석 사 학 위 논 문
Master's Thesis

# 생성형 AI 기반 레퍼런스 재조합을 통한 그래픽 디자인 아이디에이션 과정 지원

CreativeConnect: Supporting Reference Recombination for Graphic Design Ideation with Generative AI

2024

최 다 은 (崔 多 恩 Choi, DaEun)

한 국 과 학 기 술 원

Korea Advanced Institute of Science and Technology

석 사 학 위 논 문

# 생성형 AI 기반 레퍼런스 재조합을 통한 그래픽 디자인 아이디에이션 과정 지원

2024

최 다 은

한 국 과 학 기 술 원

전산학부

# 생성형 AI 기반 레퍼런스 재조합을 통한 그래픽 디자인 아이디에이션 과정 지원

최 다 은

위 논문은 한국과학기술원 석사학위논문으로
학위논문 심사위원회의 심사를 통과하였음

2023년 12월 12일

심사위원장   김 주 호   (인)

심 사 위 원   이 의 진   (인)

심 사 위 원   이 탁 연   (인)

# CreativeConnect: Supporting Reference Recombination for Graphic Design Ideation with Generative AI

DaEun Choi

Advisor: Juho Kim

A dissertation submitted to the faculty of
Korea Advanced Institute of Science and Technology in
partial fulfillment of the requirements for the degree of
Master of Science in Computer Science

Daejeon, Korea
December 12, 2023

Approved by

_____

Juho Kim
Professor of Computer Science

The study was conducted in accordance with Code of Research Ethics[1].

---

**초 록**

그래픽 디자이너들은 종종 레퍼런스의 재조합을 통해 영감을 얻는다. 본 연구의 관찰 연구(N=6) 결과에 따르면 그래픽 디자이너는 이 과정에서 주로 개념적인 키워드에 초점을 맞추며, 키워드의 발견, 확장, 재조합 탐색 과정을 거친다. 동시에 디자이너 스스로의 창의력을 발휘할 수 있는 여지를 충분히 확보하기를 원한다. 본 연구에서 제안하는 CreativeConnect 시스템은 디자이너가 레퍼런스 이미지에서 유용한 키워드를 발견하고, 이와 관련된 새로운 키워드를 추천하며, 사용자가 선택한 키워드로 다양한 재조합 옵션을 생성하고, 그 결과를 텍스트 설명과 스케치로 보여준다. 또한, 이를 위한 생성형 AI 파이프라인을 함께 제안한다. 본 시스템을 이용한 사용자 연구(N=16)에 따르면, CreativeConnect 시스템은 사용자가 레퍼런스 이미지에서 더 많은 키워드를 발견하고, 이를 기반으로 더 많은 재조합 옵션을 탐색하는 데 도움이 되었다. 그 결과, 사용자는 베이스라인 시스템에 비해 CreativeConnect을 사용하였을 때 더 많은 디자인 아이디어를 생성하고, 자신의 디자인 아이디어를 더 창의적이라고 인식했다. 본 연구에서는 창의성 지원 과정에서 CreativeConnect을 다른 유형의 창의성 작업을 지원하기 위해 확장할 수 있는 방법에 대해서도 제안하였다.

**핵 심 낱 말** 창의성 지원 도구, 디자인 아이데이션, 레퍼런스 재조합, 생성형 인공지능

**Abstract**

Graphic designers often get inspiration through the recombination of references. Our formative study (N=6) reveals that graphic designers focus on conceptual keywords during this process, and want support for discovering the keywords, expanding them, and exploring diverse recombination options of them, while still having room for their creativity. We propose CreativeConnect, a system with generative AI pipelines that helps users discover useful elements from the reference image using keywords, recommends relevant keywords, generates diverse recombination options with user-selected keywords, and shows recombinations as sketches with text descriptions. Our user study (N=16) showed that CreativeConnect helped users discover keywords from the reference and generate multiple ideas based on them, ultimately helping users produce more design ideas and higher self-reported creativity, compared to the baseline system without generative pipelines. While CreativeConnect was effective in ideation, we discussed how CreativeConnect can be extended to support other types of tasks in creativity support.

**Keywords** Creativity support tool, Design ideation, Reference recombination, Generative AI

# Contents

# List of Tables

# List of Figures

# Chapter 1. Introduction

References play a crucial role in creative thinking, serving as valuable sources for graphic designers to both grasp the landscape of the existing ideas and ignite novel ones [30, 52, 73, 62]. They offer diverse visual, conceptual, and functional stimuli, allowing individuals to explore various creative directions and draw lessons from established successful examples [3]. One effective method to generate new ideas with references is making a combination of existing examples, which is often called combinatorial creativity [7, 79, 87, 9]. In practice, this is often done through *reference recombination*, which is the process of extracting the elements or aspects of multiple references, considering connections of the extracted elements with those from other references [30], and blending those to ultimately gain novel design ideas [2].

However, each step of recombination requires significant effort from designers. To discover sources for recombination, designers need to dissect the references into individual elements and analyze them to figure out which combinations of elements are worth mixing. Additionally, to find effective methods of blending those elements into a new design idea, they must engage in exploratory efforts by drawing multiple sketches. This takes a long time and multiple iterations, especially for those who are less experienced in the design process, as they have difficulty identifying various factors from references and integrating references from disparate domains compared to professionals [4].

Previous research has provided support for these individual steps. Several approaches have been proposed [39, 35, 34] to decompose the references or show connections between them, aiding users in identifying the sources for recombination. However, these approaches do not guide how to incorporate extracted elements into a design. Also, many approaches have attempted to help users blend different concepts or images into a novel one [12, 13, 91, 84]. However, these approaches primarily emphasize generating precise combinations that effectively incorporate all elements harmoniously, rather than aiming to produce diverse combinations for exploratory purposes. Another thread of research focuses on searching by genetic recombination [45, 14, 82, 88], but these techniques were focused on widening the range of the design exploration, rather than offering inspiration on how to effectively combine specific design elements.
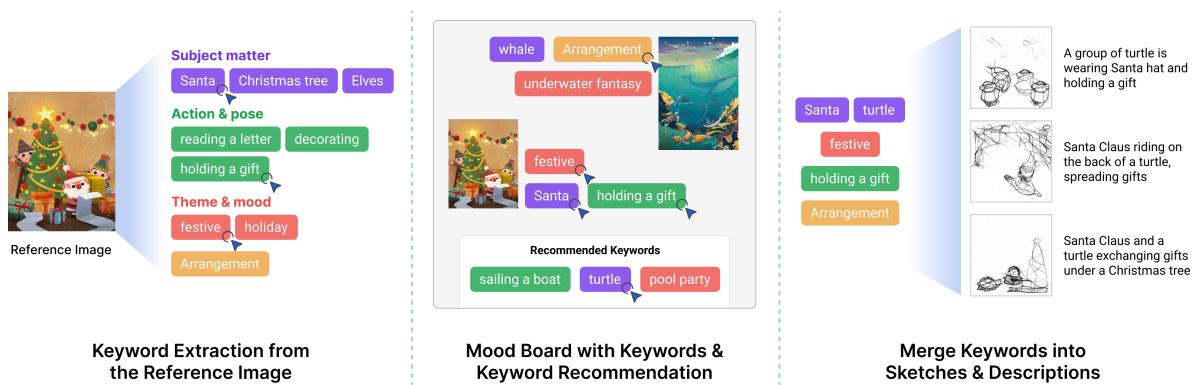


Figure 1.1: Three main features of CreativeConnect supports for enhancing reference recombination.

Through a formative study with six novice graphic designers and design students, we aimed to understand the process of reference recombination and identify challenges that they have during the process. There were two distinctive stages of ideation, 1) conceptual ideation, aiming to effectively convey the design topic, and 2) visual ideation, deciding style-wise details on top of the selected concept. We decided to focus on the conceptual one as the recombination of references tends to be more prevalent in this stage. During the conceptual ideation, designers extracted four types of elements from the reference—subject matter, action & pose, theme & mood, and compositional aspects (arrangement). Then, participants tried to brainstorm more elements related to the extracted ones and combine them in several ways. However, due to the high effort required to manually recombine them, they were concerned that they couldn't try out all of the possible methods. They also mentioned that support to help ideation should not be in an overly completed form as it can diminish their own input. With these observations, we propose four design goals for a reference recombination support system: (1) enable users to effortlessly specify the four types of conceptual elements from the reference image, (2) recommend relevant elements, (3) provide many recombinations as much as possible, and (4) intentionally keep the generated output partially unfinished, fostering a space for user creativity.

Based on the design goals, we propose CreativeConnect, a system that supports the ideation of visual designers by helping them extract elements from the reference images and generate a wide range of recombinations of those elements. Using CreativeConnect, users can easily discover and select elements from the reference image based on the four element types and can get recommendations for more relevant keywords. Once the user has chosen the keywords to combine, they get a variety of recombination options presented as pairs of sketch images and one-line descriptions. We introduced novel pipelines with generative models to automate the extraction of keywords from images, generate recombination options, and transform them into descriptions and sketches.

We conducted a within-subjects study with 16 design students, and compared CreativeConnect with the baseline, which consisted of a mood board with manual keyword notes, a layout diffusion model, and ChatGPT. Results showed that CreativeConnect could support the two stages of the reference recombination process, discovering elements and generating design ideas by recombining them. Participants were also able to produce more design ideas in a given time, and perceived that CreativeConnect helped them come up with more creative sketches compared to the baseline. They emphasized that Creative-Connect's features were especially beneficial when they wanted to explore ideas vastly different from their initial concepts. We compared the creativity support of CreativeConnect with baseline and proposed an opportunity to design a comprehensive recombination support tool that could support a wide spectrum of design needs and situations. We also found that the sketch-based output leads users to get more stimulus for their creativity by letting users further imagine through the low-fidelity. Finally, we discussed the generalizability of CreativeConnect in terms of the designer's expertise, collaborative settings, and different domains of design.

This work presents the following contributions:

- CreativeConnect, a system that aids graphic designers in their ideation by helping them extract elements from reference images and suggesting a wide range of recombinations of those elements.

- Computational pipelines with generative models that automate the extraction of keywords from images and generate recombination options in descriptions and sketches.

- Findings from a user study about how CreativeConnect can aid users in each step of recombination, leading to the generation of more design ideas and users to perceive their ideas as more creative.

# Chapter 2.  Related Work

This work aims to support designers in their reference recombination process for creativity. In this section, we review previous literature on (1) how references are used in graphic design ideation, (2) how recombination is employed for creative thinking, and (3) previous generative AI approaches for creativity.

## 2.1   Reference in Graphic Design Ideation

The creative process begins by collecting relevant inspirational materials from various sources [77, 22]. Designers leverage these collected examples to gain a comprehensive understanding of the problem space. As the process advances into idea generation, these compiled examples play a pivotal role in fostering creativity, igniting new ideas through analogical thinking [33, 28]. Recognized as one of the most challenging phases in the entire design process, previous research on creativity-supporting tools has extensively concentrated on enhancing this ideation step [24]. Previous research demonstrated that designers get valuable insights and inspirations in different ways [30], and many studies have delved into the significance of these references in design thinking, showing their potential to stimulate creativity and innovation [3, 78].

One of the primary approaches to support idea generation with references is to help designers see diverse references. Exploring diverse ideas is crucial in terms of preventing fixation [37], in which a designer becomes overly fixated on a single concept, potentially hindering creativity and innovation. Therefore, Zhang et al. [89] have utilized a Generative Adversarial Network (GAN) for exploring diverse images, while Matejka et al. [60] developed the Dream Lens to assist in exploring generative 3D design solution space.

Another avenue of research is to help designers manage their inspirations drawn from references, particularly through the use of mood boards [22]. Prior research demonstrated that building a mood board can enhance the comprehension and interpretation of ephemeral elements in design [26], which is beneficial for both defining and resolving design challenges [8], and ultimately leading to a boost in creativity [58]. Therefore, many computational systems have been proposed to help designers to build an interactive mood board, such as Funky Wall [59], SemanticCollage [49], and May AI [48].

While this paper primarily focuses on the recombination of references, we have integrated two significant insights from prior research about design references. First, we emphasize the importance of offering users a diverse array of images to support their creative process. Second, we have incorporated the concept of a mood board as a valuable tool for organizing references within our system.

## 2.2   Recombination for Creative Thinking

In the creative thinking process, new ideas often come through the combination of the existing examples [7, 79]. It was shown that creativity often arises from forging new associations among previously unrelated frames [50]. There are two crucial components in this process: recognizing the differences between existing concepts and blending them [64, 2]. Also, the diversity of the given examples is important for building novel associations between them during this process [63]. Observations of designers' creative processes showed that designers often maintain multiple small components and keep employing them to

generate new variations through a process akin to recombination [23]. Many computational systems were also proposed for building recombinations and verified to be effective in tasks such as chair design [87] or text-based ideation [9].

One practical implementation of this concept in terms of design ideation is genetic exploration. Genetic exploration involves generating novel solutions by merging elements from preexisting designs to widen the range of references. This approach has been applied in diverse domains such as garden design [45], 3D modeling [14, 70], architecture [82], and 2D graphics [88]. However, the primary aim of this approach is to enrich the reference in the information-gathering stage by utilizing existing references, rather than supporting designers to generate their own ideas from those recombinations in the next stage.

In recombination, it is also critical to decompose the reference and get elements that are worth combining. Therefore, several tools have been developed to facilitate this process, especially by automatically decomposing the original source and showing the fine-grained aspects. CollageMachine [44] decomposes websites and makes them into an interactive collage. MetaMap [39] provides a decomposed view of the reference image into three dimensions (semantic, color, and shape) and lets users explore more references using it. Hope *et al.* [34] divides the product's information into fine-grained functional parts, allowing users to combine the inspiring part. MoodCubes [35] offers a new mood board experience by decomposing multimedia references into constituent elements and using it to provide suggestions for new inspirational materials. They may not, however, directly discuss the exact strategies for merging these outputs as a new design idea. On the other hand, VRicolage [80] enables users to decompose objects into different parts, motions, or colors, and mix them. However, this process was more of utilizing collected assets, rather than generating a new idea from recombination.

Additionally, there were many previous approaches to support the process of mixing the reference images or concepts. For example, VisiBlends [12] and VisiFit [13] introduced a novel pipeline to blend two objects to convey integrated meaning. ICONATE [91] supports users to generate a new icon by mixing different icons, and PopBlends [84] automatically suggests conceptual blends of reference images. FashionQ [38] supports this blending in the domain of fashion design. Artinter [15] supports recombining style elements from the reference to facilitate communication. Nevertheless, these approaches primarily focus on seamlessly merging entire references rather than breaking them down to the element level. This approach may not fully align with the creative recombination process, which often begins by identifying specific elements to combine within the provided examples. 3DALL-E [57] presents a recombination workflow for generating a new idea, which suggests diverse low-level keywords and combines them into a prompt for text-to-image models. This approach, however, differs from our definition of reference recombination as the keywords are from LLM's understanding of the world rather than the design references.

## 2.3   Generative AI Approaches for Creativity

Before looking into AI systems for creativity, it's important to know how visual designers perceive AI for supporting their design tasks. Ko *et al.* [47] looked into how graphic designers use large-scale text-to-image generation models (LTGMs) to help with their creative works and suggest the design guidelines for building creative supporting systems using them.

Recently, diffusion-based techniques [74, 72, 65] and CLIP embedding [71] enable people to represent their ideas to visual materials quickly and easily using text prompting. There were also many previous approaches to incorporate inputs with additional modalities, such as layout [55, 92, 10] or sound [53, 81].

Techniques to add extra conditions and styles for more granular control have been proposed as well [90, 61]. There is also a thread of research on modifying the generated image to align with user intent better, such as adding style [27], latent-space manipulation [42, 43], human-prompt editing [6], and editing a specific part in generated images [75, 25].

With those novel ML techniques, the creative landscape is also continuously being reshaped, offering innovative solutions and enriching the artistic experience. Promptify [5] stands out as an iterative prompt refine tool, letting users get closer to their intended result by clearing unintended outcomes. PromptPaint [16] allows users to go beyond language to mix prompts to express challenging concepts, supporting for iterative shaping of the image. On the other aspect, the interplay between humans and AI is also fast evolving. The concept that Karimi *et al.* [41, 40] proposes is a generative AI system that helps designers by collaborating during the design phase, instead of taking over the design process. Oh *et al.* [66] and Framer [51] proposed a user-AI collaborative interface to allow a co-drawing experience.

While there has been a lot of research on expressing user intention to ML models accurately to get a better image or collaborate with AI during the design execution phase, it is less relevant to the ideation task of expanding the variety of ideas. Especially, it is still unknown how to design interaction with generative AI models to inspire graphic designers by recombining the references.

# Chapter 3.  Formative Study

To understand how designers recombine design references for ideation and what challenges they encounter during the process, we conducted a formative study. As indicated by prior research [4], novice tends to encounter more challenges in getting inspiration from references and combining them. Therefore, we recruited novice designers as they are expected to have an understanding of the overall design process but still struggle with many challenges in ideation through recombination. Six participants (6 female; age M=25.3 and SD=3.32) were recruited through an online recruitment posting. Two were professional UI/UX designers with 1 year of experience each, and one was a freelance brand designer. Three were students majoring in industrial design, with two at the graduate level and one in their fourth year of undergraduate studies. All participants reported that they had experience in at least three different graphic design projects before.

## 3.1   Study Process

The study included two parts: (1) an observation on reference searching and idea sketching, and (2) a semi-structured interview. For the first part of the study, the task was to draw an illustration for one of three different design topics: "Tourism service for kids", "Pet grooming service", or "Eco-friendly restaurant". We let participants choose one of these topics. They were first given 10 minutes to search for reference images that they wanted to utilize in their ideation process. For each reference they chose, they were asked to describe what aspects of the references they found appealing. Then, participants sketched their design ideas using their preferred method for 30 minutes. Three participants used pen and paper to sketch their ideas, while the other three used a tablet and digital drawing software. They were asked to generate at least three different design ideas, and they were prompted to describe how they integrated their references into each sketch. After that, we conducted a semi-structured interview to ask about their challenges in generating multiple ideas using references.

Following each study, two of the authors independently coded the recombination methods employed by participants in their ideation tasks and the semi-structured interview results. The coded data were then discussed collaboratively. After conducting six studies, codes were saturated and no further study sessions were conducted.

## 3.2   Findings

Through the observation of participants' design processes, we discovered that the recombination of different references primarily occurs during the initial stages of design ideation, with a specific focus on the conceptual aspects of the reference images rather than the visual elements. We identified four distinct categories of elements employed in this process. We also found specific challenges associated with it and observed that the system supporting this process should reserve a degree of incompleteness to encourage creativity.

### 3.2.1 Early-Stage Design Ideation Focuses on Conceptual Aspects

All six participants said that they refer to the references in two distinct stages: conceptual ideation and visual development. During the conceptual ideation stage, designers focus on elements that could effectively convey the design topic, such as objects or mood. After looking at those elements, they generated multiple drafts by combining them in several ways. On the other hand, the visual development stage revolved around adding visual details like color and texture to complete the sketches derived from the conceptual ideation. During this stage, designers often had a clear direction in mind and referred to a specific set of references that aligned well with their chosen direction, with less emphasis on exploring different recombinations of diverse references. This aligns with findings from previous research [32], which indicates that artists engage in a spectrum of reference usage in their creative process, ranging from detailed recreation (visual development) by tracing images to interpretive inspiration for high-level components (conceptual ideation). In summary, designers recombine references primarily for conceptual ideation, which is usually the first step of the design ideation, suggesting that a system supporting the reference recombination process should focus on how to facilitate this early-stage step.

### 3.2.2 Types of Elements Used for Recombination

During the conceptual ideation phase, participants tried to extract specific elements from references and incorporate them into their design concepts. They employed a variety of approaches for this. The simplest approach that was observed in all participants was to utilize objects in a reference in their own sketch. For example, for drawing an illustration for "Tourism service for kids", one participant took an image of a paper plane from a reference to convey the image of playful children and a tour service at the same time. Five participants extracted the abstract semantic meaning or overall theme conveyed by references. For example, after looking at an image of a person holding a pamphlet and deep in thought, one participant said that the keyword "imagination" could effectively capture the concept of kids. So they came up with a design concept about children imagining various travel destinations. Another approach observed in three out of six participants was to draw the action of a character in a reference. For instance, from a reference depicting an animal and a person holding hands, a participant got the concept of children holding hands together. Lastly, five participants extracted the composition from the references. For example, by looking at a reference where leaf shapes were arranged together to form a shovel, one participant came up with the idea of using multiple tree trunk shapes to represent the structure of a building.

### 3.2.3 Challenges During Finding Elements

We identified some opportunities to support the process of extracting elements from the references. There were many cases where the elements that designers initially found appealing in the reference search phase differed from the elements they eventually utilized in their design concepts. In the interview, participants said that upon closer examination of the references, they discovered new elements of interest and incorporated them. This means that designers couldn't immediately extract elements upon viewing the reference and it often required several examinations to uncover such elements, which is time-consuming.

Another observation was that participants often came up with new keywords based on what they had already found for further brainstorming at the element level. For instance, P3 identified "toy blocks" from one reference and "train" from another reference, then came up with the new keyword "toy train" and incorporated it into their final idea. However, this process was often more challenging compared to

finding elements directly from the reference images. P4 highlighted an opportunity for system support for this by mentioning that "I usually talk with others about my ideas, which leads me to discover new keywords related to the original one. Just like that, I think it would be nice if the system could recommend a new keyword to expand the design idea I have now."

### 3.2.4 Challenges During Recombining Elements

After finding out the elements from the references that they want to utilize in their own design ideas, another challenge became apparent. While there can be numerous ways to combine these elements, participants are often frustrated as they can't sketch out all the possibilities to determine if they are viable. Three out of six participants expressed anxiety about not being able to consider all possible combinations. P3 stated, "I always feel anxious that there might be a better way, but I can't think of it." P6 also mentioned that "The more options I explore, the more I become confident about my final design idea. I want some faster way to explore alternatives as much as possible." Four out of six participants said that they rely on their imagination to envision numerous recombination possibilities within their minds, as it is too time-consuming and effortful to sketch out all. However, two participants expressed frustration that, although combinations seemed good in their minds, they might not come together as effectively in actual sketches.

### 3.2.5 System Support should be Incomplete

Designers tend to deliberately exclude visual details during conceptual ideation. Participants said that when recombining the references for conceptual inspiration, they did not pay attention to visual details, and several participants said that they even needed to intentionally exclude those details. P2 stated, "When combining different concepts, colors and textures often become messy, so I deliberately use the same brush for all elements." P3 agreed this with another viewpoint, by expressing concern about becoming overly fixated on frequently recurring visual details while exploring conceptual recombinations. We also asked the participants if they could get recommendations for different recombination options, what form would be preferred. Four participants mentioned that they would prefer incomplete outputs such as a sketch or even a textual description of the idea so they could focus on the concept itself. The main reason for this was the concern that the model would compromise their creativity or lead them to perform unintentional plagiarism.

## 3.3 Design Goals

Based on the findings of the formative study, we identified four design goals to build a system to support designers' reference recombination process during early-stage ideation.

DG 1. **Facilitate Element Extraction from References.** To help users efficiently find the elements that would be used for the recombination, the system should help users discover the overlooked elements. Based on our observation, elements that users want to extract from references are (1) subject matters (e.g., objects, characters, landscapes), (2) action & pose, (3) theme & mood, and (4) arrangement.

DG 2. **Suggest Diverse and Relevant Elements.** To help users explore more elements on top of what they found from the references, the system should provide some recommendations of relevant

elements that users might like.

DG 3. **Generate Diverse Recombination Options.** To help users explore diverse recombination possibilities, the system should show users a varied range of recombination options and reduce their anxiety over not considering all feasible combinations. This goal highlights the system's ability to propose combinations that the user might not have considered independently.

DG 4. **Present Recombination in an Incomplete Format.** To align with designers' preference for conceptual sketches over highly detailed artwork during the initial ideation phase, the system-generated outputs should be intentionally incomplete, such as sketches. This emphasizes the importance of allowing users to inject their own creativity into the images.

# Chapter 4.  System

With derived design goals, we implemented CreativeConnect (Figure 4.1), an AI-powered design tool that supports graphic designers in coming up with novel design ideas by recombining reference images in early-stage conceptual ideation. CreativeConnect mainly consists of a mood board where users can import reference images and select what they like about the reference. When the image is imported, the system extracts keywords from it according to the four categories defined in the formative study (Section 3.2.2). This helps users easily discover and select keywords (DG 1). Selected keywords are then displayed on the mood board along with the images. CreativeConnect offers further keyword recommendations based on the keywords users have added on the board or their specific selections (DG 2). Also, when the user chooses a set of keywords that they want to recombine, the system generates multiple drafts with diverse ways of combining them (DG 3). All of the system-generated images are produced in line sketches with one-line descriptions so that users can further reinterpret by themselves (DG 4).

## 4.1  User Scenario

To demonstrate our system, we show how Sarah, a junior illustration designer, uses CreativeConnect to generate ideas for her design project. Sarah recently accepted a new commission to draw an illustration for the cover of a children's book titled, "A Christmas Dinner in the Underwater World". As the given topic is an unusual combination of two themes, she felt difficulties in getting inspiration from the references and mixing them to come up with ideas, so she decided to explore references with the help of CreativeConnect.

### 4.1.1  Getting User Inputs on the Design Reference

Sarah first uploads 10 reference images she got from her client into CreativeConnect. Looking through the references, she is intrigued by the one where two scuba divers swim with a turtle. When she chooses the image, CreativeConnect shows some keywords that can be found in the image, divided into four categories – subject matter, action & pose, theme & mood, and arrangement (Figure 4.1 (a)). As she found the scuba diver concept interesting, she clicks on the subject matter category. In the keyword list, she finds "scuba diver" and clicks it. She also finds "coral reef" on the list, which she didn't recognize before. She looks at the references again and thinks coral reefs would look great in her illustration, so she clicks "coral reef" as well. Similarly, she looks through the list of the keywords in the "action & pose" and "theme & mood" categories and selects "swimming" and "adventure" from each list. She also liked the overall composition of the image, so she clicks its "arrangement" as well. She also works on selecting keywords that she likes on other references.

### 4.1.2  Mood Board with the User-selected Keywords & Keyword Recommendation

As Sarah selects the keywords she finds useful from each image, the canvas of the CreativeConnect offers a dynamic mood board that shows the references with user-selected keywords, capturing her creative goal and preferences (Figure 4.1 (b)). As she freely moves the images to organize them, the

**(a)** Keyword Extraction panel

**(b)** Interactive Mood Board & Keyword suggestion panel

**(c)** Keyword Merge panel
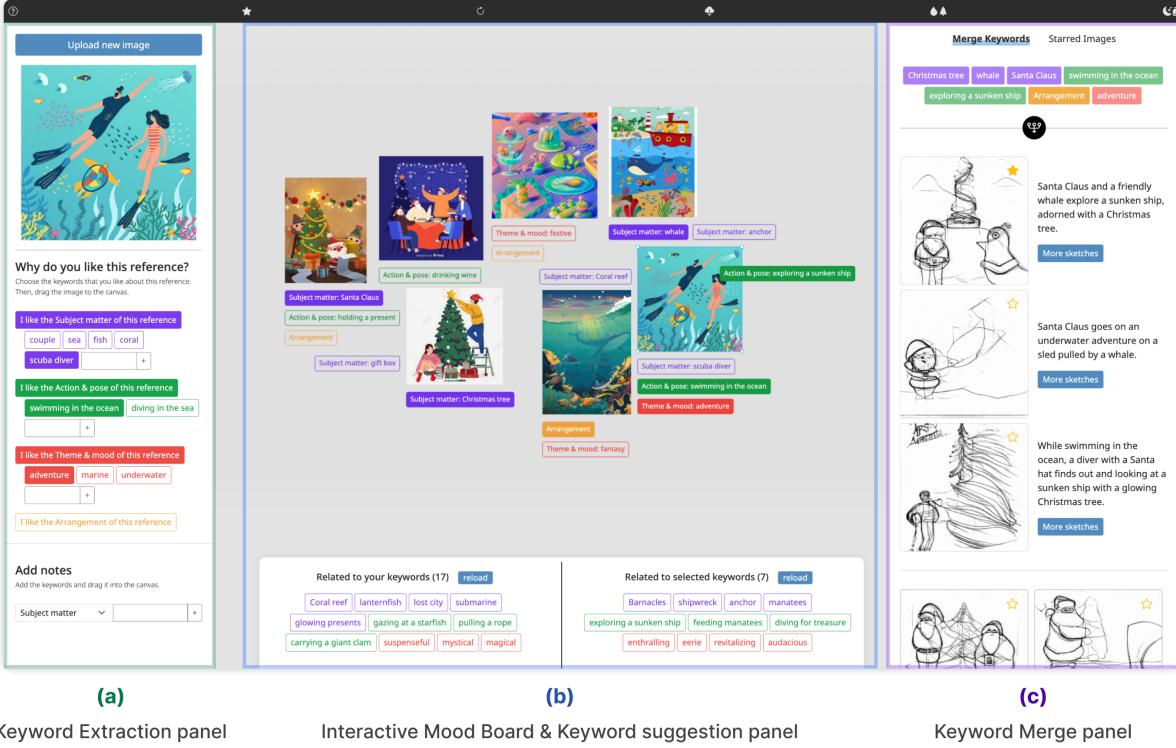
Figure 4.1: Screenshot of CreativeConnect. (a) **Keyword Extraction Panel**: The system automatically extracts keywords in four categories (subject matter, action & pose, theme & mood, and arrangement) from the reference image. Users can select these keywords, or add keywords manually. (b) **Interactive Mood Board with Keyword Suggestion Panel**: Users can organize the reference images on the mood board, along with the selected keywords. Users can also import system-recommended keywords shown below, suggested based on all keywords on the board or the keywords that users selected on this mood board. (c) **Keyword Merge Panel**: When users select keywords that they want to recombine on the mood board, the system generates sketches and their respective descriptions including all of the selected keywords. Users can view more generated sketches by clicking the "More Sketches" button.

selected keywords move along with the image. By looking at the keywords, Sarah wants to come up with additional ideas for character actions that align with the adventurous theme, similar to swimming or scuba diving. Therefore, she selects "subject matter: scuba diver", "action & pose: swimming", and "theme & mood: adventure" to get system recommendations with these keywords. CreativeConnect shows a set of keywords, such as "action & pose: exploring sunken ship", and "subject matter: anchor". Sarah finds those keywords valuable, so she drags them into the mood board.

### 4.1.3 Recombining Design References using Keywords

From the set of keywords on the mood board, Sarah now selects some keywords that she wants to include in her design and uses the system to make a first draft. She selects "Christmas tree" and "Santa Claus" for a Christmas dinner theme, and "whale", "swimming", "exploring the sunken ship", and "adventure" for the underwater theme. She also selects the "arrangement" of one of the images with an interesting composition.

After clicking the merge button, CreativeConnect generates three different drafts, each showcasing a unique and different way of incorporating these keywords (Figure 4.1 (c)). Each draft contains a one-line text description of the image concept and a sketch-style generated image based on the description and the arrangement that Sarah selected. Sarah appreciates the generated results as the way each draft combining the keywords would be difficult to think of by herself and she likes how all three drafts are distinct from each other. Also, the sketch format allows her to further imagine design concepts rather than fixating on the concept and details in the generated results.

Among the drafts, Sarah finds one description interesting: "Santa Claus goes on an underwater adventure on a sled pulled by a whale." However, she feels dissatisfied with the generated sketch, so she presses the "More Sketches" button. Then, CreativeConnect generates five more sketches with the same description, but in a slightly different way. She gets some good design ideas from the new sketches and starts working on her draft.

## 4.2 Technical Details

CreativeConnect was built as a web-based system with a ReactJS[1]-based front-end client and a Flask[2]-based back-end server. We implemented ML pipelines for extracting the keywords from the references and merging keywords into recombinations. The technical details of these pipelines are discussed in the following sections. Some examples of outputs from the pipeline are presented in Figure 10.1 in the Appendix.

### 4.2.1 Extracting Keywords from Reference Images (Figure 4.2. (a))

Based on the findings from our formative study, our pipeline is designed to extract keywords from a provided reference image in four categories: subject matter, action & pose, theme & mood, and arrangement. To achieve this, we follow a multi-step process.
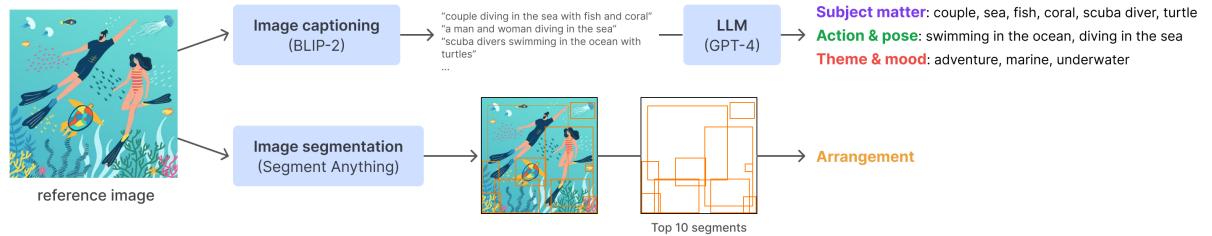
To identify the subject matter, action & pose, and theme & mood within the image, we initially employ an image captioning model BLIP-2 [54] to generate textual descriptions of the image contents. For a comprehensive understanding of the entire image, we divide it into $3 \times 3$ segments and generate

---

[1]https://react.dev/
[2]https://flask.palletsprojects.com/

Figure 4.2: Technical pipeline of CreativeConnect. (a) **Keyword extraction from image**: The caption generated from the image captioning model goes into the LLM to extract subject matter, action & pose, and theme & mood. The segmentation model is used to detect the image's arrangement. (b) **Keyword-based image generation**: the LLM generates descriptions based on the given keywords, and the layout variator generates similar arrangements. The image generation model generates the image, and the style transfer model converts this into a sketch.

captions for each segment as well as the whole image. These captioning results are then processed by GPT-4 [68], a Large Language Model (LLM), to extract lists of subject matter, action & pose, and theme & mood present in the image captions. Prompts used for this are in the Supplementary Material.

For the arrangement, we utilize the Segment Anything model [46] to identify the top ten prominent segments within the image. We then used the approach from LLM-grounded Diffusion [56] to score these segments. Bounding boxes around these segments provide information about the image's overall structure, such as where the items are placed and where large negative spaces are.

Additionally, for generating the recommendations of the relevant keywords, we also used GPT-4, and the prompts used for this are in the Supplementary Material.

### 4.2.2 Generating Recombinations (Figure 4.2. (b))

When the user selects a set of keywords to generate a new recombination, our system generates a range of options to mix those keywords.

The system first generates three textual descriptions that encompass the selected subject matter, action & pose, and theme & mood keywords. Then, it extracts the list of the objects that must be drawn on the image for this description. For this, we used few-shot prompting with GPT-3.5-turbo [67]. The used prompt is in the Supplementary Material. For the arrangement, we developed a layout variator to create layouts similar to the selected image's arrangement while aligned with the generated text description. The layout variator first applies an empirically defined random variation of -50 to 50 pixels on each bounding box component (i.e., x, y, w, h) in the original arrangements. Then, it randomly selects boxes depending on the number of objects that need to be drawn and sorts highly similar layouts first using the similarity metric. The similarity is calculated by summing the IoU and the complement of the min-max normalized centroid distance between the closest pairs of bounding boxes. Following this similarity, the top five arrangements are utilized for the recombination generation. The most similar layout is used for generating the image in the initial iteration, and other layouts are used when the user requests more sketches. A few shot prompting with GPT-3.5-turbo is used to map between the arrangements and the objects to create the best image possible. The full prompt used is in Supplementary Material.

However, when the user does not select any arrangement from the references, the system generates a broader range of diverse layout options. A few-shot prompting pipeline using GPT-3.5-turbo generates the three most appropriate layouts for the given text description and object list. This pipeline is built based on the previous work [56], and the full prompt for this is in Supplementary Material.

Given the textual description and the list of the objects mapped with the generated layout, the system generates images with a layout diffusion model [55]. Following our design goal, the system converts the generated image into a simple line sketch using the U-Net structured style transfer model [20].

## 4.3 Technical Evaluation

We evaluated ML-based pipelines, especially for keyword extraction, keyword recommendation, and textual description generation by merging keywords.

### 4.3.1 Keyword Extraction Pipeline

We built a dataset consisting of 100 images and tags of each image categorized by the subject matter, action & pose, and theme & mood. We asked 20 people with expertise in design or HCI to annotate five images each. On average, there were 5.03, 1.87, and 2.29 keywords collected per image in the category of subject matter, action & pose, and theme & mood, respectively.

Using this dataset as ground truth, we evaluated the prediction result from the keyword extraction pipeline. Keywords in subject matter and action & pose categories were matched manually one by one between similar ones. The precision and recall of our pipeline were 94.2% and 58.2% in subject matter, and 35.3% and 51.3% in action & pose. Although some salient keywords in the dataset were missing, the pipeline provided quite accurate keywords in the subject matter. The predicted action & pose keywords were not perfectly aligned with the dataset tags, but they were still acceptable in user-side because they were perceived as similar to users even if they were not completely accurate (e.g. for an image of a cat standing straight, our pipeline predicted "stretching arms", while the ground-truth is "dancing"). For theme & mood keywords, we calculated the cosine similarity of mean embedding vectors [85] of ground-truth and predicted result to compare the semantic similarity. This was because for theme and mood, even if words are not exactly the same, there can be many other words that can be accepted as similar. The similarity of the ground truth and prediction was 0.826, which means the keyword extraction model estimates the theme & mood words quite closely. Examples of the predictions are presented in the Appendix (Figure 10.2).

### 4.3.2 Keyword Recommendation Pipeline

We evaluated the keyword recommendation pipeline based on whether there was a proper level of similarity between the original keywords and the recommended keywords. This was because it would be useless if the recommendation is too similar or too irrelevant to the original keywords.

We randomly sampled three to ten keywords from each image-keyword pair in the dataset and made 100 sets of keywords. Then, from the pipeline, we got the recommendations for each set. To verify whether these recommendations have a proper range of diversity, we generated two comparison groups of keywords, the irrelevant group and the synonym group. The irrelevant group consists of random keywords from the dataset, and the synonym group is generated by paraphrasing the keyword in each set. NLTK [1] and GPT-3.5 were employed to find synonyms. Then, we used the text embedding [85] to calculate the cosine similarity of each group with the original keywords.

The similarity of the irrelevant and synonym groups to the original keywords was 0.624 and 0.774, respectively, and the recommended keywords had a similarity of 0.696, which is in the middle. This shows that our recommendations are less similar to original keywords than synonyms, but more similar compared to the irrelevant phrases.

### 4.3.3 Recombination Generation Pipeline

The recombination generation pipeline gets a user selection of a set of keywords and generates three different descriptions of the possible image that includes those keywords. As the pipeline aims to provide diverse options, we evaluated the diversity of the description generation model.

Similar to section 4.3.2, we built 100 sets of keywords randomly extracted from the dataset. For each set, we generated three descriptions using our pipeline, calculated the cosine similarities between those three, and averaged them. Here, we calculated diversity as $1 - similarity$. To validate our

description generator, we prepared two more description sets, one consisting of explicitly unrelated descriptions that were randomly acquired from the dataset, and the other one consisting of descriptions that merely paraphrase one of the generated descriptions using paraphraser with T5-based model [83]. The diversity within the random and paraphrased groups were 0.801 and 0.209 respectively, while the generated descriptions from our pipeline show a diversity of 0.395. This shows that generated output is more diverse than just paraphrasing and less diverse than random ones, which means that the pipeline generates descriptions of a reasonable amount of diversity.

We didn't evaluate the later part of this pipeline which is about generating images and transforming them into sketches, as we used models from previous research [56] without any customization or adaptation.

# Chapter 5. Evaluation

We conducted a within-subjects comparative study with 16 participants. As our design goals encompassed two steps of the reference recombination – (1) Finding elements (DG 1 and DG 2) and (2) Recombining elements (DG 3 and DG 4), we first observed how CreativeConnect supported each of these steps. We also evaluated whether CreativeConnect eventually improves designers' idea generation results and how it supports the creative process.

- RQ1. How does CreativeConnect support the two steps of the recombination process—finding elements from the references and recombining elements?

- RQ2. Can CreativeConnect help users generate better quality and quantity of design ideas?

- RQ3. How do users utilize CreativeConnect's outputs in their ideation process?

The baseline system shared a similar interface with CreativeConnect but without the key features of CreativeConnect—extracting keywords from the reference, suggesting relevant keywords, and generating recombination options. In this baseline system, users could manually leave keyword notes on each reference image, create sketches by specifying layouts and prompts to the image generation model, and use ChatGPT[1]. To assess the efficacy of the design of CreativeConnect's features and pipelines rather than the effect of AI functionalities, the same AI functionalities are also included in the baseline system. After observing prevalent use cases of AI in design processes through recent survey [69] and videos [93, 21, 76], we included both the language model and the image generation model in the baseline system to simulate real-world scenarios of designers with AI tools. To prevent the model performance from affecting the study results, the baseline included a model closely aligned with the CreativeConnect pipeline. Instead of the GPT models, we provided GPT-3.5-based ChatGPT, and for image generation, we provided the same layout diffusion model as CreativeConnect. The screenshot of the baseline interface is presented in Figure 10.3 in the Appendix.

## 5.1 Participants

We recruited 16 participants (10 female, 6 male; age M=24.81 and SD=3.78) through an online recruitment posting. To figure out whether the CreativeConnect can handle the challenges found in the formative study with novice designers, our participants were set as a similar group to the formative study. We required participants to have a degree in design or art and have participated in at least 3 different design projects. 11 participants were students majoring in design—5 participants at the graduate level and 6 participants at or above the third-year undergraduate level. The other 5 participants have graduated—2 majored in design, 1 minored in design, while others pursued majors in media arts and painting.

All of the participants also reported themselves as having enough sketching skills to express their design ideas, since we asked participants to draw their ideas during the task. The study was conducted for 2 hours, and we compensated participants with 70,000 KRW (approximately 53 USD).

---

[1]OpenAI. (2023). ChatGPT (August 3 Version). https://chat.openai.com/
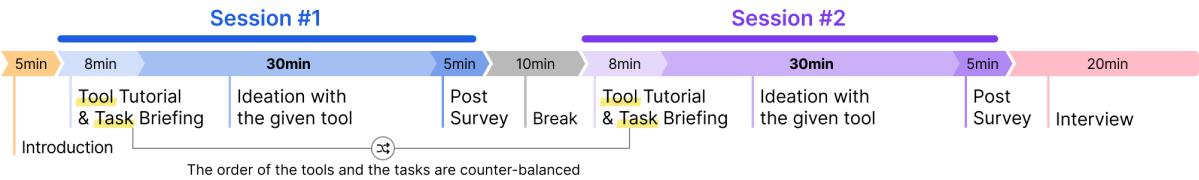
Figure 5.1: User study process. The 2-hour user study consists of two sessions with different tools, each including a 30-minute ideation phase utilizing the given tool. The order of the tool and the design tasks are counterbalanced. After the two sessions, they had 20-minute semi-structured interviews about their experiences.

## 5.2 Study Procedure

The whole process of the user study is shown in Figure 5.1. Participants were asked to perform design ideation tasks twice in two settings: CreativeConnect and baseline. The task was to draw an illustration for the cover of a children's book named "Starry Safari: Exploring Alien Jungles" or "A Christmas Dinner in the Underwater World". They were also provided with 10 reference images for each topic. The order of topics and tools was counterbalanced for each participant.

For the first five minutes of each round, participants were provided a tutorial on the given system interface and tried out the system with sample images to get used to the system. They were then given the topic and the reference images, and started ideation using the given tool for 30 minutes. If the participants came up with a design idea that they wanted to further develop, they were asked to sketch it on the paper using a pen. After each round, they completed the post-task survey. Between the two rounds, they could get a 10-minute break if they want. After both rounds, we conducted a 20-minute semi-structured interview to ask about the difference between the two conditions and the effect of the tools on their ideation process. The interview questions are presented in Appendix 10.2.2.

## 5.3 Measures

The post-survey after each round included questions about the usefulness of the given system for the different steps of the ideation, which are (1) organizing the references, (2) discovering useful elements from the reference, (3) exploring multiple ideas, (4) discovering new ideas, and (5) exploring multiple ideas. The survey also included five questions about satisfaction with the overall outcome, quantity, quality, diversity, and creativity of participants' own sketch results. We also included five questions from [86] to assess participants' self-perceived experience of using the AI system. Participants answered these questions only for the image generation feature and ChatGPT after the baseline session, and for the keyword extraction, keyword recommendation, and image/description generating features after the CreativeConnect session. Also, the survey included the Creativity Support Index [11] and NASA-TLX questionnaire [29].

We also gathered the usage logs (i.e., participant actions with timestamps) to get quantitative metrics for user behaviors. We used this data to calculate the time taken for each sketch, the number of images generated, the number of inputs provided to the image-generating model, etc. Also, every time the participants completed the sketch, the system prompted participants to rate how well the given tool assisted them in producing the idea.

Additionally, we conducted an expert evaluation of the participant's sketches. We recruited two

experts, both graduating with a bachelor's degree in art and having 6 and 1.5 years of art education each. We asked them to evaluate two factors in the 7-point Likert scale: (1) the creativity of each sketch and (2) the diversity of ideas within a set of sketches. We randomly chose three sketches drawn by each participant on each design topic, and a total of 96 sketches (3 sketches x 16 participants x 2 conditions) were evaluated. The evaluators rated the sketches individually, and for cases of significant score differences (more than 3 points), we asked evaluators to re-evaluate them. While re-evaluating, they were given each other's comments and scores and could choose to change their original score or leave it. They also had to leave comments about their decision as well. There were 9 sketches that required re-evaluation, and all of the conflicts were resolved after one round of re-evaluation. After that, we used the average score for the two evaluators' scores for the result analysis.

# Chapter 6. Results

Results showed that CreativeConnect helped participants with both finding and recombining elements for reference recombination. Also, it was shown that users with CreativeConnect were able to generate more design ideas in a given time and perceived their ideas as more creative compared to the baseline. We also found some differences between CreativeConnect and baseline in terms of how users utilize the tool for their creative process.

## 6.1 Support for Different Recombination Steps

To answer RQ1, we examined survey questions and log analysis results divided into three steps of reference recombination: (1) discovering keywords from the reference, (2) exploring different recombinations, and (3) generating new concepts. For all survey questions, we used a Wilcoxon signed-rank test as all of them were ordinal data on a 7-point Likert scale. For the usage log analysis, we conducted a two-sample t-test or two-sample paired t-test to compare between CreativeConnect and baseline.

### 6.1.1 Finding Keywords from the Reference

Participants perceived that CreativeConnect helped discover valuable keywords from the given reference images. As shown in Figure 6.1, participants found out that CreativeConnect is significantly more useful for discovering valuable elements from the reference that can be used for their ideation (M=6.13, SD=1.31) compared to the baseline system (M=3.75, SD=1.98 / p=0.001, W=0.0). Regarding how CreativeConnect and baseline helped with organizing references, the rating was not significantly different, but with a slightly higher average rating for CreativeConnect (Baseline: M=4.38, SD=1.96 / CreativeConnect: M=5.31, SD=1.58 / p=0.121, W=23.5).

Usage logs also showed that CreativeConnect effectively encouraged participants to explore and extract different keywords. In comparing the numbers of the keyword notes that participants left in both conditions using a two-sampled paired t-test, participants with CreativeConnect added more keyword notes (M=34.69, SD=10.74) compared to the baseline system (M=13.19, SD=10.53 / p¡0.0001, t=5.52). Also, as shown in Figure 6.2, participants with baseline typically extracted keywords exclusively during their initial sketch, thereafter relying solely on the previously extracted keywords without actively discovering additional keywords. In contrast, participants using CreativeConnect consistently added more keywords throughout the whole process. While they also extracted the most keywords at the beginning, they continued to extract new keywords from references for every new sketch. One participant (P15) drew out all the sketches at the end after developing multiple design ideas at once, instead of drawing a sketch immediately after developing each design idea. As it cannot be clearly identified which keyword notes correspond to each specific sketch in P15's usage log, this data was excluded only from this analysis, which is about the actions related to each sketching turn. The raw usage log data for all participants, including P15, is provided in Appendix 10.2.3.
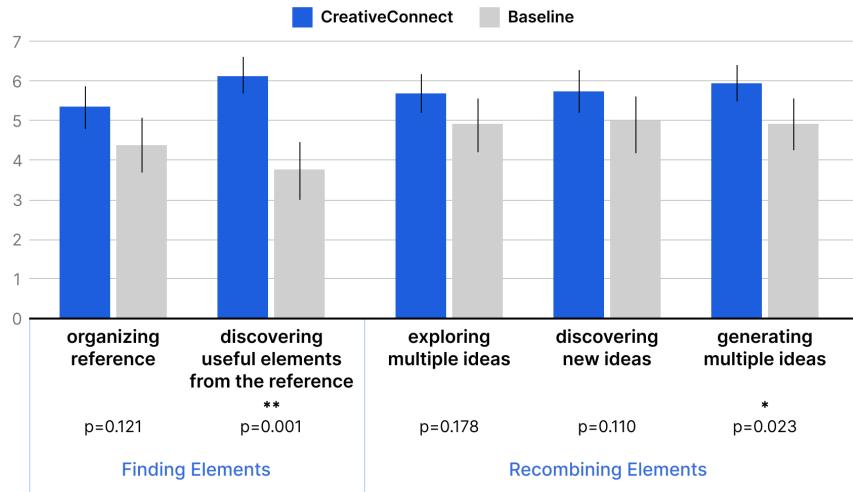
Figure 6.1: Survey result on the user-perceived efficiency during each step of the recombination step with 95% confidence interval. CreativeConnect is significantly useful in discovering elements from the reference image, and also in generating multiple ideas.
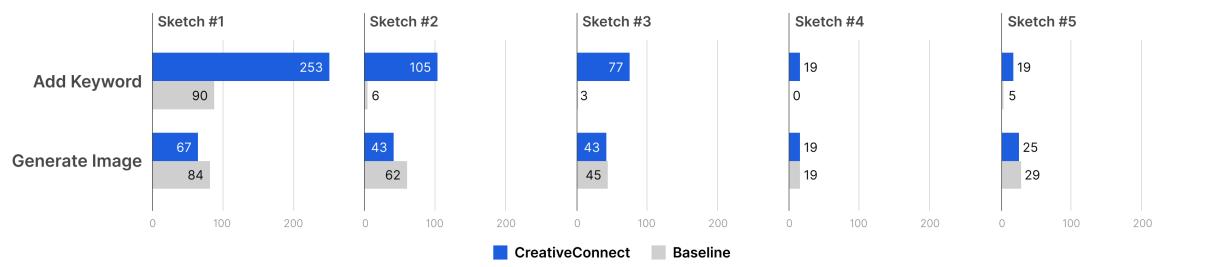


Figure 6.2: Comparison of the count of two different actions (adding keywords, generating image through generation model) taken to generate each sketch (from the first sketch to the fifth sketch) in CreativeConnect and baseline. The results show that users use the add keyword action more in CreativeConnect compared to the baseline where users only add keywords for the initial sketches. There is no significant difference in the count of generated image action.

### 6.1.2 Recombining elements

The survey's findings indicated that the CreativeConnect can be useful for recombining different elements into new design ideas. Participants said that CreativeConnect is significantly more helpful (M=5.94, SD=1.34) than the baseline system (M=4.88, SD=1.89 / p=0.023, W=10.5) for them to generate multiple ideas from the collected elements (Figure 6.1). However, participants' perception of how much the system helped explore multiple ideas was not significantly different in both conditions, although CreativeConnect had a slightly higher average rate (Baseline: M=4.88, SD=1.93 / CreativeConnect: M=5.69, SD=1.35 / p=0.178, W=22.0). Also, it was not significant in terms of discovering novel ideas, but the average was slightly higher in CreativeConnect (baseline: M=5.00, SD=1.75 / CreativeConnect: M=5.75, SD=1.48 / p=0.110, W=19.0).

We also looked into how participants used the given image generation model for recombining elements into a design idea. As shown in Table 6.1, there was also no significant difference between the two conditions in the number of generated images. However, in terms of exploring diverse recombinations using the model, CreativeConnect showed distinct advantages, as evident from the unique patterns observed when users interacted with the generation model under two conditions. Using the baseline system, users could provide separate inputs for overall image descriptions and for each object. CreativeConnect allowed users to select multiple keywords to merge. In both conditions, participants had the option to input multiple phrases together to combine them. We conducted an analysis to evaluate how diverse phrases inputted together into the model. Out of a total of 347 input sets (202 from the baseline, 145 from CreativeConnect), 14 sets (11 from the baseline, 3 from CreativeConnect) consisted of only one input, and they were excluded from the analysis since our objective was to compare the semantic similarity between phrases that were provided to the model together. For the remaining 333 input sets (191 from baseline, 142 from CreativeConnect), we computed the semantic similarity between all pairs of phrases within each input set and calculated the mean and minimum similarity. The mean similarity represents the overall similarities between phrases provided as input together, while the minimum similarity represents the most diverse pairs within the set. Finally, we conducted a two-sample t-test for each metric.

As shown in Table 6.1, the input sets created within CreativeConnect showed significantly lower similarity between the keywords (M=0.222, SD=0.094) compared to the sets made within the baseline (M=0.263, SD=0.166 / p=0.008, t=2.66) when they are calculated based on the minimum similarity. This difference is also similar when they are calculated based on the mean similarity, but it was slightly not significant (Baseline: M=0.356, SD=0.148 / CreativeConnect: M=0.330, SD=0.075 / p=0.051, t=1.95). This means that participants with CreativeConnect actively sought to create unique recombinations with greater semantic similarity, ultimately leading to the exploration of diverse and distinct recombinations compared to the baseline condition.

## 6.2 Ideation Results

To answer RQ2, we analyzed the design idea sketches that participants drew during the study session through expert evaluation, usage log, and survey results. Same as the RQ1, the Wilcoxon signed-rank test was used for survey questions. For expert evaluation results and the log analysis results, we used a two-sample t-test. For pairwise data, such as a comparison between the number of sketches drawn in each condition by each participant, we conducted a two-sample paired t-test.

|  |  | CreativeConnect | | Baseline | | Statistics | |
|---|---|---|---|---|---|---|---|
|  |  | mean | std | mean | std | p | Sig. |
| Image Generation Model Usage (Per session) | # of generated image | 57.06 | 17.91 | 46.69 | 23.52 | 0.119 | - |
| | # of user inputs to the model | 9.31 | 4.57 | 10.56 | 4.76 | 0.468 | - |
| Semantic Similarity within Input Sets | Semantic Similarity (Mean) | 0.330 | 0.075 | 0.356 | 0.148 | **0.051** | - |
| | Semantic Similarity (Min) | 0.222 | 0.094 | 0.263 | 0.166 | **0.008** | ** |

Table 6.1: Number of image generation model usage and the semantic similarity between user inputs in CreativeConnect and baseline. (-: $p > .05$, *: $p < .050$, **: $p < .010$, ***: $p < .001$)
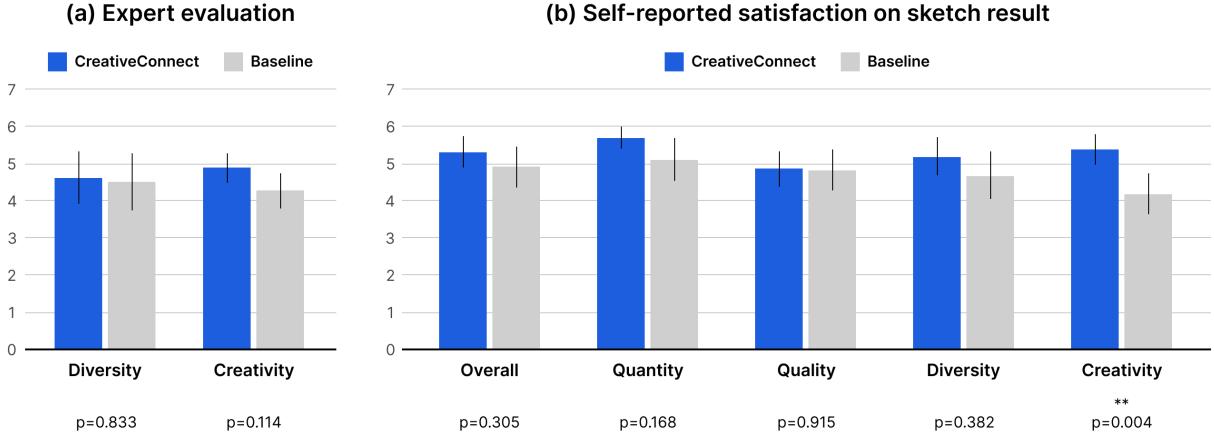


Figure 6.3: Evaluation results of user-drawn sketch with 95% confidence interval. (a) Expert evaluation on the diversity and creativity for CreativeConnect and baseline condition. (b) Self-reported satisfaction on sketch result in terms of quantity, quality, diversity, creativity, and overall for CreativeConnect and baseline condition.

### 6.2.1 Creativity & Diversity of the Final Sketches

As shown in Figure 6.3 (b), the survey results showed that participants perceived their sketch as more creative when they were using CreativeConnect (M=5.38, SD=1.09) compared to the baseline (M=4.19, SD=1.64 / p=0.004, W=0.0). During the interview, 12 out of 16 participants said that they felt they could be more creative with the support of CreativeConnect rather than the baseline, especially when they're having a hard time coming up with a new idea in the early ideation stage. In terms of other factors including overall satisfaction, quantity, quality, and diversity of the sketches, there were no significant statistical differences between the two conditions.

As shown in Figure 6.3 (a), however, the expert evaluation does not show a significant difference between the two conditions. The creativity score of the expert evaluation were slightly better in CreativeConnect (M=4.854, SD=1.418) compared to the baseline (M=4.344, SD=1.708 / p=0.114, t=1.59), but it was not significant according to the two-sample t-test results. In terms of diversity, there was also no significant difference between them (Baseline: M=4.625, SD=1.607 / CreativeConnect: M=4.75, SD=1.418 / p=0.833, t=0.23). There were possible reasons that expert evaluation was different from the survey results. First, even though the experts were asked to focus on the idea as much as possible, the

|  | CreativeConnect | | Baseline | | Statistics | |
| --- | --- | --- | --- | --- | --- | --- |
|  | mean | std | mean | std | p | Sig. |
| # of sketch per session | **5.56** | 1.63 | 5.06 | 1.73 | 0.041 | ∗ |
| time per sketch (min) | 5.01 | 2.87 | 5.39 | 3.03 | 0.403 | - |

Table 6.2: Number of sketches drawn by the participants per session and the average time taken for sketches (-: p > .05, ∗: p < .050, ∗∗: p < .010, ∗∗∗: p < .001)

participants' sketch skills were inevitably reflected in the evaluation, and some of the comments left by the evaluators were actually about the sketch skills. There is also a possibility that deviations according to the design topic may have been affected. In fact, sketches about the topic of underwater Christmas were rated higher on average.

### 6.2.2 Efficiency of the Ideation Process

As shown in Table 6.2, the two-sample pairwise t-test result showed that participants came up with more sketches in the same 30-minute ideation session with the support of CreativeConnect (M=5.56, SD=1.63) than with the baseline (M=5.06, SD=1.73 / p=0.041, t=2.24). This result indicates that CreativeConnect can be useful for efficient ideation. The interview results also demonstrated that CreativeConnect could be useful when they have to come up with a lot of ideas in a limited set of references and time, which is a common scenario in professional design tasks where clients provide references and designers must provide drafts with them.

### 6.2.3 Perceived Workload

As shown in Table 6.4, there was no difference between the two conditions in terms of the perceived workload. While CreativeConnect has additional complications such as requiring users to specify keywords to give inputs to the image generation model, this does not cause users to feel overwhelmed while performing the task.

## 6.3 Impact on User's Creative Process

### 6.3.1 Source of the Inspiration

To investigate how users use the output from CreativeConnect and baseline system for generating new design ideas differently, we asked participants to pick one sketch that they think is the most creative for each study session and explain how they got the inspiration for that.

There are five different inspirational sources found in two conditions. In both conditions, many participants got their ideas from the generated images or text descriptions. In CreativeConnect condition, more than half of the participants said that their best ideas are inspired by these generated images or text (Table 6.3). As illustrated in Figure 6.4 (a), participants utilized keywords from both reference images and recommendations and merged them using the system. What was notable is that they got the generated images and tried to reinterpret them in their own way, rather than accepting what is drawn there. Participants with the baseline system were also influenced by the images generated, but the number was slightly less (Table 6.3), and the way they were influenced was slightly different. They

|  |  |  | CreativeConnect | Baseline |
|---|---|---|:---:|:---:|
| Source of Inspiration | Within the tool | Generated image/description | 9 | 5 |
|  |  | Recommended keywords | 1 | - |
|  |  | ChatGPT answers | - | 2 |
|  | Outside of the tool | Own creativity | 1 | 1 |
|  |  | Reference images | 5 | 8 |
|  |  | Avg. tool assistance rating | **5.625** | 4.563 |

Table 6.3: Count of the inspiration sources for the most creative sketches chosen by the participants and the average rating of the efficiency of the tool assistance for drawing those sketches. Figure 6.4 illustrates the example use cases of inspiration within the tool.

tend to refer to the visual compositions or details of the shapes and apply them into their own sketch. P7 mentioned the reason for this, "While putting prompts into the image generation model (in the baseline), I already had the concept that I wanted. Therefore, I refer to the expression method of it, rather than trying to find something new out of it."
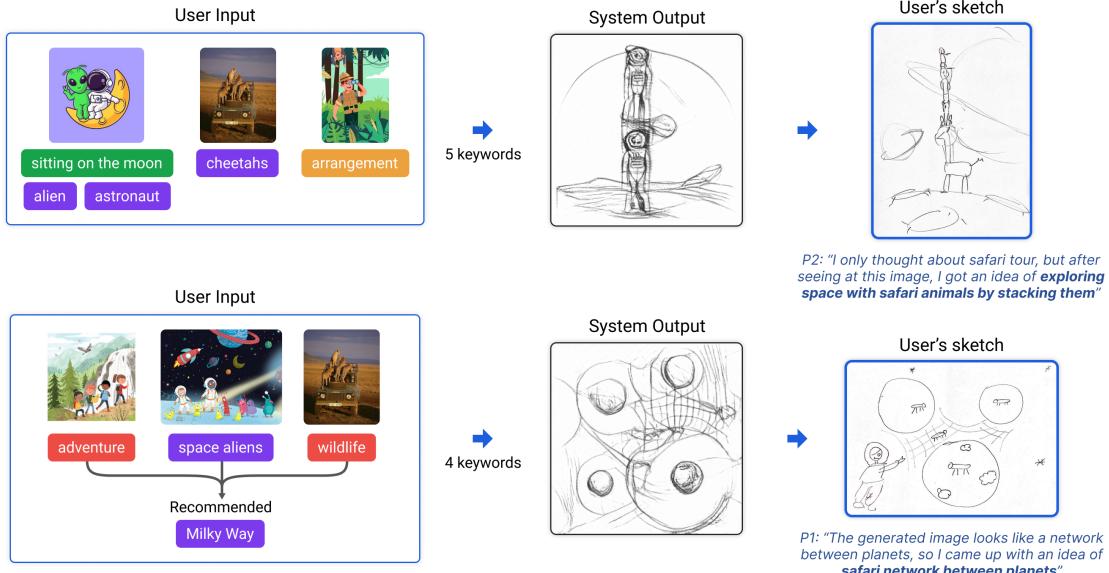
One thing noticeable is that participants were influenced more by the given reference images when they were using the baseline. This shows that CreativeConnect can potentially make users less directly influenced by reference images, ultimately preventing them from fixating on them. P16 explicitly pointed out this by saying, "When using CreativeConnect, I gave less focus to given images, and as I can expand to a lot of ideas only with a small number of references, I didn't even use all of them." P14 mentioned, "This (baseline) tool feels like a notepad that manages references, so I kept referring to the reference images themselves."

As shown in Figure 6.4 (d), there was also a participant who got an idea from CreativeConnect's recommended keywords. In baseline, instead of this keyword recommendation feature, they could use ChatGPT, and 2 participants said that they got their inspiration from this. However, the number of this usage was relatively small (Table 6.3), mainly because of the challenges of using it for visual tasks. Participants mentioned difficulties in both formulating prompts and leveraging the language-based output for their design during the interview.
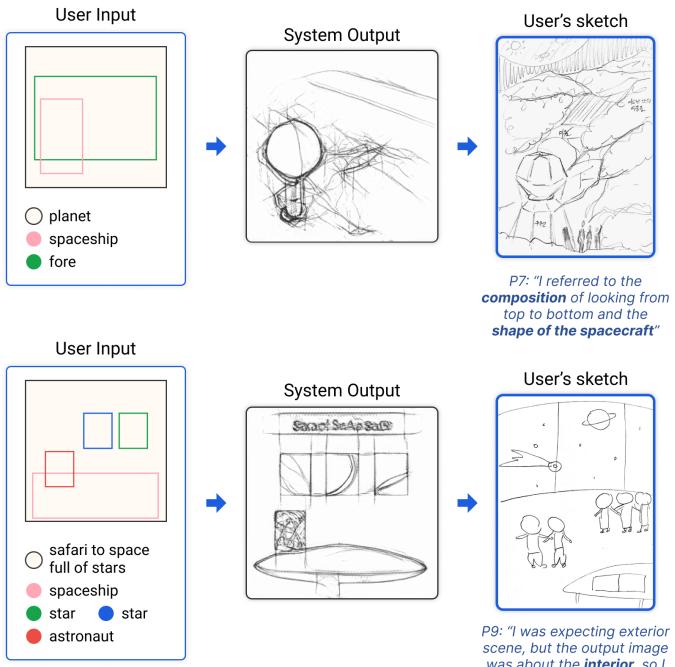
This difference in sources of inspiration affected the results of the user's rating of how effective the assistance of the tool was. We conducted a two-sample pairwise t-test to compare participants' ratings on the tool's usefulness for generating their favorite ideas, and it was shown to be higher in CreativeConnect (M=5.63, SD=1.41) compared to the baseline (M=4.56, SD=1.89 / p=0.045, t=2.18) (Table 6.3), indicating that users perceived the features of CreativeConnect are more useful for coming up with their best ideas, compared to the baseline system.

The survey results about the perceived experience of using the AI-based system also showed a more specific reason for this helpfulness. As shown in Table 6.4, CreativeConnect is shown to be significantly better for thinking through what kind of outputs users want to complete for the given task (baseline: M=5.00, SD=1.97 / CreativeConnect: M=6.13, SD=1.02 / p=0.045, W=14.5). This shows that participants don't think of the results of the CreativeConnect's image generation model as their final results, but more as a guide to think about what they want. It leads users to think in diverse ways. P9 mentioned that "In baseline, the result came out exactly what I thought, so I replicated the output. However, CreativeConnect shows me various high-level ways to combine things, so I could explore those methods
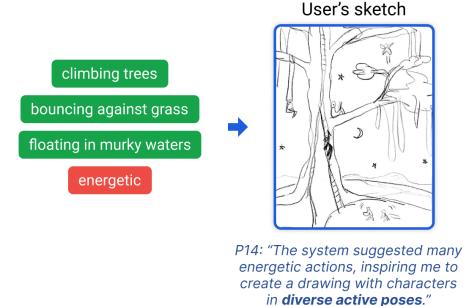
**(a) Inspired by generated images (CreativeConnect)**

User Input
sitting on the moon
cheetahs
arrangement
alien    astronaut

5 keywords

System Output

User's sketch

P2: "I only thought about safari tour, but after seeing at this image, I got an idea of **exploring space with safari animals by stacking them**"

User Input
adventure    space aliens    wildlife
Recommended
Milky Way

4 keywords

System Output

User's sketch

P1: "The generated image looks like a network between planets, so I came up with an idea of **safari network between planets**"

**(b) Inspired by generated images (baseline)**

User Input
○ planet
● spaceship
● fore

System Output

User's sketch

P7: "I referred to the **composition** of looking from top to bottom and the **shape of the spacecraft**"

User Input
○ safari to space full of stars
● spaceship
● star    ● star
● astronaut

System Output

User's sketch

P9: "I was expecting exterior scene, but the output image was about the **interior**, so I referred to that concept"

**(c) Inspired by recommended keywords (CreativeConnect)**

climbing trees
bouncing against grass
floating in murky waters
energetic

User's sketch

P14: "The system suggested many energetic actions, inspiring me to create a drawing with characters in **diverse active poses**."

**(d) Inspired by ChatGPT answers (baseline)**

Q: What should I draw if I want to make a children book cover that book title is 'christmas dinner in underwater world'?

A: (...) **Title and Author**: Place the title "Christmas Dinner in Underwater World" in a playful and colorful font at the top of the cover. (...)

User's sketch

P6: "ChatGPT's response made me realize the existance of title on the book cover, sparking ideas to **make something with this title**."

Figure 6.4: Examples of user input, system-generated output, user-drawn sketches inspired by the system output, and corresponding user quotes. In both conditions, users were inspired by the generated images. However, CreativeConnect users were more inspired by the overall concept of the image, while baseline users referred to a specific composition or the detail of the object from the image. Also, CreativeConnect users often got inspired by the recommended keywords, while baseline users had ChatGPT support instead of it.

|  |  | CreativeConnect | | Baseline | | Statistics | |
|---|---|---|---|---|---|---|---|
|  |  | mean | std | mean | std | p | Sig. |
| Self-perceived experience on ML model | Match goal | 5.00 | 1.63 | 4.63 | 1.96 | 0.5805 | - |
|  | Think through | 6.13 | 1.02 | 5.00 | 1.97 | 0.0454 | ∗ |
|  | Transparent | 4.81 | 1.80 | 4.38 | 1.67 | 0.4488 | - |
|  | Controllable | 4.75 | 1.95 | 4.06 | 1.84 | 0.2976 | - |
|  | Collaborative | 5.38 | 1.59 | 4.94 | 2.08 | 0.4809 | - |
| NASA-TLX | Mental | 3.69 | 1.82 | 4.19 | 1.94 | 0.39 | - |
|  | Physical | 1.81 | 1.22 | 2.50 | 2.10 | 0.10 | - |
|  | Temporal | 2.81 | 1.83 | 3.50 | 2.28 | 0.23 | - |
|  | Effort | 3.63 | 1.82 | 3.94 | 2.05 | 0.63 | - |
|  | Performance | 5.31 | 1.08 | 5.06 | 1.39 | 0.78 | - |
|  | Frustration | 2.63 | 1.93 | 3.50 | 1.75 | 0.14 | - |
| Creativity Support Index | Enjoyment | 5.91 | 1.00 | 5.09 | 1.78 | 0.077 | - |
|  | Exploration | 5.38 | 1.54 | 4.81 | 1.56 | 0.211 | - |
|  | Expressiveness | 5.44 | 1.18 | 4.53 | 1.75 | 0.032 | ∗ |
|  | Immersion | 4.69 | 1.99 | 4.69 | 1.82 | 1 | - |
|  | Results Worth Effort | 5.47 | 1.27 | 5.25 | 1.71 | 0.591 | - |
|  | Collaboration | 5.19 | 1.25 | 4.41 | 1.71 | 0.016 | ∗ |

Table 6.4: Survey results of self-perceived experience on ML features, NASA-TLX questionnaire, and Creativity Support Index. (-: p > .05, ∗: p < .050, ∗∗: p < .010, ∗∗∗: p < .001)

and expand those processes on my own."

### 6.3.2 Creativity Support Index

According to Table 6.4, users tend to prefer CreativeConnect significantly more than the baseline in terms of expressiveness and collaboration, but in the other criteria, there was no significant difference between the two systems. Through the post-interview, we found out that participants felt different types of creativity support in each system. Participants said that the baseline was helpful when they had an overall idea in their mind and wanted to get support for expressing it in the sketch. On the other hand, participants said that the CreativeConnect is helpful for their creativity when they have no idea in their minds. These differences will be explained in more detail in section 7.1.

# Chapter 7.  Discussions

We propose a novel AI-infused creativity support tool CreativeConnect, which assists graphic designers in generating their design ideas by recombining reference images. Based on our findings, we suggest some design implications for future creativity support tools.

## 7.1  CreativeConnect vs. baseline - Two Different Types of Creativity Support

The results show that CreativeConnect supports the early-stage conceptual ideation with reference recombination process successfully, by aligning well with the four design goals that we derived from the formative study. Participants could easily extract keywords (DG 1) and utilize keyword recommendations as a source of new inspirations (DG 2) which led them to make more keyword notes. Also, they explored diverse keyword recombinations (DG 3) which led them to make more design ideas in a given time. Additionally, they perceived their idea as more creative as CreativeConnect provided the output as an incomplete sketch and let participants inject their creativity into it (DG 4). However, participants couldn't feel the difference in the overall degree of creativity support between the two tools. The interviews revealed that this was because CreativeConnect and baseline both provided valid creativity support, but in a distinct way based on users' current needs.

In the baseline system, users should specify all the details of the generated image, so they appreciated the transparency and control. The system faithfully reproduced user input by that control, resulting in a final output that closely mirrors the concept in their mind. These generated outputs helped users actualize their existing ideas, more supporting *implementation* [18, 17]. Dynamic Brushes [36] or Framer [51] had a similar approach to creativity in terms of this.

Conversely, CreativeConnect stimulates creativity by providing *inspiration* [18, 17]. Instead of requiring users to provide detailed input, CreativeConnect accepts keywords and deliberately refrains from exact expression, generating a wider range of outcomes, potentially with serendipity. How CreativeConnect can provide participants with this creative leap can be explained by Cross' descriptive model of creative design [19]. The keyword extraction feature actively supports *emergence*, allowing designers to find unrecognized properties of the existing design. The keyword recommendation also supports *mutation*, helping designers to generate new ideas by modifying existing designs partially. P15 metaphorically likened this process to having someone nearby constantly talking with them with fresh variations in the form of keywords. Furthermore, the keyword merging feature enhances *combination*, where new ideas are generated by combining features from existing designs. Therefore, CreativeConnect could be potentially useful for addressing a common challenge known as "artist's block" or "creative block", similar to the "writer's block" experienced by writers [31]. CreativeConnect could provide proper support when designers find themselves creatively stuck, breaking creative inertia by sparking novel ideas and opening new creative avenues.

These differences can be valuable design implications for future creativity support tools as designers require different types of creativity support in different stages of the ideation process. By dynamically adjusting the type of support based on the user's context, such a tool can offer a more personalized and

effective creative experience. For instance, when the system detects that a user is in the exploration phase, it can employ an approach similar to CreativeConnect, encouraging the generation of diverse and abstract ideas. Conversely, when the user demonstrates a desire to refine and develop a particular concept, the tool can provide baseline-like features to ensure greater control and fidelity in the generated output. This adaptable approach acknowledges the multifaceted nature of the creative process and supports users with the right tools at the right moment, ultimately enhancing their creativity. Also, by integrating those inspiration and implementation support into a single tool, it can enable a seamless transition between generating diverse ideas and refining specific concepts, fostering a more iterative and efficient creative workflow.

## 7.2   The Role of Low-fidelity Output for Creativity Support

The post-interview showed that adopting low-fidelity output can facilitate further imagination beyond what the system provided. In both CreativeConnect and the baseline, we deliberately employ a low-fidelity sketch output. During the interview, 12 out of 16 participants said that they preferred the sketch output rather than a complete image, as it allowed users room for imagination and interpretation. The image converted into a sketch omits small details and retains only the larger forms, which in turn generates large empty space. This emptiness encourages users not just to perceive the generated image but to see it as room for further development and makes users deeply engaged in further ideation. Some participants even expressed opposition to completed images for the ideation stage, as they believed that an abundance of details in reference images makes them fixated on that specific design idea and hinders them from utilizing the images in their own ideas. P2 said, "I usually get completed artworks from Pinterest [1] as a reference, and I found myself unavoidably looking at the unique style of that designer, wanting to replicate it. This time, I liked that I could maintain my own style while exploring different references of concepts." From our findings, we argue that adopting low-fidelity output should be an option to consider when designing creativity support systems to prevent fixation and facilitate the user's creativity in ideation. For example, a design reference tool can dynamically adjust the levels of details of the provided images based on the user's current design stage. When the user wants references for overall concepts, the system can convert reference images to a simple black line drawing or even present it solely as a textual description. Conversely, when the user has determined a specific concept and is exploring different visual details, the system can offer the original images with full details.

## 7.3   Generalizability of CreativeConnect in Different Context

CreativeConnect is designed to support novice designers, such as design students with a general understanding of the design process but struggle with reference recombination. However, our user study revealed some insights applicable to different expertise levels. We observed that participants with limited sketching skills were satisfied more with the baseline system, as it was more aligned with their intentions and suitable for the aid for the actual sketching. Therefore, for users less familiar with artistic expression, an AI tool's output should prioritize alignment with the users' original intent, rather than abstraction. Conversely, for experts accustomed to extracting inspiration from references and combining them into their original idea [4], CreativeConnect could serve as a tool for serendipity rather than helping them with the process of keyword extraction and recombination. For example, P16 said that suggested keywords

---

[1] https://www.pinterest.com/

and merged images acted as prompts to remind them of some aspects initially overlooked. Therefore, features should be redesigned to encourage reflection and creative exploration, such as highlighting the part of the generated images that were not present in existing references but emerged through our system features.

The user study results showed that CreativeConnect could be also applied for other design contexts such as collaborative projects. According to the CSI survey results (Section 6.3.2), participants indicated that CreativeConnect would be significantly helpful for collaborating with other designers. This was because CreativeConnect is designed to follow the sequential steps of leaving keyword notes and merging them, and it keeps track of these processes on the mood board and the merging panel. Therefore, participants said that simply showing CreativeConnect screen could be used to share their creative processes with other designers, making it easier for them to understand each other's thought processes and quickly reach an agreement on the design direction. One future work direction can be incorporating features of CreativeConnect to collaborative mood board tools [15, 49, 48] and studying the benefits of keyword-based recombination features.

CreativeConnect could also be applied to other design domains. Our design goals and the feature design of CreativeConnect are primarily tailored to the illustration design task. Illustrations predominantly about conveying design topics through visual subject matters and do not usually include other modalities such as text (common in poster or publication design) or motion & interaction (common in UI/UX and motion graphic design). However, even in other design domains, the recombination process of extracting elements from the reference and recombining them would be an effective strategy. To apply the recombination approach to another design domain, we have to first identify what elements designers in that domain focus on when looking at references and use those different categories of elements as keywords in the pipeline of CreativeConnect.

# Chapter 8. Limitations and Future Work

Our work has several limitations which can be potentially addressed in future work.

- In our user study, the ideation tasks were conducted for 30 minutes in each condition, which was shorter than the actual design process. Therefore, it was difficult to observe how the behavior changed over a long time. Future work can be done to incorporate CreativeConnect with real-world design projects and see how their behavior patterns differ from lab studies.

- Our pipeline generates an image description containing all of the keywords selected by the user as a method of recombination. However, there can be various ways of recombination other than this, such as blending objects or indirectly expressing some keywords through visual details such as colors. Further work can be done on these various recombination methods and how to support them.

- As CreativeConnect and baseline both leverage generative AI including LLM and layout diffusion model, the result may be influenced based on users' familiarity with AI. Since this study did not explore those dimensions, future research can be conducted to find how creativity supporting tools with AI features may have varying effects depending on the user's knowledge level of AI or prior experiences of using AI.

# Chapter 9.  Conclusion

This paper proposed CreativeConnect, a system designed to support graphic designers in the reference recombination process, allowing them to generate novel design ideas. Building on our formative study observations, CreativeConnect offers users assistance in identifying key elements within reference images. It also provides a diverse range of recommendations for relevant keywords and recombination options. Notably, the low-fidelity sketch-based output of CreativeConnect was shown to encourage creativity by enabling further imaginative exploration. Our user study demonstrated that CreativeConnect efficiently supported both steps of finding and recombining elements and helped participants to come up with more design ideas and perceive their ideas as more creative compared to the baseline. While CreativeConnect represents a promising step towards comprehensive recombination support tools for designers, we also suggested an opportunity to expand such systems to cater to address a wider spectrum of design needs and situations.

# Chapter 10.  Appendix

## 10.1   Technical Details

### 10.1.1   Prompt: Extracting Keywords from Image Captions

**System Prompt:**

You will be provided with multiple sentences to describe an illustration.  Your task is to extract a list of Subject matter, Action & pose, and Theme & mood.

Subject matters are one-word, describing the specific physical objects, characters, or landscape that the user wants to include in their illustration.  Example subject matters include pencil, children, or wave.  For subject matters, no adjectives should be included.  They should always be a noun.

Actions & poses are word-level or phrase-level actions that the character or the object in the illustration performs. Example actions & poses include riding a bus, standing still, or traveling.

Themes & moods are words not directly present in the illustration, but those that can potentially convey the overall theme or mood of the illustration.  Example themes & moods include imaginative, eco-friendly, or sad.  They should be adverbs, preferably one or two words.

If you are provided a sentences including some style such as cartoon, illustration, image, photo, exclude it.  For other example, 'an illustration of a woman sitting at a table' caption is extracted to 'woman', 'table', 'sitting at a table', 'cozy'.  The 'illustration' is not contained.  Eliminate the changed forms of the same word, such as plurals.  Only include roots.  For example of 'trees' and 'tree', only include 'tree'."

**Examples:**
**User:**
a card with chinese writing with colorful objects on it
a red and orange background with a blank paper, chinese, pencils, stationery and more
an image of a classroom scene with various supplies
**Assistant:**
Subject matter: card, Chinese writing, colorful objects, red and orange background, blank paper, Chinese, pencils, stationery, classroom, supplies.
Action & pose:
Theme & mood: education, learning, multiculturalism
**User:**
a man and a woman sitting at a table eating food
a woman sitting in a chair in a room with potted plants
a man sitting at a table eating a hot dog
a illustration of a woman sitting at a table in a kitchen with potted plants
an illustration of a woman sitting at a table
an illustration of a refrigerator next to a chair and a table
an illustration of two people sitting at a table with their feet on a chair
a cartoon of a person sitting in a chair in a shower
**Assistant:**
Subject matter: man, woman, table, food, chair, potted plants, hot dog, kitchen, refrigerator, feet
Action & pose: sitting at a table, eating food
Theme & mood: cozy, heartwarming
**User:**
a painting of a white barn in a field of flowers
a painting of a blue house with a tree next to it
a painting of a blue house with a roof and trees
a painting of a window of a house with flowers

a painting of a field of flowers in front of a house

**Assistant:**

Subject matter: painting, white barn, field, flowers, blue house, tree, roof, window

Action & pose:

Theme & mood: rural, peaceful, nature

**User:**

the album cover of the beatles abbey road

a man and a woman standing in front of a car

a man and a woman walking down a street

a group of people walking across a crosswalk

**Assistant:**

Subject matter: Beatles, Abbey Road, man, woman, car, street, group, people, crosswalk

Action & pose: standing in front of a car, walking down a street, walking across a crosswalk

Theme & mood: urban, nostalgia

## 10.1.2 Prompt: Recommending Relevant Keywords

**System Prompt:**

We are trying to support novice designers' ideation process by semantically combining different parts of illustration references. You will be provided with the topic of the ideation, and multiple keywords users like in the illustrations they found as references. There are three types of keywords: Subject matter, Action & Pose, and Theme & Mood.

Subject matters are one-word, describing the specific physical objects, characters, or landscape that the user wants to include in their illustration. Example subject matters include pencil, children, or wave. For subject matters, no adjectives should be included. They should always be a noun. Come up with more than four new keywords for Subject matter.

Actions & poses are word-level or phrase-level actions that the character or the object in the illustration performs. Example actions & poses include riding a bus, standing still, or traveling.

Themes & moods are words not directly present in the illustration, but those that can potentially convey the overall theme or mood of the illustration. Example themes & moods include imaginative, eco-friendly, or sad. They should be adverbs, preferably one word.

Your task is to expand on the keywords being given, by combining multiple keywords or looking for synonyms that can inspire new creations or ideas. For example, the subject matter "pencil" can be combined with the action & pose "traveling" to inspire a new action & pose "writing a diary". You can combine as many keywords at once. Another example is to generate "hair salon" from "hair dryer", "comb", and "scissors". For combinations that result in theme & mood, make them as abstract as possible. An example is to make "adventurous", "gusty" from "riding on ship" and "tent".

Come up with new keywords for each keyword type with creative combinations. Only use the original keywords provided for creating new keywords. Do not just paraphrase original keywords. Do not suggest similar keywords to the original ones.

Important: Include at least one subject matter for each combination. Subject matter and theme & mood should be a SINGLE WORD. Combinations among subject matters are highly recommended. New keywords should be śurprisingćompared to original ones. It means the character of your suggested word should have low similarity.'

**Examples:**

**User:**

Subject matter: camping, tent, tree, animals, Eiffel tower, family

Action & pose: riding on a bus, riding on a ship

Theme & mood: playful, imaginative

**Assistant:**

Subject matter: bear, sleeping person, safari, cruise

Action & pose: traveling, setting up camp, dancing jazz

Theme & mood: adventurous, serene, joyful, romantic

**User:**

Subject matter: boy, dinosaur, flower

Action & pose: watching television

Theme & mood: fantasy, playful

**Assistant:**

Subject matter: wind mill, volcano, movie screen

Action & pose: exploding strongly, riding a dinosaur, flying away to the sky

Theme & mood: vast, whimsical, rustic, frenetic

**User:**

Subject matter: dreamy scene, boy

Action & pose: playing with dino toys

**Assistant:**

Subject matter: universe, Saturn, astronauts

Action & pose: imagining adventures, floating on the space, role-playing, daydreaming

Theme & mood: jolly, imaginative, impactful

**User:**

Subject matter: Christmas tree

Action & pose: dancing around the Christmas tree

Theme & mood: family-bonding

**Assistant:**

Subject matter: Fireplace, wooden sled, Snowman, jazz, piano

Action & pose: melting, giving present, body-warming

Theme & mood: jubilant, sparkling, heartwarming

**User:**

Subject matter: sea turtles, Christmas tree, marine life

Action & pose: swimming, dancing around the Christmas tree

Theme & mood: fantasy, underwater, family-bonding

**Assistant:**

Subject matter: Sea horse, Christmas lights, coral, mermaid

Action & pose: floating on the wave, blinking eye, singing under the sea

Theme & mood: ethereal, aquatic, charming, panoramic

**User:**

Subject matter: kid, cat

Action & pose: laying on top of a suitcase, playing hide and seek

Theme & mood: Rustic, vivid, exhilarating

**Assistant:**

Subject matter: Birdcage, attic, trunk, blue bird

Action & pose: jumping on boxes, chasing birds, hiding in a suitcase

Theme & mood: quaint, mischievous, lively, nostalgic

## 10.1.3 Prompt: Generating Recombinations in Text Descriptions

**System Prompt:**

The user wants to draw an illustration, with the assistance of you. You will be provided with multiple keywords users wanted to include in their illustrations. There are three types of keywords: Subject matter, Action & pose, and Theme & mood.

Subject matters are one-word, describing the specific physical objects, characters, or landscape that the user wants to include in their illustration. Example subject matters include pencil, children, or wave. For subject matters, no adjectives should be included. They should always be a noun.

Actions & poses are word-level or phrase-level actions that the character or the object in the illustration performs. Example actions & poses include riding a bus, standing still, or traveling.

Themes & moods are words not directly present in the illustration, but those that can potentially convey the overall theme or mood of the illustration. Example themes & moods include imaginative, eco-friendly, or sad. They should be adverbs, preferably one word.

Your task is to generate three descriptions of the illustration that the user can draw based on the given keywords. The three descriptions should be significantly different from each other. Each description should include three things: "Caption" and "Objects".

"Caption" is a simple description of the overall image of the description. This should include some objects in the

"Objects" list. Keep it concise. Do not make it long and do not include unnecessary adjectives.

"Objects" is a list of the objects depicted in the illustrations, and a short description of them. The objects should be one of the given "Subject matters", or something related to them. You don't have to include all given subject matters. The given action and concept should also be considered for generating the object's detail.

**Examples:**

**User:**

Subject matter: ball, cat, dog

Action & pose: jumping

Theme & mood: playful, peaceful

**Assistant:**

1.

Scene: In the backyard, cats and dogs play happily with a ball which is their favorite toy.

Objects: [(ball, a brightly colored ball), (dog, a dog jumping with its tongue sticking out), (cat, an agile cat in mid-jump)]

2.

Scene: A beach where a dog is united in its pursuit of a bouncing ball.

Objects: [(ball, a sandy beach ball), (dog, a spotted dog leaping for the ball)]

3.

Scene: In a residential living room, the cat and dog both defy gravity in their jumps to catch the toy.

Objects: [(toy, a small rubber toy with vibrant stripes), (dog, a bouncy terrier flying mid-air), (cat, a Siamese cat also in mid-jump)]

**User:**

Subject matter: dinosaur, drink, hat, chair

Action & pose: relaxing

Theme & mood: sunny, relaxed

**Assistant:**

1.

Scene: A dinosaur wearing a hat, lounging with a drink on a chair under the sun.

Objects: [(dinosaur, a large but friendly looking dinosaur), (drink, a fruity cocktail with a tiny umbrella), (hat, a wide-brimmed straw hat), (chair, a comfortable looking lounge chair)]

2.

Scene: Dinosaur relaxes in a sophisticated environment with a drink and plops a hat on the chair next to it.

Objects: [(dinosaur, a dinosaur in a suit), (drink, a fancy drink in a crystal glass), (hat, a stylish trilby), (chair, a plush velvet chair)]

3.

Scene: A kid with a hat is playing on a picnic mat with a toy dinosaur.

Objects: [(kid, a kid with pink shirts), (dinosaur, a small green toy dinosaur)]

**User:**

Subject matter: dog, teeth

Action & pose:

Theme & mood: care

**Assistant:**

1.

Caption: A caring veterinarian examining a dog's teeth.

Objects: [(dog, a dog sitting calmly), (teeth, a pair of dental tools), (vet, a veterinarian wearing a lab coat)]

2.

Caption: A child diligently brushing their dog's teeth.

Objects: [(dog, a dog lying on its back), (teeth, a toothbrush and dog toothpaste), (child, a child brushing the dog's teeth)]

3.

Caption: A dental hygienist showing a dog owner the correct brushing technique for their pet's teeth.

Objects: [(dog, a dog standing on a dental examination table), (teeth, a dental mirror and toothbrush), (hygienist, a dental hygienist demonstrating the brushing technique)]

### 10.1.4 Prompt: Matching Layout with Objects

**System Prompt:**

You are an intelligent bounding box matcher. I will provide you with a caption that describes an illustration, a list of the objects that are included in the illustration, and a list of bounding boxes. Your task is to match bounding boxes to each object to make the illustration most balanced and realistic.

Each bounding box is in the format of (object name, [top-left x coordinate, top-left y coordinate, box width, box height]). The bounding boxes are represented as a proportion. The top-left corner has coordinate [0, 0]. The bottom-right corner has coordinnate [1, 1]. The bounding boxes should not go beyond the image boundaries.

**Examples:**

**User:**

A realistic image of landscape scene depicting a green car parking on the left of a blue truck, with a red air balloon and a bird in the sky

[air balloon, car, bird, truck]

[0.041, 0.783, 0.442, 0.179], [0.525, 0.699, 0.408, 0.263], [0.261, 0.458, 0.222, 0.264], [0.525, 0.458, 0.279, 0.195]

**Assistant:**

[('car', [0.041, 0.783, 0.442, 0.179]), ('truck', [0.525, 0.699, 0.408, 0.263]), ('balloon', [0.261, 0.458, 0.222, 0.264]), ('bird', [0.525, 0.458, 0.279, 0.195])]

**User:**

A realistic top-down view of a wooden table with two apples on it

[apple, apple, wooden table]

[0.219, 0, 0.562, 1], [0.402, 0.138, 0.195, 0.195], [0.402, 0.667, 0.195, 0.195]

**Assistant:**

[('wooden table', [0.219, 0, 0.562, 1]), ('apple', [0.402, 0.138, 0.195, 0.195]), ('apple', [0.402, 0.667, 0.195, 0.195])]

**User:**

A realistic scene of three skiers standing in a line on the snow near a palm tree

[skier, skier, skier, palm tree]

[0.487, 0.131, 0.142, 0.441], [0.661, 0.131, 0.143, 0.441], [0.836, 0.131, 0.142, 0.441], [0.795, 0.613, 0.183, 0.387]

**Assistant:**

[('skier', [0.487, 0.131, 0.142, 0.441]), ('skier', [0.661, 0.131, 0.143, 0.441]), ('skier', [0.836, 0.131, 0.142, 0.441]), ('palm tree', [0.795, 0.613, 0.183, 0.387])]

**User:**

An oil painting of a pink dolphin jumping on the left of a steam boat on the sea

[dolphin, steam boat]

[0.273, 0, 0.245, 1], [0.032, 0.455, 0.135, 0.420]

**Assistant:**

[('steam boat', [0.273, 0, 0.245, 1]), ('dolphin', [0.032, 0.455, 0.135, 0.420])]

**User:**

Immersed in his imagination, a boy is indoors enacting a prehistoric tale using four toy dinosaurs.

[dino toys, dino toys, dino toys, boy, dino toys]

[0.250, 0.218, 0.566, 0.563], [0.074, 0.556, 0.137, 0.284], [0.074, 0.76, 0.137, 0.284], [0.659, 0.041, 0.254, 0.134], [0.464, 0.840, 0.195, 0.120]

**Assistant:**

[('boy', [0.250, 0.218, 0.566, 0.563]), ('dino toys', [0.074, 0.556, 0.137, 0.284]), ('dino toys', [0.074, 0.76, 0.137, 0.284]), ('dino toys', [0.659, 0.041, 0.254, 0.134]), ('dino toys', [0.464, 0.840, 0.195, 0.120])]

**User:**

Two pandas in a forest without flowers

[panda, panda]

[0.114, 0.399, 0.183, 0.441], [0.733, 0.106, 0.150, 0.441]

**Assistant:**

[('panda', [0.114, 0.399, 0.183, 0.441]), ('panda', [0.733, 0.106, 0.150, 0.441])]

### 10.1.5 Prompt: Generating Layout based on Image Caption

**System Prompt:**
You are an intelligent bounding box generator. I will provide you with a caption for an illustration and a list of the objects. Your task is to generate the bounding boxes for the objects based on the caption. The images are of size 512x512. The top-left corner has coordinate [0, 0]. The bottom-right corner has coordinnate [512, 512]. The bounding boxes should not go beyond the image boundaries. Each bounding box should be in the format of (object name, [top-left x coordinate, top-left y coordinate, box width, box height]). If an object must appear several times by the provided caption, multiple bounding boxes may be added for the object.

**Examples:**
**User:**
A realistic image of landscape scene depicting a green car parking on the left of a blue truck, with a red air balloon and a bird in the sky
[car, truck, air balloon, bird]
**Assistant:**
[('car', [0.041, 0.549, 0.412, 0.31]), ('truck', [0.525, 0.553, 0.408, 0.313]), ('balloon', [0.129, 0.016, 0.283, 0.264]), ('bird', [0.578, 0.082, 0.279, 0.195])]
**User:**
A realistic top-down view of a wooden table with two apples on it
[wooden table, apple]
**Assistant:**
[('wooden table', [0.039, 0.289, 0.922, 0.422]), ('apple', [0.293, 0.441, 0.195, 0.195]), ('apple', [0.547, 0.441, 0.195, 0.195])]
**User:**
A realistic scene of three skiers standing in a line on the snow near a palm tree
[skier, palm tree]
**Assistant:**
[('skier', [0.01, 0.297, 0.271, 0.328]), ('skier', [0.543, 0.375, 0.236, 0.308]), ('skier', [0.289, 0.338, 0.242, 0.303]), ('palm tree', [0.789, 0.205, 0.201, 0.49])]
**User:**
An oil painting of a pink dolphin jumping on the left of a steam boat on the sea
[steam boat, dolphin]
**Assistant:**
[('steam boat', [0.453, 0.439, 0.502, 0.291]), ('dolphin', [0.041, 0.486, 0.369, 0.24])]
**User:**
A cute cat and an angry dog without birds
[cat, dog]
**Assistant:**
[('cat', [0.1, 0.131, 0.529, 0.632]), ('dog', [0.589, 0.232, 0.412, 0.445])]
**User:**
Two pandas in a forest without flowers
[panda]
**Assistant:**
[('panda', [0.059, 0.335, 0.414, 0.441]), ('panda', [0.516, 0.338, 0.434, 0.432])]
**User:**
Immersed in his imagination, a boy is indoors enacting a prehistoric tale using two toy dinosaurs.
[boy, dino toys]
**Assistant:**
[('boy', [0.367, 0.076, 0.301, 0.486]), ('dino toys', [0.15, 0.469, 0.201, 0.254]), ('dino toys', [0.543, 0.475, 0.188, 0.248])]

## 10.1.6   Example Outputs from the Technical Pipeline

Figure 10.1 shows some examples of inputs and outputs for three technical pipelines of CreativeConnect — (1) element extraction pipeline, (2) keyword recommendation pipeline, and (3) recombination generation pipeline. Figure 10.2 shows some more examples from the keyword extraction pipeline.
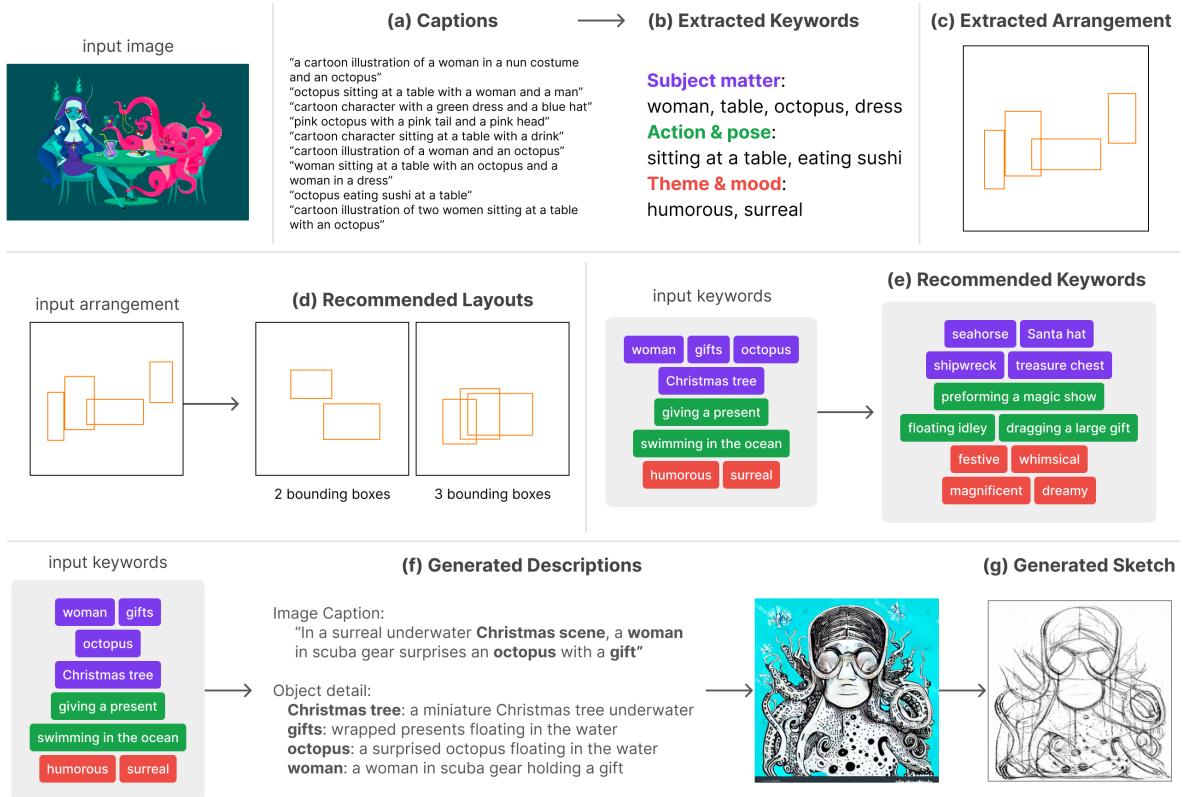
Figure 10.1: Examples of inputs and outputs given to CreativeConnect's pipelines. From the input image, (a) captions describing a given image and (c) arrangement of an image are acquired. Using captions as input, (b) keywords that are found in this image are extracted and categorized in subject matter, action & pose, and theme & mood. From the (c) extracted arrangement, a layout variator generates two different recommended layouts similar to the original one, containing 2 and 3 bounding boxes respectively. The keyword recommendation pipeline is used for (e) recommending keywords that are relevant to the input keywords. Based on the input keywords, (f) descriptions are generated with image captions and the details of objects. The final (g) generated image is created from the descriptions and converted into a sketch style.



Figure 10.2: Examples of the reference image with corresponding human-labeled ground truth labels and predicted keywords from the keyword extraction pipeline of CreativeConnect. Keywords predicted by our system can often be more descriptive and innovative than the ground-truths, which are highlighted in bold.
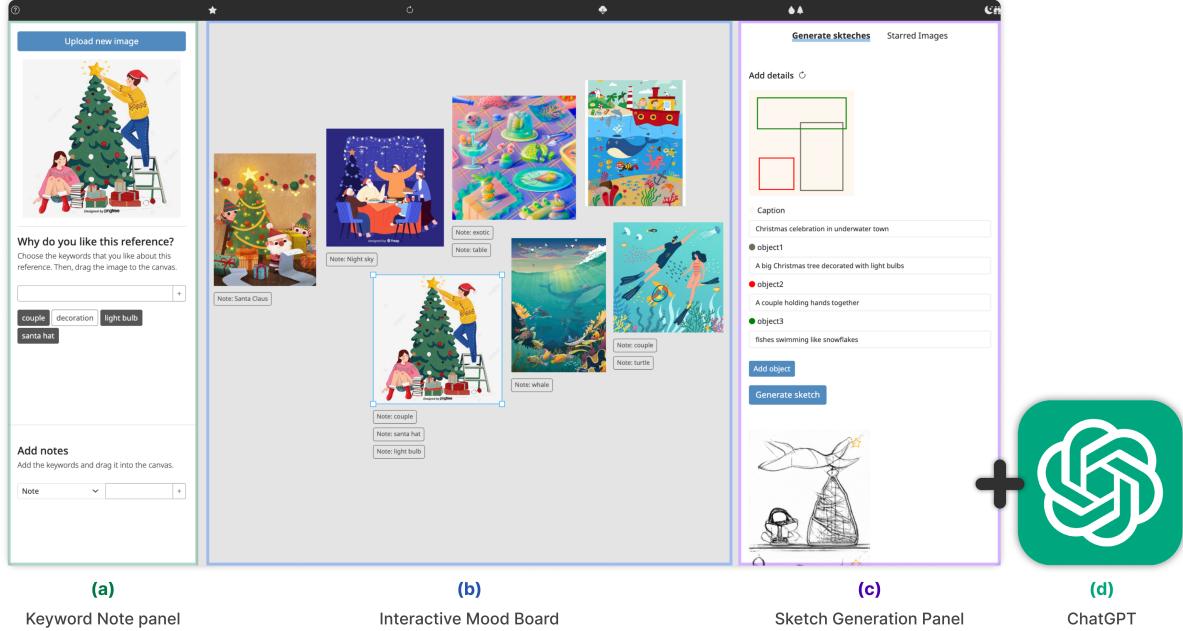
Figure 10.3: Screenshot of baseline system. (a) **Keyword Note Panel**: Users can add keywords manually on each image. (b) **Interactive Mood Board**: Users can organize the reference images on the mood board, along with the added keyword notes. (c) **Sketch Generation Panel**: Users can configure the overall layout of the generated image by manipulating boxes in the layout controller (beige) on the top of the panel. Additionally, users can provide prompts for the entire image and specific parts. They can specify more image details by clicking the "Add object" button. To get the generated sketches, users can click the "Generate sketch" button. (d) **ChatGPT**: Users were also provided with ChatGPT on a separate screen.

## 10.2 User Study

### 10.2.1 Baseline System Interface

Figure 10.3 shows the interface of the baseline system used for the user study. The baseline system looks similar to the CreativeConnect. There is no keyword extraction feature in the left panel, but it allows participants to manually add keyword notes. In the center, there is the same interactive mood board with the CreativeConnect, but there is no keyword suggestion panel. The right panel enables users to manually configure the layout and prompts for image generation, instead of selecting the keywords to combine. Other features such as mood board interactions (zoom, add/delete images) and saving favorite sketches were provided the same as the CreativeConnect. Participants could employ ChatGPT for various purposes other than this interface.

### 10.2.2 Interview questions

These are the questions used for the semi-structured interview after the two idea generation sessions with baseline and CreativeConnect tools.

1. Can you share the idea sketch that you think most creative in each topic, and what was the main source of inspiration for those ideas?

2. Comparing the baseline and CreativeConnect, what were the main differences you noticed in for the idea generation process?

3. In each three main stages of idea generation—finding reference elements, exploring ideas, and generating sketches—did you find one tool to be more helpful than the other, and why?

4. Were there any differences in your typical approach to idea generation when using these tools? If so, how was it different from the typical work process?

5. Which functionalities did you find most beneficial in both tools and in what scenarios were they particularly useful?

6. Were there any situations or specific sketches where the tools were especially useful or not useful?

7. In terms of image generation methods, what were the main differences between baseline and CreativeConnect, and when did you feel each method was more helpful?

8. How did you feel about the output in sketch format, and do you think the tool's effectiveness would differ if outputs were presented as a completed image rather than a sketch?

9. How did you incorporate the generated images into your final idea sketch?

### 10.2.3   Additional User Study Results: Raw Usage Log

Figure 10.4 shows the full usage log for all 16 participants of the user study, showing the timestamps of 3 types of user actions (adding keyword notes, generating images, and completing a design idea sketch).

Figure 10.4: Usage log for all participants in both CreativeConnect (Top) and baseline (Bottom) condition. The red triangle indicates the timestamp when the participants complete each sketch. The pink dot is the point when the participant added new keywords for the reference image. The sky blue dot is the point when the participant gave input into the image generation model. As shown in the figure, P15 first conducted multiple keyword-adding and image-generation actions, came up with all the design ideas, and then sketched all of them later in the session all at once. During the interview, P15 explained that they intended to focus exclusively on the sketching process, so they decided to jot down half-baked design ideas as memos in a text and draw them collectively. Unfortunately, the collected usage log only records the point when the overall sketch is completed and does not capture the individual instances of writing each memo. Therefore, we were unable to analyze which actions affected each design idea, so the usage data of P15 was excluded from the analysis of the relationship between action types and each sketching turn (Figure 6.2).

# Bibliography

[1] S. Bird, E. Klein, and E. Loper. *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.", 2009.

[2] M. A. Boden. Creativity and artificial intelligence. *Artificial Intelligence*, 103(1):347–356, 1998. Artificial Intelligence 40 years later.

[3] N. Bonnardel. Creativity in design activities: The role of analogies in a constrained cognitive environment. In *Proceedings of the 3rd conference on Creativity & cognition*, pages 158–165, 1999.

[4] N. Bonnardel and E. Marmèche. Towards supporting evocation processes in creative design: A cognitive approach. *International Journal of Human-Computer Studies*, 63(4):422–435, 2005. Computer support for creativity.

[5] S. Brade, B. Wang, M. Sousa, S. Oore, and T. Grossman. Promptify: Text-to-image generation through interactive prompt exploration with large language models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, UIST '23, New York, NY, USA, 2023. Association for Computing Machinery.

[6] T. Brooks, A. Holynski, and A. A. Efros. Instructpix2pix: Learning to follow image editing instructions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18392–18402, 2023.

[7] D. T. Campbell. Blind variation and selective retentions in creative thought as in other knowledge processes. *Psychological review*, 67(6):380, 1960.

[8] T. Cassidy. The mood board process modeled and understood as a qualitative design research tool. *Fashion Practice*, 3(2):225–251, 2011.

[9] J. Chan, S. Dang, and S. P. Dow. Comparing different sensemaking approaches for large-scale ideation. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, page 2717–2728, New York, NY, USA, 2016. Association for Computing Machinery.

[10] M. Chen, I. Laina, and A. Vedaldi. Training-free layout control with cross-attention guidance, 2023.

[11] E. Cherry and C. Latulipe. Quantifying the creativity support of digital tools through the creativity support index. *ACM Trans. Comput.-Hum. Interact.*, 21(4), jun 2014.

[12] L. B. Chilton, S. Petridis, and M. Agrawala. Visiblends: A flexible workflow for visual blends. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–14, New York, NY, USA, 2019. Association for Computing Machinery.

[13] S. R. L. B. Chilton, E. J. Ozmen, and S. Ross. Visifit: Ai tools to iteratively improve visual blends. 2019.

[14] O. Chouchoulas and A. Day. Design exploration using a shape grammar with a genetic algorithm. *Open House International*, 32:26–35, 06 2007.

[15] J. J. Y. Chung and E. Adar. Artinter: Ai-powered boundary objects for commissioning visual arts. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, DIS '23, page 1997–2018, New York, NY, USA, 2023. Association for Computing Machinery.

[16] J. J. Y. Chung and E. Adar. Promptpaint: Steering text-to-image generation through paint medium-like interactions. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, UIST '23, New York, NY, USA, 2023. Association for Computing Machinery.

[17] J. J. Y. Chung, S. He, and E. Adar. The intersection of users, roles, interactions, and technologies in creativity support tools. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference*, DIS '21, page 1817–1833, New York, NY, USA, 2021. Association for Computing Machinery.

[18] J. J. Y. Chung, S. He, and E. Adar. Artist support networks: Implications for future creativity support tools. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference*, DIS '22, page 232–246, New York, NY, USA, 2022. Association for Computing Machinery.

[19] N. Cross. Descriptive models of creative design: application to an example. *Design Studies*, 18(4):427–440, 1997. Descriptive models of design.

[20] Datasculptor. Image2linedrawing, 2023. Hugging Face Spaces.

[21] Design with Canva. How to Use ChatGPT to Design Like a Pro. https://www.youtube.com/watch?v=VmBLuvBf0xE, jul 27 2023. [Online; accessed 2023-12-04].

[22] C. Eckert and M. Stacey. Sources of inspiration: a language of design. *Design studies*, 21(5):523–538, 2000.

[23] J. Frich, M. M. Biskjaer, L. MacDonald Vermeulen, C. Remy, and P. Dalsgaard. Strategies in creative professionals' use of digital tools across domains. In *Proceedings of the 2019 Conference on Creativity and Cognition*, CC '19, page 210–221, New York, NY, USA, 2019. Association for Computing Machinery.

[24] J. Frich, L. MacDonald Vermeulen, C. Remy, M. M. Biskjaer, and P. Dalsgaard. Mapping the landscape of creativity support tools in hci. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–18, New York, NY, USA, 2019. Association for Computing Machinery.

[25] R. Gal, Y. Alaluf, Y. Atzmon, O. Patashnik, A. H. Bermano, G. Chechik, and D. Cohen-Or. An image is worth one word: Personalizing text-to-image generation using textual inversion, 2022.

[26] S. Garner and D. McDonagh-Philp. Problem interpretation and resolution via visual stimuli: the use of 'mood boards' in design education. *Journal of Art & Design Education*, 20(1):57–64, 2001.

[27] L. A. Gatys, A. S. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2414–2423, 2016.

[28] A. Goel. Design, analogy, and creativity. *IEEE Expert*, 12(3):62–70, 1997.

[29] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati, editors, *Human Mental Workload*, volume 52 of *Advances in Psychology*, pages 139–183. North-Holland, 1988.

[30] S. R. Herring, C.-C. Chang, J. Krantzler, and B. P. Bailey. 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*, CHI '09, page 87–96, New York, NY, USA, 2009. Association for Computing Machinery.

[31] B. Hirst. How artists overcome creative blocks. *The Journal of Creative Behavior*, 1992.

[32] J. Holinaty, A. Jacobson, and F. Chevalier. Supporting reference imagery for digital drawing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pages 2434–2442, October 2021.

[33] K. J. Holyoak and P. Thagard. *Mental leaps: Analogy in creative thought*. MIT press, 1996.

[34] T. Hope, R. Tamari, D. Hershcovich, H. B. Kang, J. Chan, A. Kittur, and D. Shahaf. Scaling creative inspiration with fine-grained functional aspects of ideas. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery.

[35] A. Ivanov, D. Ledo, T. Grossman, G. Fitzmaurice, and F. Anderson. Moodcubes: Immersive spaces for collecting, discovering and envisioning inspiration materials. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference*, DIS '22, page 189–203, New York, NY, USA, 2022. Association for Computing Machinery.

[36] J. Jacobs, J. Brandt, R. Mech, and M. Resnick. Extending manual drawing practices with artist-centric programming tools. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–13, New York, NY, USA, 2018. Association for Computing Machinery.

[37] D. G. Jansson and S. M. Smith. Design fixation. *Design studies*, 12(1):3–11, 1991.

[38] Y. Jeon, S. Jin, P. C. Shih, and K. Han. Fashionq: An ai-driven creativity support tool for facilitating ideation in fashion design. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery.

[39] Y. Kang, Z. Sun, S. Wang, Z. Huang, Z. Wu, and X. Ma. Metamap: Supporting visual metaphor ideation through multi-dimensional example-based exploration. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery.

[40] P. Karimi, N. Davis, M. L. Maher, K. Grace, and L. Lee. Relating cognitive models of design creativity to the similarity of sketches generated by an ai partner. In *Proceedings of the 2019 Conference on Creativity and Cognition*, C&C '19, page 259–270, New York, NY, USA, 2019. Association for Computing Machinery.

[41] P. Karimi, J. Rezwana, S. Siddiqui, M. L. Maher, and N. Dehbozorgi. Creative sketching partner: An analysis of human-ai co-creativity. In *Proceedings of the 25th International Conference on Intelligent*

*User Interfaces*, IUI '20, page 221–230, New York, NY, USA, 2020. Association for Computing Machinery.

[42] T. Karras, S. Laine, and T. Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[43] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila. Analyzing and improving the image quality of stylegan. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[44] A. Kerne. CollageMachine: An Interactive Agent of Web Recombination. *Leonardo*, 33(5):347–350, 10 2000.

[45] K. G. Kim, R. L. Davis, A. E. Coppi, A. Cattaneo, and P. Dillenbourg. Mixplorer: Scaffolding design space exploration through genetic recombination of multiple peoples' designs to support novices' creativity. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery.

[46] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick. Segment anything, 2023.

[47] H.-K. Ko, G. Park, H. Jeon, J. Jo, J. Kim, and J. Seo. Large-scale text-to-image generation models for visual artists' creative works. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*, IUI '23, page 919–933, New York, NY, USA, 2023. Association for Computing Machinery.

[48] J. Koch, A. Lucero, L. Hegemann, and A. Oulasvirta. May ai? design ideation with cooperative contextual bandits. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–12, New York, NY, USA, 2019. Association for Computing Machinery.

[49] J. Koch, N. Taffin, A. Lucero, and W. E. Mackay. Semanticcollage: Enriching digital mood board design with semantic labels. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, DIS '20, page 407–418, New York, NY, USA, 2020. Association for Computing Machinery.

[50] A. Koestler. The act of creation: A study of the conscious and unconscious in art, 1964.

[51] T. Lawton, K. Grace, and F. J. Ibarrola. When is a tool a tool? user perceptions of system agency in human–ai co-creative drawing. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, DIS '23, page 1978–1996, New York, NY, USA, 2023. Association for Computing Machinery.

[52] B. Lee, S. Srivastava, R. Kumar, R. Brafman, and S. R. Klemmer. Designing with interactive example galleries. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, page 2257–2266, New York, NY, USA, 2010. Association for Computing Machinery.

[53] S. Lee, C. Kong, D. Jeon, and N. Kwak. Aadiff: Audio-aligned video synthesis with text-to-image diffusion, 2023.

[54] J. Li, D. Li, S. Savarese, and S. Hoi. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models, 2023.

[55] Y. Li, H. Liu, Q. Wu, F. Mu, J. Yang, J. Gao, C. Li, and Y. J. Lee. Gligen: Open-set grounded text-to-image generation. *CVPR*, 2023.

[56] L. Lian, B. Li, A. Yala, and T. Darrell. Llm-grounded diffusion: Enhancing prompt understanding of text-to-image diffusion models with large language models, 2023.

[57] V. Liu, J. Vermeulen, G. Fitzmaurice, and J. Matejka. 3dall-e: Integrating text-to-image ai in 3d design workflows. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, DIS '23, page 1955–1977, New York, NY, USA, 2023. Association for Computing Machinery.

[58] A. Lucero. Funky-design-spaces: Interactive environments for creativity inspired by observing designers making mood boards. In J. Abascal, S. Barbosa, M. Fetter, T. Gross, P. Palanque, and M. Winckler, editors, *Human-Computer Interaction – INTERACT 2015*, pages 474–492, Cham, 2015. Springer International Publishing.

[59] A. Lucero, D. Aliakseyeu, K. Overbeeke, and J.-B. Martens. An interactive support tool to convey the intended message in asynchronous presentations. In *Proceedings of the International Conference on Advances in Computer Entertainment Technology*, ACE '09, page 11–18, New York, NY, USA, 2009. Association for Computing Machinery.

[60] J. Matejka, M. Glueck, E. Bradner, A. Hashemi, T. Grossman, and G. Fitzmaurice. Dream lens: Exploration and visualization of large-scale generative design datasets. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–12, New York, NY, USA, 2018. Association for Computing Machinery.

[61] C. Mou, X. Wang, L. Xie, Y. Wu, J. Zhang, Z. Qi, Y. Shan, and X. Qie. T2i-adapter: Learning adapters to dig out more controllable ability for text-to-image diffusion models, 2023.

[62] F. Müller-Wienbergen, O. Müller, S. Seidel, and J. Becker. Leaving the beaten tracks in creative work–a design theory for systems that support convergent and divergent thinking. *Journal of the Association for Information Systems*, 12(11):2, 2011.

[63] M. D. Mumford, M. I. Mobley, R. Reiter-Palmon, C. E. Uhlman, and L. M. Doares. Process analytic models of creative capacities. *Creativity research journal*, 4(2):91–122, 1991.

[64] Y. Nagai, T. Taura, and F. Mukai. Concept blending and dissimilarity: factors for creative concept generation process. *Design Studies*, 30(6):648–675, 2009.

[65] A. Q. Nichol, P. Dhariwal, A. Ramesh, P. Shyam, P. Mishkin, B. Mcgrew, I. Sutskever, and M. Chen. GLIDE: Towards photorealistic image generation and editing with text-guided diffusion models. In K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu, and S. Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 16784–16804. PMLR, 17–23 Jul 2022.

[66] C. Oh, J. Song, J. Choi, S. Kim, S. Lee, and B. Suh. I lead, you help but only with enough details: Understanding user experience of co-creation with artificial intelligence. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–13, New York, NY, USA, 2018. Association for Computing Machinery.

[67] OpenAI. Language models are few-shot learners, 2020.

[68] OpenAI. Gpt-4 technical report, 2023.

[69] T. Palmer, J. Bowman, and T. Geoco. 2023 design tools survey - ai, Dec 2023.

[70] M. L. Pilat and C. Jacob. Creature academy: A system for virtual creature evolution. In *2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence)*, pages 3289–3297, 2008.

[71] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever. Learning transferable visual models from natural language supervision. In M. Meila and T. Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 8748–8763. PMLR, 18–24 Jul 2021.

[72] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen. Hierarchical text-conditional image generation with clip latents, 2022.

[73] D. Ritchie, A. A. Kejriwal, and S. R. Klemmer. D.tour: Style-based exploration of design example galleries. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, UIST '11, page 165–174, New York, NY, USA, 2011. Association for Computing Machinery.

[74] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10684–10695, June 2022.

[75] N. Ruiz, Y. Li, V. Jampani, Y. Pritch, M. Rubinstein, and K. Aberman. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 22500–22510, June 2023.

[76] Satori Graphics. Ai Tools That TOTALLY Innovate A Designer's Life! https://www.youtube.com/watch?v=eTffPm_e1ko, jan 17 2023. [Online; accessed 2023-12-04].

[77] B. Shneiderman. Creating creativity: User interfaces for supporting innovation. *ACM Trans. Comput.-Hum. Interact.*, 7(1):114–138, mar 2000.

[78] P. Siangliulue, J. Chan, K. Z. Gajos, and S. P. Dow. Providing timely examples improves the quantity and quality of generated ideas. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*, C&C '15, page 83–92, New York, NY, USA, 2015. Association for Computing Machinery.

[79] D. K. Simonton. Scientific creativity as constrained stochastic behavior: the integration of product, person, and process perspectives. *Psychological bulletin*, 129(4):475, 2003.

[80] E. Stemasov, D. Ledo, G. Fitzmaurice, and F. Anderson. Immersive sampling: Exploring sampling for future creative practices in media-rich, immersive spaces. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, DIS '23, page 212–229, New York, NY, USA, 2023. Association for Computing Machinery.

[81] K. Sung-Bin, A. Senocak, H. Ha, A. Owens, and T.-H. Oh. Sound to visual scene generation by audio-to-visual latent alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6430–6440, June 2023.

[82] M. Turrin, P. von Buelow, and R. Stouffs. Design explorations of performance driven geometry in architectural design using parametric modeling and genetic algorithms. *Advanced Engineering Informatics*, 25(4):656–675, 2011. Special Section: Advances and Challenges in Computing in Civil and Building Engineering.

[83] M. K. Vladimir Vorobev. A paraphrasing model based on chatgpt paraphrases. 2023.

[84] S. Wang, S. Petridis, T. Kwon, X. Ma, and L. B. Chilton. Popblends: Strategies for conceptual blending with large language models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA, 2023. Association for Computing Machinery.

[85] W. Wang, F. Wei, L. Dong, H. Bao, N. Yang, and M. Zhou. Minilm: Deep self-attention distillation for task-agnostic compression of pre-trained transformers. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 5776–5788. Curran Associates, Inc., 2020.

[86] T. Wu, M. Terry, and C. J. Cai. 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*, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery.

[87] L. Yu and J. V. Nickerson. Cooks or cobblers? crowd creativity through combination. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, page 1393–1402, New York, NY, USA, 2011. Association for Computing Machinery.

[88] L. Zaman, W. Stuerzlinger, C. Neugebauer, R. Woodbury, M. Elkhaldi, N. Shireen, and M. Terry. Gem-ni: A system for creating and managing alternatives in generative design. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, page 1201–1210, New York, NY, USA, 2015. Association for Computing Machinery.

[89] E. Zhang and N. Banovic. Method for exploring generative adversarial networks (gans) via automatically generated image galleries. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery.

[90] L. Zhang, A. Rao, and M. Agrawala. Adding conditional control to text-to-image diffusion models, October 2023.

[91] N. Zhao, N. W. Kim, L. M. Herman, H. Pfister, R. W. Lau, J. Echevarria, and Z. Bylinskii. Iconate: Automatic compound icon generation and ideation. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–13, New York, NY, USA, 2020. Association for Computing Machinery.

[92] G. Zheng, X. Zhou, X. Li, Z. Qi, Y. Shan, and X. Li. Layoutdiffusion: Controllable diffusion model for layout-to-image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 22490–22499, June 2023.

[93] X. Zhu. Research guides: Machines and society: Chatgpt for visual design, Nov 2023.

# Acknowledgment

# Curriculum Vitae in Korean

이         름: 최 다 은

생 년 월 일: 1998년 11월 03일

전 자 주 소: daeun.choi@kaist.ac.kr

## 학         력

2014. 3. – 2017. 2.    현대청운고등학교 (졸업)

2017. 3. – 2022. 2.    한국과학기술원 전산학부 (학사)

2022. 3. – 2024. 2.    한국과학기술원 전산학부 (석사)

## 경         력

2019. 3. – 2024. 2.    한국과학기술원 전산학부 조교

## 연 구 업 적

1. **DaEun Choi**, Sumin Hong, Jeongeon Park, John Joon Young Chung, Juho Kim, "CreativeConnect: Supporting Reference Recombination for Graphic Design Ideation with Generative AI," under review.

2. Kihoon Son, **DaEun Choi**, Tae Soo Kim, Young-Ho Kim, Juho Kim, "GenQuery: Supporting Expressive Visual Search with Generative Models," under review.

3. Kihoon Son, **DaEun Choi**, Tae Soo Kim, Juho Kim, "Demystifying Tacit Knowledge in Graphic Design: Characteristics, Instances, Approaches, and Guidelines," under review.

4. Jeongeon Park, **DaEun Choi**, "AudiLens: Configurable LLM-Generated Audiences for Public Speech Practice," *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology.* 2023.

5. Jeongyeon Kim, **DaEun Choi**, Nicole Lee, Matt Beane, Juho Kim, "Surch: Enabling Structural Search and Comparison for Surgical Videos," *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.* 2023.

6. Tae Soo Kim, **DaEun Choi**, Yoonseo Choi, Juho Kim, "Stylette: Styling the Web with Natural Language," *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems.* 2022.

7. Hyunsung Cho, **DaEun Choi**, Donghwi Kim, Wan Ju Kang, Eun Kyoung Choe, Sung-Ju Lee, "Reflect, not regret: Understanding regretful smartphone use with app feature-level analysis," *Proceedings of the ACM on human-computer interaction 5.CSCW2 (2021).*