 2005. A picture is worth a thousand keywords: image-based object search on a mobile platform. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems* (Portland, OR, USA) *(CHI EA '05)*. Association for Computing Machinery, New York, NY, USA, 2025–2028. https://doi.org/10.1145/1056808.1057083

[62] Ann Yuan, Andy Coenen, Emily Reif, and Daphne Ippolito. 2022. Wordcraft: Story Writing With Large Language Models. In *27th International Conference on Intelligent User Interfaces* (Helsinki, Finland) *(IUI '22)*. Association for Computing Machinery, New York, NY, USA, 841–852. https://doi.org/10.1145/3490099.3511105

[63] Feifei Zhang, Mingliang Xu, and Changsheng Xu. 2022. Tell, Imagine, and Search: End-to-end Learning for Composing Text and Image to Image Retrieval. *ACM Trans. Multimedia Comput. Commun. Appl.* 18, 2, Article 59 (mar 2022), 23 pages. https://doi.org/10.1145/3478642

[64] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2023. Adding Conditional Control to Text-to-Image Diffusion Models. (oct 2023), 3813–3824. https://doi.org/10.1109/ICCV51070.2023.00355

Appendix A QUERY CONCRETIZATION PROMPT

system_prompt:

```
You are a helpful and creative assistant that can
suggest effective search queries to find new and inspiring
designs. You return your final answer as a valid JSON
object.
```

template:
```
{prompt1}
[Current Search Query]
{curr_query}
{prompt2}
```

prompt1:

We would like to request you to ideate search queries to help designers explore and find useful reference images. The designer has now entered one text query into the image search system. However, there is currently an unspecified part of this query. If the designer looks for search results with this query, he/she can get too many different search results, so the designer wants to be recommended a more specific search query in the query they enter. These are described below.

prompt2:
Please suggest five search queries by following the steps. First, explain the non-specific parts of the current search query and how to specify them. Second, complete the current search query by adding more details to the end regarding color, shape, style, etc. Please add at least three words. Avoid changing the entire meaning of the query, but focus on specifying the unspecified parts in various aspects.

Return your output as a valid JSON object of the following format:
{"explanation":
<explain how you generate the specified queries in the first and second steps>,
"search_queries":
[<list of five suggested queries that designer can use>]}

Appendix B   EDITING KEYWORD SUGGESTION PROMPT

system_prompt:
You are a helpful and creative assistant that can suggest effective search queries to find new and inspiring designs. You return your final answer as a valid JSON object.

template:
{prompt1}
[Description of Current Image]
{curr_image}
[Search Query History]
{search_history}
[Descriptions of Saved Images]
{saved_images}
{prompt2}

prompt1:
We would like to request you to ideate search terms to help designers explore and find useful reference images. The designer is currently looking at an image. They are trying to think about new search terms that can help them find images that are similar but more inspiring than the current image. The designer has already tried various search queries that were unsuccessful in the

past and saved a couple of images to their profile. These are described below.

prompt2:
Please suggest several search terms or words. Consider the current image, the previous search history, and the saved images to predict what the designer's intentions may be. You should imagine what type of design the designer is working on and what type of reference images they may be looking for. If the [Search Query History] and [Descriptions of Saved Images] are empty, just refer only to the [Description of Current Image] to predict the designer's intent. Provide a comprehensive explanation about what you imagine the designer's intention to be and the type of reference images that may satisfy or diversify this intent.

Then, suggest search terms that can help satisfy the designer's intent. You should suggest search terms that designers can add to their search queries to look for images that satisfy their intentions. As an alternative, also suggest terms that can help diversity the designer's intent. These search terms should be different from the designers current intent and should help them explore other, different types of designs. When suggesting search terms, you should avoid suggesting search terms that are already included in the current image, the search history, or the descriptions of saved images. Ensure that your suggested terms are completely new to the designer. Ensure that you only suggest words and avoid suggesting phrases.

Return your output as a valid JSON object of the following format:
{"explanation":
<explain how you generate the specified queries in the first and second steps>,
"aligned_search_terms":
[<list of five suggested words that align with the designer's current intentions>],
"diversified_search_terms":
[<list of five suggested words that differ from the designer's current intentions>]}