LumiMood: A Creativity Support Tool for Designing the Mood of a 3D Scene

Jeongseok Oh
jungseok17@gm.gist.ac.kr
Gwangju Institute of Science and Technology
Gwangju, Republic of Korea

Seungju Kim
dibo3440@gm.gist.ac.kr
Gwangju Institute of Science and Technology
Gwangju, Republic of Korea

SeungJun Kim*
seungjun@gist.ac.kr
Gwangju Institute of Science and Technology
Gwangju, Republic of Korea

Figure 1: (a) LumiMood is an AI-driven creativity support tool for mood designing of game scenes. (b) LumiMood takes the current scene as the input and stylizes the mood of the scene by adjusting the lighting and the post-processing effects. (c) Designers can utilize LumiMood to intentionally evoke specific emotions in users through their scene design.

ABSTRACT

The aesthetic design of 3D scenes in game content enhances players' experience by inducing desired emotions. Creating emotionally engaging scenes involves designing low-level features, such as color distribution, contrast, and brightness. This study presents LumiMood, an AI-driven creativity support tool (CST) that automatically adjusts lighting and post-processing to create moods for 3D scenes. LumiMood supports designers by synthesizing reference images, creating mood templates, and providing intermediate design steps. Our formative study with 10 designers identified distinct challenges in mood design based on the participants' experience levels. A user study involving 40 designers revealed that using LumiMood benefits the designers by streamlining workflow, improving precision, and increasing mood intention accuracy. Results indicate that LumiMood supports clarifying mood concepts and improves interpretation of lighting and post-processing, thus resolving the challenges. We observe the effect of template based designing and discuss considerable factors for AI-driven CSTs for users with varying levels of experiences.

*Corresponding author.

CHI '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0330-0/24/05
https://doi.org/10.1145/3613904.3642440

CCS CONCEPTS

• **Human-centered computing → Systems and tools for interaction design**; • **Computing methodologies → Artificial intelligence**.

KEYWORDS

creativity support tool, affective computing, graphics design, artificial intelligence

ACM Reference Format:
Jeongseok Oh, Seungju Kim, and SeungJun Kim. 2024. LumiMood: A Creativity Support Tool for Designing the Mood of a 3D Scene. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 21 pages. https://doi.org/10.1145/3613904.3642440

1 INTRODUCTION

Designing an immersive and emotionally engaging environment is one of the main tasks in the development of digital games [35]. The emotional tone of a game is crafted through a complex interplay of elements, including narrative, sound, and visual aesthetics [5, 20]. Among these, studies have shown that visual stimuli are the significant factors that trigger a range of emotional responses [17], affecting the players" experience [27], shaping their motivations [23, 31], and improving their overall engagement [64]. In addition, passive visual stimuli such as lighting, color, and static scenes can induce the desired emotion in users [33]. In affective computing studies, visual stimuli are categorized into two features: high-level and low-level features [76, 78, 81, 82]. High-level features contain semantic attributes associated with specific concepts or objects, such as 3D models in game scenes. Low-level features include the overall attributes of a visual stimulus, such as color distribution, contrast, and brightness. A combination of high-level semantics and low-level color features guides users to establish a specific emotional connection with the visual stimuli [76, 82]. Therefore, both these factors should be carefully considered when designing game scenes.

With the advances achieved in computer vision, creativity support tools (CSTs) employ artificial intelligence (AI) models for automated generation of 3D models and their positional relationships in 3D scene development [19, 40]. These CSTs reduce the effort required for scene design by substituting labor-intensive 3D modeling procedures. Although studies have successfully automated design workflows, they have primarily focused on synthesizing the shapes of 3D models [67, 75, 80], which pertains to the semantic feature domain. In contrast, less attention has been paid to the automatic generation of low-level features, such as the overall color distribution or contrast. Low-level features, such as color distribution, contrast, and brightness, set the mood of the scene and evoke specific emotions in the users. Panda et al. have noted that the semantic meanings of visual stimuli could be unrelated to the actual emotional response of viewers due to the color distributions [49]. The potential importance of low-level features in building the emotional mood should be considered in creating game scenes, but there's a scarcity in CST studies supporting the mood design of the 3D scenes. To bridge this gap, we aimed to build a creativity tool that enables the automatic design of the low level features.

The aim of this study is to develop an AI-driven CST that designs the mood of a 3D scene by adjusting the lighting and post-processing to manipulate low-level features for the emotional affection of the players. Previous studies have highlighted the importance of lighting in the aesthetics of the entire scene [15, 32], but have not explored the mood creation of a 3D scene. First, a formative study was conducted to understand the workflow and challenges involved in designing the low-level features of 3D scenes for mood direction. We identified three major challenges faced by designers: 1) abstractness of mood concepts, which makes it difficult to realize mood design; 2) labor-intensive design workflow; and 3) ambiguity in selecting the properties to adjust, which is due to the similar yet different effects of various lighting and post-processing properties.

To address these challenges in creating the mood of a 3D scene, we present LumiMood, an AI-driven CST that automatically adjusts the lighting and post-processing to design low-level features. Figure 1 illustrates the overall concept of LumiMood. When the designer inputs an initial scene along with the desired emotional affect, LumiMood suggests templates of scenes with specific moods. LumiMood is designed to function within Unity, one of the most popular game development tools in the industry. It aims to assist game designers in manipulating the mood of game scenes by managing the lighting and post-processing properties listed in Table 1. We addressed the above challenges by introducing three components of LumiMood: 1) a reference image generator that synthesizes and recommends designs, 2) a template scene creator that automatically adjusts the lighting and post-processing properties, and 3) a design step recaller that assists designers by providing the intermediate design steps. Because designers prioritize different challenges depending on their experience level, our user study involved 20 professional and 20 novice designers. We conducted two experiments: a scene replication task and mood designing task. The results suggest that AI-driven mood design can enhance the efficiency and performance of mood creation in game scenes. Additionally, we analyzed the effect of templates in the design process. Based on the results, we discuss the differences between professionals and novices working with CSTs. The main contributions of this study are as follows:

- **Understanding the challenges in mood design based on the experience level:** Through a formative study with 10 designers, we explored the significance and challenges of mood design in 3D game scenes. Our findings reveal distinct challenges faced by designers, contingent on their experience level.
- **Development of LumiMood:** We introduce LumiMood, a system that performs mood design within the Unity engine. LumiMood automatically adjusts the lighting and post-processing to facilitate mood creation.
- **User study with professional and novice designers:** Our user study involving 40 designers (20 professionals and 20 novices) provided insights into how LumiMood can enhance the design performance across different experience levels.
- **Implications for tailoring CSTs to experience level:** Our study presents a comparison between the designers based on the experience level for developing CSTs incorporating AI features for inspiring users. We also discuss the specific needs and challenges of designers with varying levels of expertise.

2 RELATED WORKS

2.1 AI-driven CSTs for automatic design generation

AI models have the potential to enhance the creativity and efficiency of users. Previous studies have demonstrated that AI can support creative tasks, such as 3D modeling [60], drawing [18], and coloring cartoons [77]. To generate new content, CSTs incorporate generative models, such as generative adversarial networks (GANs) or diffusion models [24]. GANs and diffusion models have shown great performance in image synthesis, and commercial services that support image generation, such as DALL-E [56] and Stable Diffusion [58] have been developed. Combined with natural language

Table 1: List of properties that LumiMood automatically adjusts for designing the mood of a scene. LumiMood is implemented to be expandable to allow designers to add or remove specific properties during the design process.

	Functionality	Properties
Lighting	Light sources	Light intensity, Light color, Light direction, Light position, Flare size, Flare tint, Surface tint
Post-processing	Bloom	Threshold, Intensity, Tint, Scatter
	Channel mixer	Red, Green, Blue
	Chromatic aberration	Intensity
	Color adjustments	Contrast, Color filter, Hue shift, Saturation
	Fog	Fog attenuation distance, Base height, Maximum height, Max fog distance, Albedo
	Depth of field	Focus distance
	Film grain	Intensity, Response
	Lens distortion	Intensity
	Lift gamma gain	Lift, Gamma, Gain
	Shadows midtones highlights	Tint, Start, End
	Vignette	Color, Intensity, Smoothness
	White balance	Temperature, Tint

processing, generative models can offer greater creative control to users by including prompts [30]. By utilizing the prompts, users can interact with the AI models [40], simplify complex tasks [74], and access the AI models more easily [28].

Studies have proposed tools that utilize generative models combined with prompting capabilities to improve the efficiency of designing 3D scenes. DreamFusion [52] has introduced text-based mesh generation, with a dedicated texture map that can also be guided by prompts. Similarly, the Unity engine has unveiled Muse & Sentis, which enable the automatic generation of 3D models and animations within the game engine using natural languages. Other tools, including face generation [25], mesh composition based on affective keywords [9], and avatar generation [72], also permit users to edit 3D models. These tools mainly concentrate on synthesizing the meshes and textures of an object, which are the high-level features of visual stimuli (e.g., semantics), when constructing 3D scenes. Tools such as GET3D [19] enable the stylization of colors by creating textures but allow limited editing of low-level features, such as color distribution, contrast, and brightness of the entire scene.

As studies have focused on automated design using generative models, attention has been given to finding a balance between automation and human control. With the power of generative models, CSTs can offer end-to-end automated workflows to users. Automation tools are useful in building designs with less effort, but limit the designer's freedom and control over the content creation process [26, 77]. Oh et al. proposed DuetDraw, a human–AI collaborative drawing system, and reported that using AI as a supplementary method while guaranteeing user control can improve the usability [48]. Conversely, abundant options and complexity associated with excessive control can overwhelm users [46, 48, 57]. Thus, several studies have suggested that the balance between automation and control should be carefully considered when designing CSTs. Amershi et al. noted that because AI models can sometimes produce incorrect results due to the variety of options, CSTs should allow users to understand and control the system [4].

This study aims to help designers bridge the gap between abstract mood concepts and concrete visual representations when designing the mood of a scene. In addition to developing an automated system for adjusting the lighting and post-processing, we created a visualizer that provides the intermediate steps in the design process. The visualizer provides designers a sense of control and better understanding of the design process using the automation system.

2.2 Lighting design for evoking emotions

Lighting and post-processing are considered to set the mood of the scene. Lighting designers consider lighting as the primary factor for designing the mood of a space [37]. Cinematic designers also design the moods of a scene by continuously adjusting the lights [7]. Lighting in 3D engines adds color to a scene, influences the appearance of textures, and sets the mood [42, 65]. Post-processing is used for enhancing the mood of the scene. Post-processing is performed after rendering and can alter or enhance the atmosphere created by the lighting effects, ultimately changing the overall appearance and mood of the scene [36]. Color grading, such as lift, gamma, and gain, helps establish the desired mood by using different color tones to evoke different emotions. For instance, a warm, golden hue might convey nostalgia or warmth, whereas a blue or gray tone might evoke feelings of coldness or sadness [71]. Adjusting the contrast, brightness, saturation, and sharpness through post-processing techniques can affect the mood [3]. Moreover, effects such as lens flare, light scattering, and shadow enhancement create a specific mood or atmosphere [6].

The relationship between lighting and emotion is explained by color symbolism [47]. The light settings of a scene determine the shadows, highlights, and contrasts [50], and the color of the light source determines the overall temperature of the scene [6], thereby affecting human emotions. Colors are major factors in forming human emotions [13]. They influence human affection and behavior, thus affecting decision-making and conversation [16]. For example, warm colors such as red and orange are more arousing than cool colors such as blue or green owing to their association with passion [43, 68]. Furthermore, the primary colors of light red, green, and blue evoke different emotions: disgust, neutrality, and happiness, respectively [21].

In this study, we conducted interviews with designers to explore the methods they employ to evoke emotional responses in scenes. Building upon previous studies, we focused on understanding how game designers specifically design the mood of a scene. Additionally, we examined the process of designing the lighting and post-processing, as these elements are crucial in shaping the overall mood of a scene. Through this investigation, we aimed to reveal the design workflow and gain insights into the challenges faced by designed when designing the desired mood of a scene.

2.3 Emotion models and visual emotion analysis

To measure human emotions, categorical emotion state (CES) and dimensional emotion space (DES) models are introduced. The CES model uses a discrete number of basic emotion classes such as anger, disgust, fear, joy, sadness, and surprise. The CES model has the advantage that the represented human emotion is labeled in natural language (e.g., sadness, anger) to easily comprehend the categorized results. Popular CES models used in affective computing include Ekman [14], which classifies emotions into 6 classes of anger, disgust, fear, happiness, sadness, and surprise; Mikels [44], which classifies emotions into fear, sadness, disgust, anger, amusement, contentment, awe, and excitement; and Plutchik [51], which explains emotions as combinations of joy, sadness, trust, disgust, fear, anger, surprise, and anticipation. In contrast, the DES model introduces emotional dimensions: valence and arousal (VA). Valence refers to the positive and negative states of emotion, whereas arousal refers to the intensity of excitement. Emotions are positioned on the VA plane and each emotion is scored based on valence and arousal [59].

Studies on visual emotion analysis (VEA) have presented deep learning-based methods to identify the features of visual stimuli that evoke emotions. The main challenge in VEA is overcoming the affective gap, which is the disparity between human emotions and image features [81]. To address this gap, studies have proposed affective image datasets for both CES and DES models. Some examples are Flickr and Instagram CES datasets based on Mikel's emotion model [79] and International Affective Picture System (IAPS), a DES dataset with VA labels [38]. VEA studies have used these datasets to train AI models for extracting important features from the images and explain how human emotions are structured from the visual stimuli.

In this study, we utilized Parrott's emotion model [53] to train the reference-image-generation diffusion model of LumiMood. Parrott's emotion model provides three levels of emotion labels: primary, secondary, and tertiary. At the primary level, Parrott's model classifies emotions as positive and negative and subdivides these two classes into six base emotions at the secondary level: love, joy, surprise, anger, sadness, and fear. At the tertiary level, the model further subdivides the secondary emotions into 25 fine-grained emotion classes. To adopt Parrott's emotion model, we specifically used the WEBEmo dataset, the largest CES dataset [49], and generated captions of each image inside the dataset to use the image–caption pair to train the AI model. Parrott's model allows detailed emotion information to be captured in the generated image caption, whereas other emotion models provide only one label. Further, compared to the DES models using VA values, using CES category names allows the users to direct the synthesized image with emotion keywords [83].

3 FORMATIVE STUDY

To understand how game designers create the mood of a game scenes, we conducted interviews with game designers. Our formative study aimed to capture the challenges involved in designing the moods of scenes and gain insight into how an AI-based tool can help overcome these challenges. We recruited 10 participants with at least one year of game development experience. The experience levels of the participants ranged between 1–10 years (M=4.5, SD=2.6). All participants reported using the Unity engine as their primary development environment; some had also used other programs such as Maya, Cinema 4D, and Photoshop. Table 2 provides detailed information about the participants.

3.1 Interview Process

The interviews were semi-structured and allowed both targeted questions and open-ended discussions. We designed the interview questions, shown in Table 3, to understand 1) the importance of mood design when creating a game, 2) actual method of designing the moods of game scenes, and 3) challenges in mood creation. The interviews lasting an average of one hour were conducted remotely. After the interviews, we analyzed the transcribed records using the reflexive thematic analysis method [10]. Through the interviews, we established a design objective for our system to enable designers to resolve the challenges in mood design.

3.2 Results: formative study

3.2.1 Importance of mood design. The designers responded that designing the proper mood for 3D scenes is important in game development (item1 M=6.6, SD=0.663). Because the mood of a game scene is the main factor in shaping the player's emotional response and overall game experience, the designers responded that they pay careful attention to designing the mood of the scene; *"The mood of a scene can influence players' immersion, making it a key aspect of game design" (P2)*, *"Evoking the desired emotion is especially important in designing virtual reality games or first-person shooting games" (P1)*, *"The first impression regarding the game is strongly correlated with the mood of the scene and affects all subsequent user experiences" (P6)*. The participants also highlighted the importance of a more meticulous mood design to ensure the development of a high-quality game; *"Matching the desired atmosphere and emotions in a scene with the initial concept of the game influences the overall completeness of the game" (P3)*, *"Game players are more influenced by the game's mood than by the technological completeness of the game" (P4)*.

3.2.2 Lighting and post-processing as tools for mood design. We investigated the methods used by designers to create the mood of a scene. Through observations, we found that the participants primarily used two functionalities to create the mood: lighting and post-processing. All participants reported using lighting to express the

Table 2: Participant details for the formative study. Each participant had at least 1 year of experience in game development using Unity engine.

ID	Age	Gender	Key responsibilities	Experience	Programming skills
P1	34	Male	Virtual reality scene design	10 years	Unity engine
P2	26	Male	Game level design	6 years	Unity engine, Maya
P3	25	Female	3D modeling	6 years	Unity engine, Rhino, Photoshop
P4	25	Male	Overall game development	3 years	Unity engine
P5	22	Female	Game scene design	1 year	Unity engine
P6	23	Male	3D environment design	5 years	Unity engine
P7	25	Female	3D motion graphics design	5 years	Unity engine, Cinema 4D, Octane
P8	23	Male	Overall game development	4 years	Unity engine
P9	24	Female	Virtual reality content development	4 years	Unity engine
P10	23	Female	Game scene design and programming	1 year	Unity engine

Table 3: Questions asked in the formative study. Further questions were asked based on the answers of the participants.

Item	Question	Answer Type
1	(Importance of mood design) How important is it to design the mood of a game scene?	Likert(1-7), Open-Ended
2	(Design methods & workflow) How do you typically design the mood of a scene?	Open-Ended
3	(Challenges in mood design) What is the most challenging part of designing the mood of a scene in the entire workflow?	Open-Ended

mood of a scene. Moreover, most participants reported using post-processing to adjust low-level features, such as color distribution, contrast, and brightness. We discovered that designers have limited freedom in designing a mood with high-level semantic features because the features must be consistent with the game's concept and level of design (e.g., designers cannot place car objects when designing a medieval fantasy game); *"Lighting and post-processing are the most frequently used functions when designing the mood of a game. A mesh or texture is less used because it must strictly follow the game concept" (P2)*, *"Different moods can be directed with different lighting settings within the same meshes and textures" (P7)*.

In our interviews, we explored how designers employ lighting and post-processing to establish the desired mood in a 3D scene. The participants indicated that the colors set by the lighting and post-processing affect the mood of the scene. Designers typically adjust the properties of lighting and post-processing concurrently, with a tendency to prioritize the lighting adjustments before addressing the post-processing properties. This preference stems from the predominant use of post-processing to enhance the final aesthetics of the scene. Designers initiate mood construction by configuring the lighting properties first; *"The position and direction of the light sources determine the bright and dark areas, shaping the overall mood of the scene" (P10)*, *"The orange color of the light makes the scene warm and calm" (P1, P3)*, *"When expressing the concrete texture of a building, using blue light can create a colder feeling" (P4)*. Post-processing is used to enrich the atmosphere or generate an ambiance that lighting alone cannot achieve, by incorporating special effects; *"The vignette effect produces a nostalgic mood, guiding players' focus to the center of the screen" (P7)*, *"A warm feeling in the scene is easily created by increasing the temperature property in the white balance effect" (P8)*, *"Post-processing is responsible for rendering the scene more realistic and engaging. For instance, bloom effects are commonly employed for reflective materials to enhance their shininess" (P6)*.

3.2.3 Design workflow. In the formative study, we asked designers about their workflow for designing the mood of a scene. We found that participants iteratively improve their designs by repeating ideation, implementation, and evaluation until they achieve a satisfactory scene design. This observation aligns with a previous study that suggests lighting design requires a trial-and-error process [29]. In the ideation stage, designers first identify the visual representations of a specific mood. They either conceive the design themselves *(P8, P10)*, follow art theories *(P1, P3)*, or search for reference images *(P1-P7, P9)* to help envision the desired mood imagery. Next, the designers adjust their scene to achieve a satisfying result that matches their envisioned imagery *(P1-P10)*. This iterative process involves adjusting the lighting and post-processing properties, such as the position and color of light sources. Finally, they evaluate the resulting design outcome and continue modifying their scene design until they achieve a satisfying mood design. At this stage, the designers self-evaluate their work *(P2-P5, P7, P9, P10)* or ask for feedback from others *(P1, P6, P8)*. We set this iterative design workflow as the baseline for our study.

3.2.4 Challenges in mood design. Most designers described mood designing as challenging, requiring a deep understanding of both the game engine and visual aesthetics. By analyzing the interview results, we observed three major challenges in designing the mood of a scene. These challenges arise when the designer translates abstract mood concepts into concrete understanding (Figure 2).

- **Challenge1: Translating abstract mood concepts into visual representations.** In the formative study, participants noted that mood designing is creatively demanding. Because mood concepts do not have specific shapes, envisioning the actual visual representation of a mood requires translating

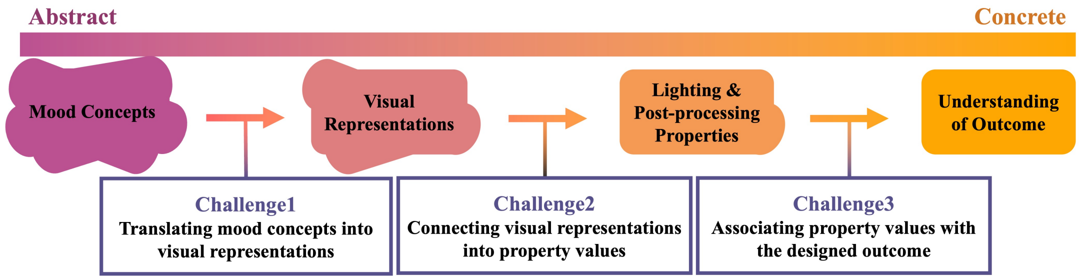


Figure 2: Designers translate abstract concepts into concrete ones. Challenges arise in each translation.

the abstract concepts into actual visual representations; *"Designing the mood is more challenging than designing 3D objects, as the mood lacks a specific shape" (P2), "Envisioning the appropriate mood for a given scene is difficult. Finding suitable reference images that fit my case is also challenging" (P6), "When trying to create a scary atmosphere, besides simply darkening the scene, I am unsure about the specific design elements I should use" (P9).*

- **Challenge2: Connecting visual representations to property values.** Designers follow a trial-and-error process to create the visual representation of a scene. To determine the precise property value to make the scene resemble the envisioned imagery, designers need to set an initial value, evaluate its alignment with their intended imagery, and then repeat the process of adjusting the value until they identify the optimal setting that closely matches their desired visual representation. This iteration process is considered labor-intensive and difficult to adopt, as the designers have to adjust and evaluate their intermediate design outcome constantly; *"I have to set the properties of lighting and post-processing until I get a satisfactory result. This takes so much time" (P8), "There are many properties that affect the visual aesthetics of the scene. I have to tweak all of them until I get a satisfactory result" (P10).*
- **Challenge3: Associating the property settings with design outcome.** To achieve successful design, designers need a comprehensive understanding of how each property affects the different parts of the scene. This understanding necessitates interpreting the algorithms behind the lighting and post-processing effects. Consequently, designers find it difficult to select the appropriate properties to adjust to achieve their desired outcome during the design process. Hence, the designers reported that they use only the properties they understand; *"Even experienced designers have a hard time understanding what will happen when they change certain parameters" (P1), "Most designers around me do not understand the exact functionalities of each parameter. I too use only a few functionalities that I fully understand" (P5).*

3.3 Discussion: formative study

From the formative study, we discovered three challenges designers face when creating the mood of a scene. We analyzed the interview results by grouping the participants according to their experience level to see how the challenges differ based on the experience level.

3.3.1 Prioritization of challenges differs with experience level. As the accumulation of experience is the main source of development of expertise [12], we categorized the participants into two groups according to the experience level. We assigned participants with at least 5 years of game development experience to the high-experience group and those with less than five years of experience to the low-experience group, thereby dividing the participants into the top 50% and bottom 50%. Thus, each group comprised five participants. We analyzed our interview results for both the high-experience group (*P1*, *P2*, *P3*, *P6*, *P7*) and low-experience group (*P4*, *P5*, *P8*, *P9*, *P10*).

We identified that the order of prioritization of the importance of the three challenges mentioned above differed between the groups. Challenge1, the challenge of translating abstract mood concepts into visual representations, was the major issue for the high-experience group because they considered the issue of abstractness of the mood concepts more challenging than Challenge2 or Challenge3; *"Getting the inspiration for the design is the most time-consuming part." (P7).* The low-experience group was mostly concerned with Challenge2, as the most frequently mentioned challenge was the burden of iterating during the design workflow; *"Designing with the actual properties is the hardest part, as it requires intensive iteration." (P9), "I often get lost while setting the properties." (P8).* Both groups consistently reported facing Challenge3 each time they had to identify the property to modify.

Our findings regarding the prioritization of challenges based on the experience level align closely with previous studies. Ahmed et al. observed a similar pattern among experienced and novice designers, where professionals perform a preliminary evaluation of the ideation results while novices directly implement their designs [2]. This evaluation stage performed by experienced designers makes it difficult to solve Challenge1, which involves translating abstract mood concepts into concrete visual representations. Additionally, Cross et al. noted that experts attempt to frame the overall problem at the beginning of the design process [12], making the abstractness of mood concepts challenging for the high-experience group. In contrary, novices found Challenge2, the trial-and-error process, to be the hardest. A previous study indicated that because novice designers lack sufficient experience to evaluate the intermediate processes, improving the design quality through iteration is challenging for them [2]. Thus, Challenge2 was the most significant problem for the low-experience group.

3.3.2 Design objectives. To address the challenges, we present LumiMood, an AI-driven CST for mood design of a 3D scene. We identified the following design objectives:

- **Support bridging between abstract mood concepts and their concrete visual representations (Challenge1).** Designers struggle with transforming the abstract concepts of mood designs into concrete visual elements. Providing reference images to guide mood design can provide inspiration and relieve ambiguity in envisioning the scene.
- **Support workflow for efficiency in adjusting and evaluating the designed scene (Challenge2).** The designers have to adopt a trial-and-error process for setting the lighting and post-processing properties, which is labor-intensive and time-consuming. Generating templates for lighting and post-processing properties would decrease the workload and time consumption of the design process by allowing the designers to develop their design from a roughly adjusted scene.
- **Support guides for better understanding and control of lighting and post-processing properties (Challenge3).** Designers have reported difficulty in selecting the appropriate properties of the lighting and post-processing effects to achieve the desired outcome during the design process. By viewing the intermediate stages that show how the appearance of a scene changes when certain properties are modified, designers can better understand the effect of each lighting and post-processing property.

4 LUMIMOOD

4.1 System overview and LumiMood workflow

LumiMood is an automatic mood design tool for 3D scenes, which works within the Unity engine. It adjusts the lighting and post-processing effects to design the mood of a given 3D scene to achieve the desired emotional effect on game players. We added material tint as a lighting property when LumiMood is used to adjust a scene. To address the three challenges observed in our formative study, LumiMood includes three main components: a reference image generator (Generator), template scene creator (Creator), and design step recaller (Recaller). The Generator addresses Challenge1, which arises from the abstractness of mood concepts. It is a finetuned diffusion-based model that synthesizes affective images for a given emotion class and can be used as a reference image for designing. The Creator addresses Challenge2, which arises from the labor-intensive trial-and-error design process. It creates a template scene that resembles the reference affective image synthesized by the Generator to the extent possible. The Recaller addresses Challenge3, which relates to the selection of properties to modify for incorporating the desired mood into the design. It provides the intermediate steps in LumiMood's design process, thus enhancing the designer's understanding of the effects of each lighting and post-processing property while offering control over the entire design workflow. Designers can use LumiMood to generate template scene designs and develop their designs based on the template scenes suggested by LumiMood. The three components of LumiMood are integrated with the Unity engine, as shown in Figure 3.

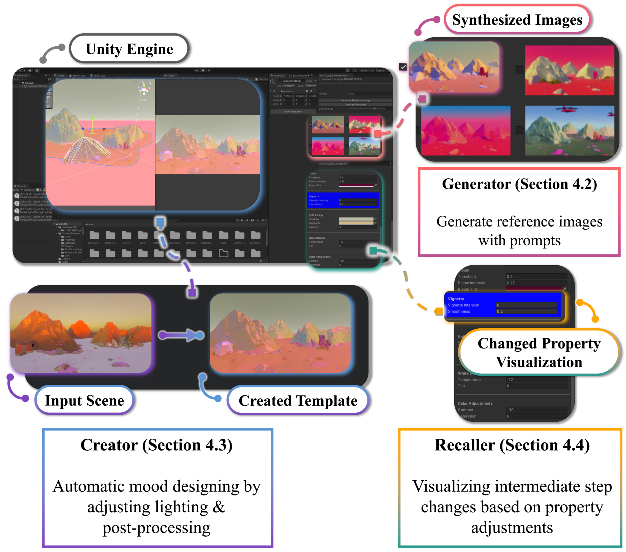


Figure 3: Overall framework of LumiMood. LumiMood uses the current scene as input to generate the mood design. When requested by the designer, LumiMood synthesizes images that can be used as reference images. The designer can then generate the template scene with the Creator, which automatically adjusts the lighting and post-processing to make the scene resemble the reference image. Once the template scene is generated, the designer can traverse through LumiMood's intermediate design steps using the Recaller.

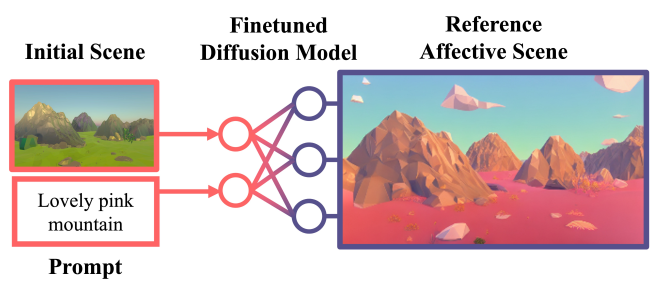


Figure 4: A brief schematic of the Generator. The Generator receives a prompt about the mood that the designer wants to convey in the current scene and synthesizes an image that the designer can use as reference.

4.2 Generator: Synthesizing the reference images

The first component of our system is the Generator, an AI model that synthesizes the reference affective images based on the emotions provided as keywords. The Generator addresses Challenge1, where designers struggle with the abstractness of mood concepts. As shown in Figure 4, the Generator receives the image of the

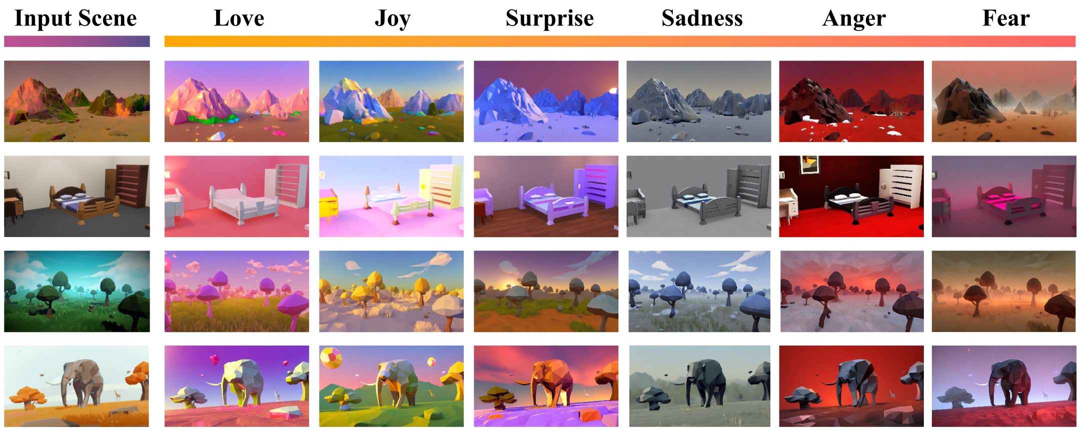


Figure 5: Example of images synthesized by the Generator. The images were synthesized using basic keywords of emotion classes: love, joy, surprise, sadness, anger, and fear.

current scene and a natural language prompt that describes the desired mood that the designer aims to create. Subsequently, it generates specific mood designs that evoke the intended emotion through image-to-image translation. Figure 5 shows the examples of translated images generated based on an input image and basic keywords comprising emotion classes, such as love, joy, surprise, sadness, and fear.

4.2.1 Image generation model. We first trained the diffusion model using the WEBEmo dataset, which is based on Parrott's emotion model. To train the Generator, we generated a caption containing emotion keywords for each image in the dataset and fine-tuned the diffusion model using image–caption pairs. Captions were generated using ClipClap [45], which leverages CLIP [54] and GPT [55] for generating natural language captions. Each image in the WEBEmo dataset was captioned, and the resulting descriptions were concatenated with three levels of image classes corresponding to the hierarchical classification of emotions in Parrott's model. For example, when captioning an image with the base class "love," we added the corresponding 2-class primary class "positive" and 25-class fine-grained tertiary information. When generating the images, the diffusion model used the given emotion class as a positive prompt and other emotion classes as negative prompts. This approach allowed the model to generate images that aligned with the desired emotion.

4.2.2 Prompting. The Generator is fine-tuned using the image–caption pair data, allowing it to guide the image synthesis using natural language prompts. We designed a UI element as an add-on inside the Unity engine, making it easy for designers to type prompts directly into the engine. When a designer requests reference images, the typed prompt is sent to the Generator backend, which begins synthesizing the translated images. Because the Generator is optimized based on the diffusion network pre-trained on the LAION dataset [61], the resulting image typically retains the semantic meaning of the input. However, in specific instances, the model may not accurately preserve the semantic meaning (e.g., coloring the sea as red). To mitigate this, we designed LumiMood to generate four affective images for each synthesis request; these images are subsequently displayed in the add-on window. Designers can select an image or modify the prompt to regenerate the images if none of the four images is suitable. After selecting an image and proceeding to template creation, the Creator is triggered. Using the selected image as the ground truth, the Creator automatically adjusts the lighting and post-processing properties.

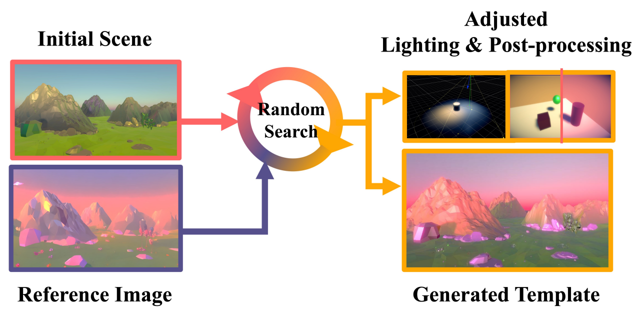


Figure 6: A brief schematic of the Creator. The Creator receives a reference image and adjusts the lighting and post-processing properties to create a template scene that resembles the reference image.

4.3 Creator: Adjusting the lighting and post-processing properties

The second component of LumiMood is the Creator, which was designed to resolve Challenge2. This challenge involves the labor-intensive baseline workflow required to connect visual representations to actual property values. The Creator adjusts the lighting and post-processing to stylize the low-level features of the scene, such as color distribution, contrast, and brightness, to evoke the desired emotions in the game players. It takes the reference image synthesized by the Generator as the input and outputs a template scene that appears as similar as possible to the reference image. Figure 6 demonstrates the operation of the Creator, and Figure 7

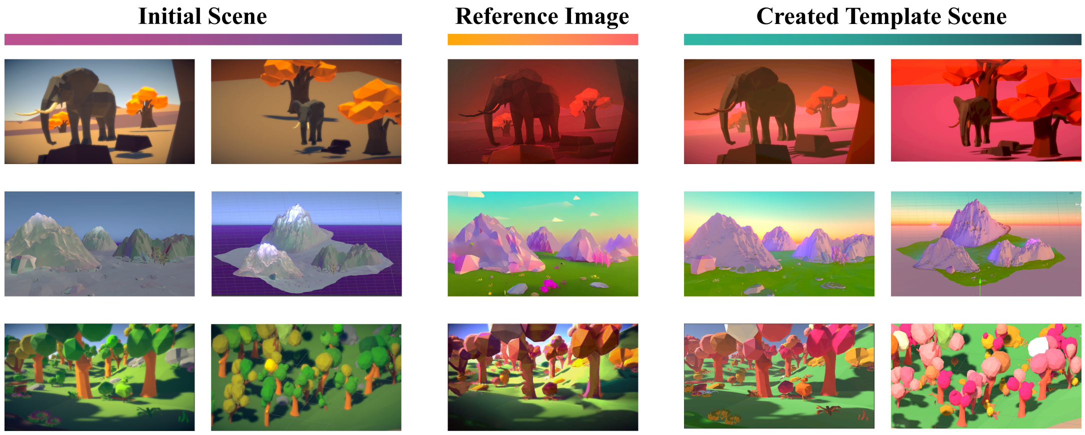


Figure 7: Example images of the created template scenes of the Creator. The template scenes are created based on the given initial scene by adjusting the lighting and post-processing to resemble the reference image, which were synthesized by the Generator.

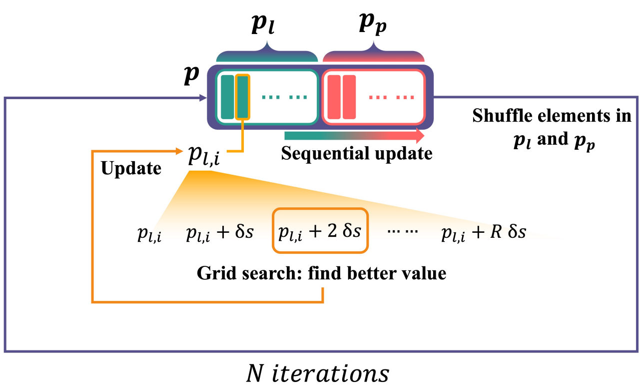


Figure 8: Grid search algorithm implemented in the Creator. The algorithm continuously searches for better property values that resemble the reference image obtained from the Generator.

shows the examples of template scenes created using the reference images synthesized by the Generator.

4.3.1 Template creation algorithm. The Creator works by reducing the disparity between the current scene and a reference image, ultimately creating a template scene that closely resembles the reference image. We used a mean squared error (MSE) function to quantify the difference and implemented an optimization algorithm to lower the MSE between the scene and reference image by adjusting the lighting and post-processing properties. The lighting and post-processing adjustment process can be defined as follows:

$$\mathbf{p} = \arg\min_{\mathbf{p}\prime} \frac{1}{WH} \sum_{i=0}^{WH} (Y_{i,\mathbf{p}\prime} - Y_{i,GT})^2, \quad (1)$$

where Y_{GT} is the reference image, $Y_{\mathbf{p}\prime}$ is the rendered image of the game scene projected onto the main camera with the lighting and post-processing properties of $\mathbf{p}\prime$, and W and H are the width and height of the reference image, respectively. The equation for the optimized property $\mathbf{p}$ cannot be solved mathematically because formulating the rendered image $Y_{\mathbf{p}}$ in terms of $\mathbf{p}$ requires prior knowledge of the relationship between the properties and appearance of the resulting scene [41]. Therefore, we implemented the grid search algorithm to computationally determine the values of $\mathbf{p}$ that would eventually evoke the desired emotions. Figure 8 depicts the grid search algorithm. The algorithm traverses the property space and determines the values of $\mathbf{p}$. The algorithm is structured as follows:

(1) **Initialization.** The template creation algorithm begins by aggregating all lighting and post-processing properties in the game scene. It then stores all initial properties as $\mathbf{p}$, consisting of $\mathbf{p}_l$ for lighting and $\mathbf{p}_p$ for post-processing. These represent the initial vectors for lighting and post-processing, respectively.
(2) **Grid search.** The algorithm explores the property space for a total of N iterations. At the start of each iteration, the properties within $\mathbf{p}_l$ and $\mathbf{p}_p$ are randomly shuffled. Each iteration comprises a lighting optimization phase followed by a post-processing optimization phase. The grid search algorithm first optimizes the values of the lighting properties before optimizing the post-processing values. This ordering is deliberate, as the objective of post-processing is to enhance the lighting results.
(3) **Lighting optimization** At each iteration, the algorithm sequentially changes an element $p_{l,i}$ of vector $\mathbf{p}_l$ by adding a step value δs. The step value δs is determined by the grid resolution R, which can be set by the users. For instance, if the user sets R to 20, the δs of the hue value of the light would be 0.05, as the red value ranges from 0 to 1. After changing the value of $p_{l,i}$, the algorithm renders the scene and compares it with Y_{GT} using the MSE metric, searching for $\mathbf{p}_l\prime$ that satisfies $MSE_{\mathbf{p}_l} > MSE_{\mathbf{p}_l\prime}$. Whenever a value is

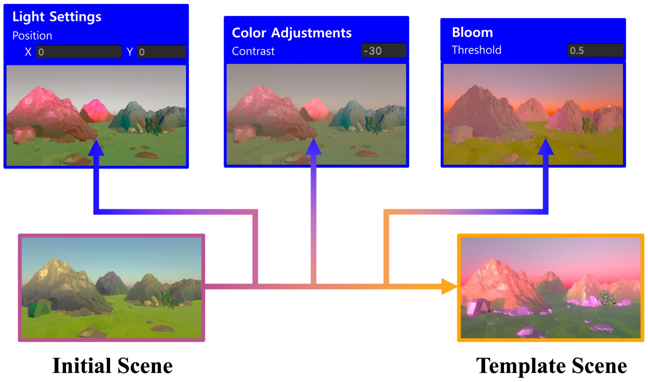


Figure 9: A brief schematic of the Recaller. The Recaller backtraces the steps of the Creator and depicts the changes in the scene caused by a change in a property.

changed, it is recorded in a stack memory to allow for the recalling of the property values.

(4) **Post-processing optimization** After optimizing the lighting, the algorithm proceeds to refine the values of the vector $\mathbf{p}_p$ using the same grid search algorithm. The updated vectors, $\mathbf{p}_l\prime$ and $\mathbf{p}_p\prime$, collectively form $\mathbf{p}\prime$, which denotes the entire updated property vector derived through the grid search algorithm.

(5) **Random traversal** Following the calculation of $\mathbf{p}\prime$, the algorithm compares it with $\mathbf{p}$. If $\mathbf{p}$ and $\mathbf{p}\prime$ are identical after the grid search, the $\mathbf{p}\prime$ vector either represents the global optimum or has converged to a local optimum. To validate the result, the algorithm randomly adds or subtracts $\frac{1}{2}\delta s$ to or from an element in the $\mathbf{p}\prime$ vector and repeats step (2) with additional m iterations. If the MSE decreases within m iterations, the algorithm continues running. If not, it reverts to the $\mathbf{p}\prime$ state. The algorithm terminates when it cannot find a better value after the random traversal or after N iterations.

Users can set the values for N and R. A higher N and R would enhance the resemblance between the template scene and reference image but with high computation time.

4.4 Recaller: Display of intermediate steps

The third component of LumiMood is the Recaller, which is designed to resolve Challenge3. This challenge arises owing to properties that are difficult to understand. Previous studies have shown that providing an intermediate step can increase the designer's control and understanding of the entire design workflow, thus improving the user experience [77]. After the Creator adjusts the lighting and post-processing, the Recaller retraces the Creator's design process and computes the effect of each property on the template scene design. Users can use this information to track the change in each property and create the desired template scene. Thus, the Recaller offers designers greater control and comprehension of the results of each property change. Figure 9 depicts the Recaller.

4.4.1 *Property visualizer.*

The Recaller automatically collects the properties related to lighting and post-processing and presents them in the property visualizer as a shortcut. The Recaller uses this property visualizer window to describe the design steps of the Creator, highlighting the changed lighting and post-processing properties in the intermediate design steps.

4.4.2 *Recall algorithm.*

After the Creator modifies the lighting and post-processing properties of the template scene in N iterations, the Recaller calculates and sorts the most influential properties. As shown in Figure 9, when the designer recalls the previous design steps, the Recaller demonstrates the changes in the scene in the order of the most influential properties. When the Creator adjusts the given scene, the lighting and post-processing properties are recorded for every iteration. Using this record, LumiMood calculates the impact score of each property on the adjusted template by calculating the average value of the changes in the MSE made by each property over N iterations.

$$Impact(p_i) = \frac{1}{N}\sum_{n=1}^{N}(MSE_{\mathbf{p}_{n-1}} - MSE_{\mathbf{p}_{n-1}+\mathbf{e}_i p_i}) \tag{2}$$

Equation 2 represents the extent to which each property has affected the appearance of a scene. In this equation, p_{n-1} is the property vector on the nth iteration and $\mathbf{e}_i$ is a unit vector with all zero elements except for the ith element, which is set to 1. Thus, $\mathbf{p}_{n-1} + \mathbf{e}_i pi$ represents the property vector at the $n-1$th iteration, except for the ith element, which corresponds to the value of the nth property. The system sorts each property according to its impact score. When the user revisits the design process of the algorithm, the system displays the results step-by-step, starting from the first element in the sorted properties.

5 USER STUDY

The user study consisted of two tasks to confirm the effectiveness of LumiMood: Task1 (T1), which requires replicating a scene, and Task2 (T2), which requires designing a mood. These tasks were inspired by a study conducted by Karlik et al. [29], which introduced a matching task for replicating the lighting of a given scene and an open-ended task for freely designing the lighting of a given scene. We compared two workflow conditions—the baseline workflow used in the formative study and the LumiMood workflow where designers use LumiMood for designing—through T1 and T2. The goal of T1 was to observe how LumiMood streamlines the baseline workflow (Challenge2) and provides a better understanding of the lighting and post-processing properties (Challenge3). T2 was designed to determine how well the designers were able to design their scenes to evoke the desired emotions in users (Challenge1). Through T1 and T2, we verified how LumiMood supported both professionals and novices in their mood design workflows. Figure 10 depicts the overall flow of the user study.

5.1 Participants

In the formative study, we discovered that the major challenges faced by game designers vary depending on their level of experience. Therefore, we conducted a user study with 40 game designers, consisting of 20 females and 20 males, including 20 professionals (11

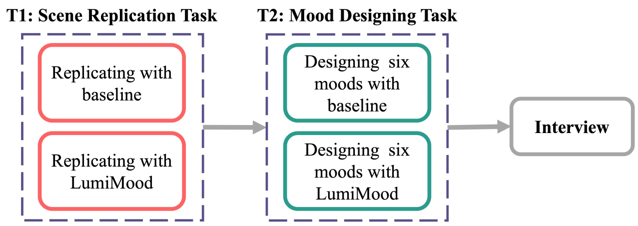


Figure 10: Flowchart of the user study. We conducted two tasks: scene replication task (T1) and mood designing task (T2).

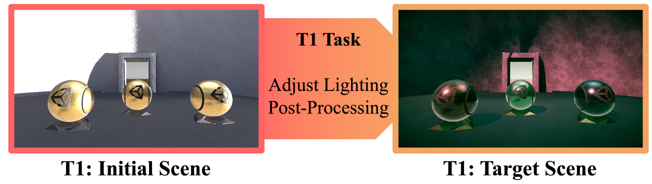


Figure 11: Task 1 involved replicating a scene. Participants were tasked with designing the mood of the initial scene to match the target scene, which was fixed. As a result, all designers had to design the initial scene in the same style to achieve this goal.

females and 9 males) and 20 novices (9 females and 11 males). None of the designers participating in the formative study was included in the user study and vice versa. The professional game designers had at least five years of experience in the game development industry, whereas novice game designers had less experience. Note that the criteria for professionals were the same as those for the high-experience group in the formative study, and the criteria for novices were the same as those for the low-experience group. All participants were required to have experience in using the Unity engine. The participants were remotely connected to a PC prepared by us in the lab, where they completed the design tasks. Before beginning the user study, the participants were asked to provide personal information, such as age, gender, and experience in game development.

5.2 T1: Scene Replication Task

The first task involved replicating a scene. The participants were instructed to design an initial scene (Figure 11-left) that closely resembled the target scene (Figure 11-right). T1 was designed to evaluate the effectiveness of LumiMood in resolving Challenge2, which relates to the labor-intensive baseline workflow, by streamlining the design process using LumiMood. T1 also addressed Challenge3, which arises from the ambiguity in selecting the properties to apply to the design. The designers created the scenes using both the baseline and LumiMood workflows. Thus, they designed the same scene twice. The order of the two workflows was counterbalanced to prevent learning effects. We evaluated the performance of each designer based on the design time and quality of the replication results. In T1, all designers were tasked with designing the same scene. This allowed for a direct comparison of the designers' workflows and understanding of lighting and post-processing properties when using the baseline workflow versus LumiMood. In the T1 questionnaire, the participants were asked about their perceived difficulty, satisfaction with the design outcome, and understanding of the properties.

5.2.1 ***Procedure***. The task T1 is depicted in Figure 11. The participants were shown an image of the target scene and asked to modify a given initial scene in Unity engine to closely resemble the target scene. The target scene was generated by randomly selecting the lighting and post-processing properties; therefore, the Generator in LumiMood was not used in this task. The procedure for T1 is as follows:

- Step1. The participants were informed about the task and LumiMood.
- Step2. The participants were given 10 min to familiarize themselves with the Unity project and UI of LumiMood.
- Step3. The participants designed the scene using two workflows —baseline and LumiMood—in counterbalanced order within 10 min.
- Step4. The participants answered a questionnaire shown in Table 4 after each task.

5.2.2 ***Questionnaire and interview***. After each workflow, the participants were asked to complete the questionnaire shown in Table 4. The questionnaire aimed to assess qualitative results, such as the perceived difficulty of the task (Item 1), satisfaction with the design outcome (Item 2), and level of understanding of lighting and post-processing properties (Item 3). After completing T1, we interviewed the participants to investigate how the use of LumiMood affected their perceived difficulty, satisfaction with the design outcome, and understanding of properties when compared with the baseline workflow.

Table 4: T1 questionnaire. Participants were asked about the perceived difficulty, satisfaction with the design outcome, and satisfaction with the workflow.

Item	Questionnaire	Answer Type
1	(Perceived Difficulty) It was challenging to modify the scene to resemble the target scene.	Likert (1-7)
2	(Outcome Satisfaction) I am satisfied with the quality of the design outcome that I created.	Likert (1-7)
3	(Property Understanding) I was able understand the effect of each property on the design outcome.	Likert (1-7)

5.2.3 ***T1: Results***. In T1, we observed the design precision, design time, and mouse/keyboard interaction counts as the quantitative measures. Design precision was measured by calculating the distance between the target scene and final designed scene. We

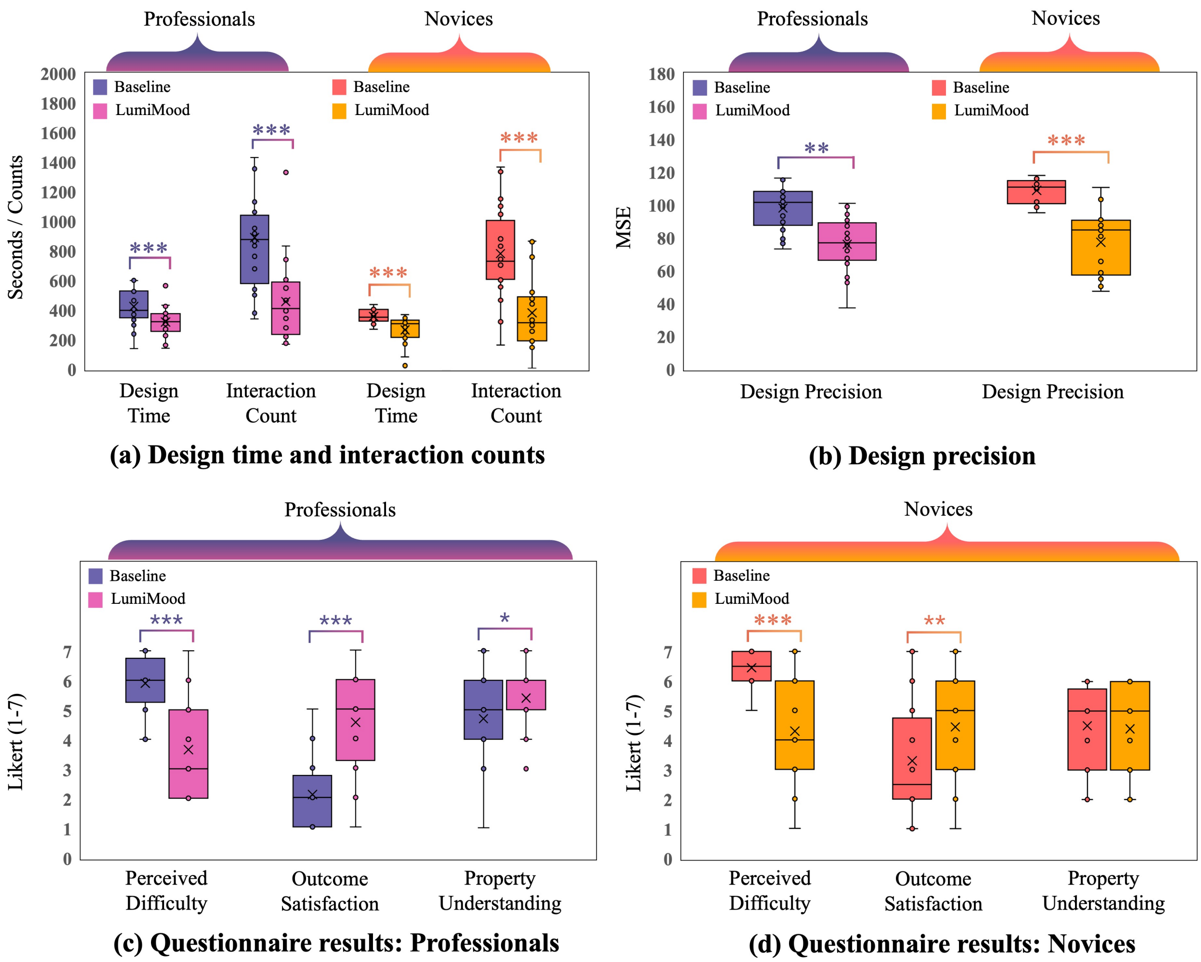


Figure 12: Questionnaire results for T1: (a) Quantitative results for design time (in seconds) and interaction counts. (b) Mean Squared Error (MSE) values of the design results. (c) Questionnaire results for perceived difficulty and satisfaction with design outcome and workflow among professionals. (d) Questionnaire results for perceived difficulty and satisfaction with design outcome and workflow among novices. Significance levels are indicated: $^{*}p$<0.05, $^{**}p$<0.01, $^{***}p$<0.001.

compared the performances and questionnaire results of the participants for the baseline and LumiMood workflows. The results are depicted in Figure 12 and the statistics are reported in Table 5.

Faster design, fewer interactions. Both professional and novice participants exhibited faster design times and fewer interaction counts. To assess the time performance, we measured the time used by the designers to complete the task, allowing a maximum time of 10 min. Because the design time and interaction counts passed the skewness and kurtosis normality tests [34], we performed a paired t-test to verify the effect of LumiMood in comparison with that of the baseline. The results are shown in Figure 12-(a). When compared with the baseline results, both professionals and novices showed a significant decrease in the design time and interactions counts. This demonstrates that LumiMood can enhance the design efficiency in reaching the final design outcome.

Higher precision on design results. Both professional and novice participants could realize a design closer to the target scene. The design accuracy was measured using the MSE metric between the target scene and final design outcome. We used a paired t-test to compare the distributions within each group. The result is shown in Figure 12-(b). When compared with the MSE for the baseline results, both professionals and novices showed a significant decrease in the MSE of the LumiMood results. Thus, LumiMood enhanced their ability to design scenes closely resembling the target scene, thereby improving the design performance of both groups. We also compared the MSE results between the professionals and novices using a two-way repeated measures ANOVA. Before using LumiMood, novices had a significantly higher MSE than professionals (p_{holm}=0.026). However, no significant differences were observed after using LumiMood (p_{holm}=0.715). Without LumiMood, the designs created by the novices were noticeably less similar to the target scene than those created by the professionals were. However,

Table 5: Statistical results of T1. Significance levels are indicated: *p<0.05, **p<0.01, ***p<0.001.

Fig12-(a)(b)	Professionals					
	Design Time		Interaction Count		Design Precision	
	Baseline	LumiMood	Baseline	LumiMood	Baseline	LumiMood
Statistics	M=423.13 SD=129.12	M=317.48 SD=100.79	M=891.05 SD=369.97	M=457.05 SD=281.59	M=97.93 SD=12.84	M=75.74 SD=16.24
Significance	t(19)=4.54, p<0.001***		t(19)=6.40, p<0.001***		t(19)=3.96, p=0.001**	
Fig12-(a)(b)	**Novices**					
	Design Time		Interaction Count		Design Precision	
	Baseline	LumiMood	Baseline	LumiMood	Baseline	LumiMood
Statistics	M=356.45 SD=43.93	M=265.94 SD=92.65	M=782.75 SD=309.12	M=379.90 SD=240.95	M=108.88 SD=7.34	M=77.28 SD=19.14
Significance	t(19)=5.05, p<0.001***		t(19)=6.94, p<0.001***		t(19)=6.58, p<0.001***	
Fig12-(c)	**Professionals**					
	Perceived Difficulty		Outcome Satisfaction		Property Understanding	
	Baseline	LumiMood	Baseline	LumiMood	Baseline	LumiMood
Statistics	Mdn=6.00, Mo=6.00 $Q1$=5.75, $Q3$=6.25	Mdn=3.00, Mo=2.00 $Q1$=2.00, $Q3$=5.00	Mdn=2.00, Mo=1.00 $Q1$=1.00, $Q3$=2.25	Mdn=5.00, Mo=5.00 $Q1$=3.75, $Q3$=6.00	Mdn=5.00, Mo=4.00 $Q1$=4.00, $Q3$=6.00	Mdn=6.00, Mo=5.00 $Q1$=5.00, $Q3$=6.00
Significance	Z=-3.38, p<0.001***		Z=3.55, p<0.001***		Z=2.05, p=0.042*	
Fig12-(d)	**Novices**					
	Perceived Difficulty		Outcome Satisfaction		Property Understanding	
	Baseline	LumiMood	Baseline	LumiMood	Baseline	LumiMood
Statistics	Mdn=6.50, Mo=3.00 $Q1$=6.00, $Q3$=7.00	Mdn=4.00, Mo=3.00 $Q1$=3.00, $Q3$=6.00	Mdn=2.50, Mo=2.00 $Q1$=2.00, $Q3$=4.25	Mdn=5.00, Mo=5.00 $Q1$=3.00, $Q3$=6.00	Mdn=5.00, Mo=5.00 $Q1$=3.00, $Q3$=5.25	Mdn=5.00, Mo=6.00 $Q1$=3.00, $Q3$=6.00
Significance	Z=-3.50, p<0.001***		Z=2.82, p=0.005**		Z=-0.35, p=0.749	
Fig13	**Professionals**		**Novices**			
	Modified Properties		Modified Properties			
	Baseline	LumiMood	Baseline	LumiMood		
Statistics	M=20.15 SD=4.67	M=22.80 SD=5.68	M=18.55 SD=6.08	M=20.05 SD=5.21		
Significance	t(19)=-3.46, p=0.003**		t(19)=-1.58, p=0.130			

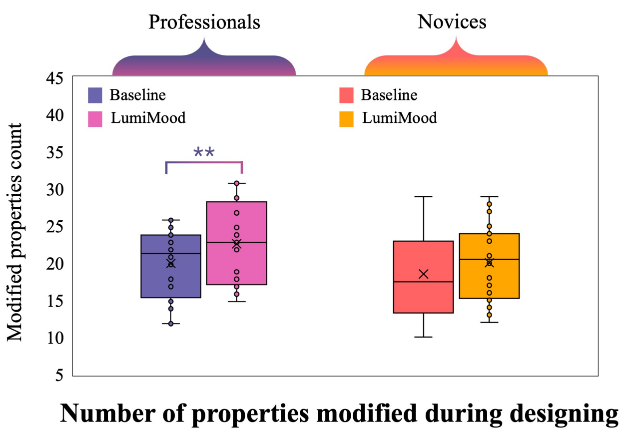


Figure 13: Number of properties modified by the designers to create the final design outcome; *p<0.05, **p<0.01, ***p<0.001.

LumiMood enabled them to design at a level similar to that of professionals. This indicates that using the LumiMood workflow can benefit both professionals and novices in their design performance, specifically in terms of precision when referencing an image.

Lower difficulty, higher outcome satisfaction. Both professional and novice participants reported that designing with LumiMood was easier and the design outcomes were more satisfactory. We collected data on the perceived task difficulty, design outcome satisfaction, and understanding of a property using a questionnaire. We conducted a Wilcoxon signed-rank test to observe the effect of LumiMood on the perceived experience of the designers. The results of the professionals are shown in Figure 12-(c), whereas those of novices are shown in Figure 12-(d). When professionals used LumiMood, we observed significant improvements in the perceived difficulty, outcome satisfaction, and understanding of design properties when compared with the baseline results. We observed similar trends for novices, where the results of LumiMood exhibited significantly improved perceived difficulty and outcome satisfaction when compared with the baseline results. We found no significant difference in the understanding of a property between the baseline and LumiMood. These results indicate that LumiMood reduced the perceived task difficulty for both professionals and novices while increasing their satisfaction with the design outcome and workflow. Only professionals benefited from the intermediate step information provided to better understand the changes corresponding to each property.

5.2.4 ***T1: Further observations.*** The results from T1 indicated that professional designers demonstrated enhanced understanding of lighting and post-processing properties when using LumiMood. No significant difference was found regarding the understanding of the properties in the questionnaire results of the novices. Therefore, we analyzed the design processes of both professionals and novices to observe their interpretation of the properties. Specifically, we examined the number of properties actively manipulated by the designers while creating their scenes. Because the target scene in T1 was fixed, we could compare the properties utilized before and after using LumiMood.

Learning how to use the properties. After using LumiMood, we observed that professionals were able to utilize a greater number of properties to achieve their goals. We observed the number of properties that designers used to create their final design outcomes. For the baseline workflow, we counted the properties changed at least once during the design process. For the LumiMood workflow, we excluded cases where a property modified by the Creator was later reset to its initial values, as it cannot be assumed that the designer understands the property well enough to create a new design. The results are shown in Figure 13. When compared with the baseline workflow, professionals tended to adjust a greater number of properties when using LumiMood. For novices, no significant difference was found in the number of properties used with LumiMood. This indicates that by using the Recaller, professionals were able to utilize a greater number of properties to achieve a design resembling the given target scene. This aligns with the questionnaire results, as novices reported no change in their understanding of the properties.

5.3 T2: Mood designing task

The second task was the mood designing task. In this task, the participants were instructed to freely design the mood of an initial scene to evoke the target emotion in the game players. The goal of this experiment was to observe how LumiMood resolves Challenge1, where designers had difficulty envisioning the abstract concepts of moods that evoke certain emotions in game players.

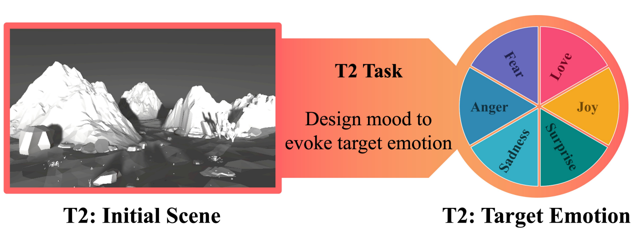


Figure 14: In task T2, participants were instructed to design the mood of the initial scene to evoke a specific emotional response from viewers.

5.3.1 ***Procedure.*** In this task, the participants were asked to design the mood of the initial scene depicted in Figure 14. The goal was to eventually evoke the given emotion keyword in viewers. Six emotion keywords based on Parrott's model were provided, and the participants conducted mood designing 12 times (6 emotion keywords × 2 workflows). The tasks were presented in a balanced Latin square order, with each task lasting up to 10 min. The participants were allowed to search the internet during this task.

- Step1. The participants were informed about the task and LumiMood.
- Step2. The participants were given 10 min to familiarize themselves with the Unity project and UI of LumiMood. The Unity project was different from that used in T1.
- Step3. Six emotion keywords were presented and the participants were asked to design the appropriate mood to evoke the given emotion.
- Step4. The participants answered a questionnaire shown in Table 6 after completing each task.

5.3.2 ***Questionnaire and interview.*** After each task, the participants completed a questionnaire designed to obtain qualitative results. The questionnaire included items such as satisfaction with Generator (Item1), Creator (Item2), and Recaller (Item3), as well as self-evaluation of the design outcome and workflow satisfaction (Item4 and Item5). Items 1, 2, and 3 were not asked when the participants used the baseline workflow. After completing T2, we interviewed the participants to obtain their opinions regarding LumiMood in comparison with the baseline workflow. Furthermore, we requested that the participants rank the Generator, Creator, and Recaller in the order of their preference.

5.3.3 ***T2: Results.*** We analyzed how LumiMood resolved Challenge1, which arises when translating the abstract concepts of moods into concrete visual representations through the observation of qualitative measures. The results are depicted in Figure 16 and the statistics are reported in Table 7.

Performance specifications of LumiMood workflow. From the template creation algorithm discussed in section 4.3.1, we configured the number of iterations N to 4 and the search resolution R to 10 to expedite template creation. On a PC with an i7-10700 CPU and RTX 2080 GPU, the average time for the Generator to produce the reference image was 9.00 s (SD=1.93), whereas the Creator required 19.85 s (SD=1.80) to return the template scene. As the Creator relies on a grid search algorithm with randomness, it may occasionally create template scenes that deviate from the reference images (e.g., the algorithm may diverge, resulting in a completely white scene). After each use of LumiMood, we posed a simple question to the designers: whether the final template scene created by LumiMood appeared similar to the reference image. Out of 240 LumiMood design cases (40 designers × 6 emotions), the template scene resembled the reference image in 211 cases (87.92%). Figure 15 illustrates the template creation process in LumiMood. In each iteration, the grid search algorithm undergoes lighting optimization and post-processing optimization sequentially.

Table 6: T2 questionnaire. The designers were asked to rate their satisfaction with using LumiMood as well as their satisfaction with the design outcome and workflow.

Item	Questionnaire	Answer Type
1 (LumiMood only)	(Generator satisfaction) The synthesized reference image expresses the intended emotion well.	Likert (1-7)
2 (LumiMood only)	(Creator satisfaction) Lighting and post-processing adjusted by the Creator express the intended emotions well.	Likert (1-7)
3 (LumiMood only)	(Recaller satisfaction) By recalling the design steps, I was able to understand how each lighting and post-processing property has effected the design outcome.	Likert (1-7)
4	(Outcome satisfaction) I think the final outcome conveys the intended emotion well.	Likert (1-7)
5	(Workflow satisfaction) I am satisfied with the design workflow.	Likert (1-7)

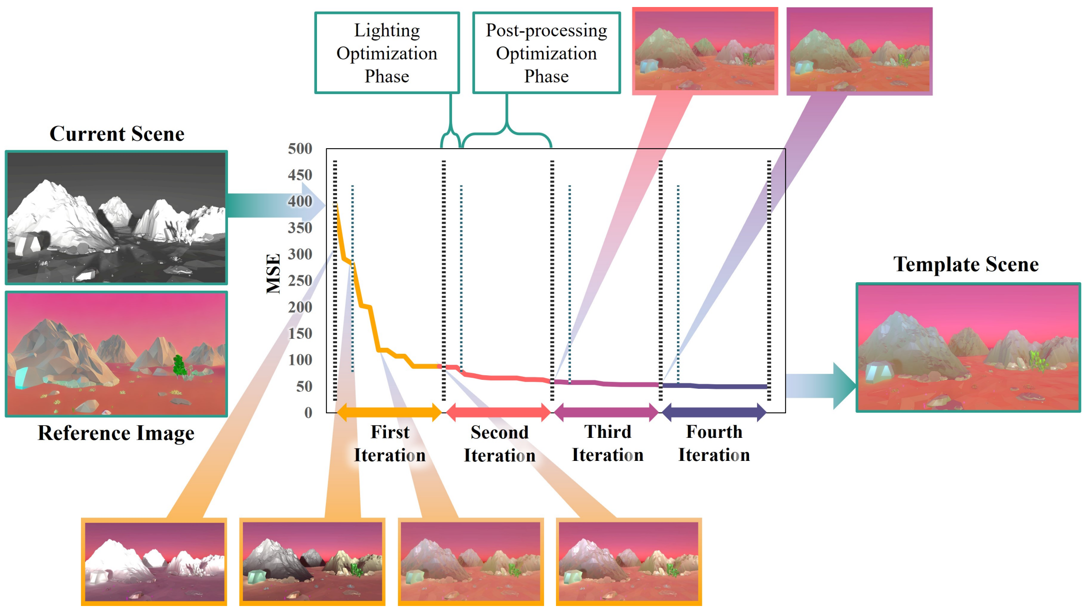
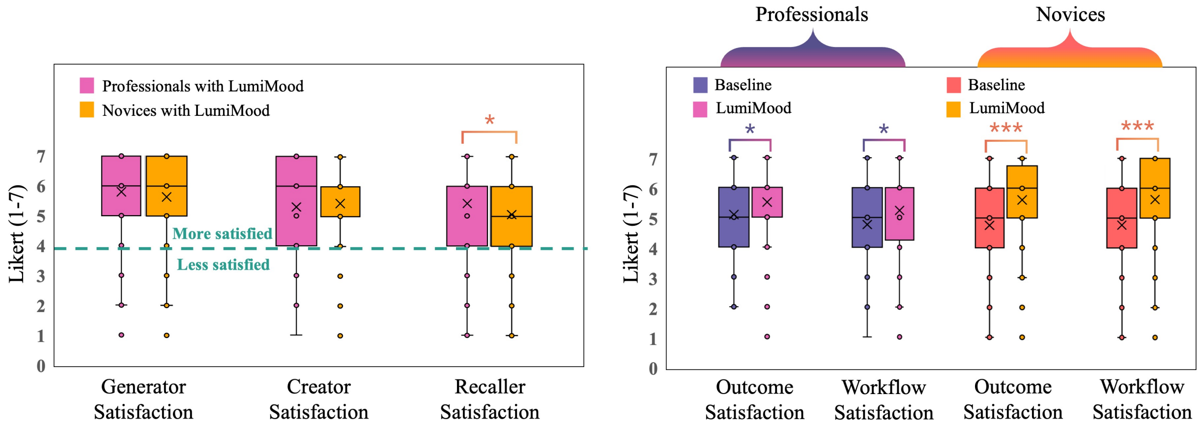


Figure 15: Change in aesthetics of the scene during the grid search iterations. In T2, we set the total number of iterations N to 4 and number of search steps to 10 for fast generation of the template scene.

High satisfaction towards Generator, Creator, and Recaller. Both the professional and novice participants expressed high satisfaction with the three components of LumiMood. The designers were asked to rate LumiMood based on the appropriateness of the results. Figure 16-(a) shows the satisfaction regarding the Generator, Creator, and Recaller. Both professionals and novices assigned high scores for satisfaction with the Generator, Creator, and Recaller. Hence, we observed that the designers in T2 were satisfied with the functionalities of LumiMood. The Mann–Whitney U test revealed a significant difference in Recaller satisfaction between novices and professionals. This indicates that providing intermediate designs was more helpful to professionals than to novices.

We also examined the correlation between the quality of Generator, Creator, and Recaller, and workflow satisfaction in T2. Pearson's r was calculated for the correlation analysis, and the results are presented in Table 8. We found that for professionals, the Creator had the highest correlation with workflow satisfaction, followed by the Generator and Recaller. A similar tendency was observed for novices, but the Recaller showed no significant correlation with workflow satisfaction. Therefore, we identified that the use of design templates is greatly associated with an increase in workflow satisfaction.

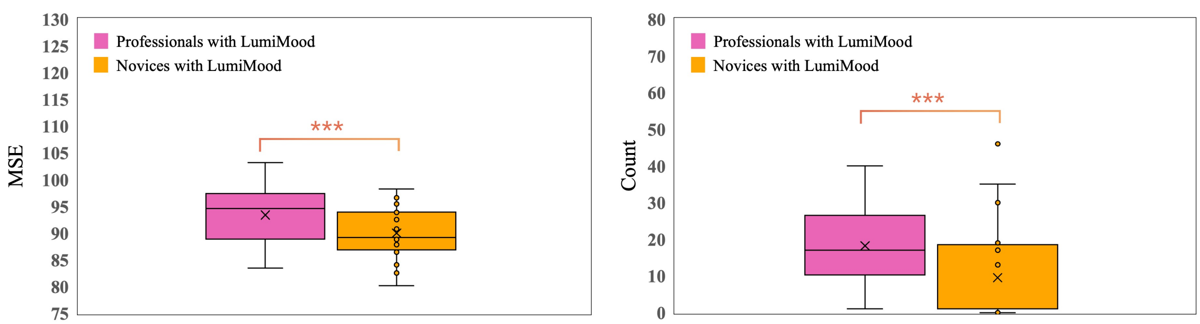
Greater satisfaction with design outcome and workflow. The satisfaction with the outcome and workflow were enhanced after using LumiMood for both professionals and novices. Figure 16-(b) shows the results of the satisfaction with the design outcome and workflow. We conducted a Wilcoxon signed-rank test to compare the results before and after using LumiMood. When professionals used LumiMood, we observed a significant improvement in the outcome satisfaction when compared with the baseline results. We observed similar trends for novices; the use of LumiMood significantly improved the outcome satisfaction when compared



(a) Generator, Creator, Recaller satisfaction

(b) Design outcome, workflow satisfaction

Figure 16: Questionnaire results from T2. (a) Satisfaction with Generator, Creator, and Recaller. (b) Satisfaction with design outcome and workflow. The error bar represents the standard deviation. Significance levels are indicated: *p<0.05, **p<0.01, ***p<0.001.



(a) Difference between design outcome and template

(b) Number of times Recaller was used

Figure 17: Differences between professionals and novices utilizing the template scene. (a) MSE between final design outcome and template scene. (b) Number of times Recaller was used while designing. Significance levels are indicated: *p<0.05, **p<0.01, ***p<0.001.

with the baseline results. Therefore, we concluded that both professionals and novices were able to produce satisfying outcomes with LumiMood, along with improved workflow satisfaction.

5.3.4 *T2: Further observations.*

In T2, the designers had the freedom to create their own designs, resulting in a broader range of design outcomes. They could either use the template scene as is or create their own design using the template scenes. The results of T2 showed that all three components of LumiMood influenced the overall workflow satisfaction for both professionals and novices while showing notable differences in the satisfaction with the Recaller. Consequently, we observed variations in the design process between professionals and novices when incorporating the template scene into their workflows.

Utilizing the template scenes for designing. The novices relied more on the template scenes, whereas the professionals actively used the intermediate steps to create their own designs. We examined how designers incorporated the suggested template scene into their final design outcome. To assess this, we measured the difference between the appearance of the suggested template scene and the final design outcome by using the MSE metric. The independent t-test revealed that the difference was significantly higher for the professionals than for the novices. Professionals developed their designs further, resulting in a larger difference between the template scene and final outcome. This indicates that novices are more likely to accept the design suggested by LumiMood, whereas professionals tend to further refine their scene using the template scene.

The interaction counts of the Recaller's intermediate steps also suggest that novices rely more on the template result than professionals do. The intermediate steps provided by the Recaller enable designers to trace back the design steps of LumiMood and start

Table 7: Statistical results of T1. Significance levels are indicated: *p<0.05, **p<0.01, ***p<0.001.

Fig16-(a)	**Generator Satisfaction**		**Creator Satisfaction**		**Recaller Satisfaction**	
	Professionals with LumiMood	Novices with LumiMood	Professionals with LumiMood	Novices with LumiMood	Professionals with LumiMood	Novices with LumiMood
Statistics	Mdn=6.00, Mo=7.00 $Q1$=5.00, $Q3$=7.00	Mdn=6.00, Mo=7.00 $Q1$=5.00, $Q3$=7.00	Mdn=6.00, Mo=6.00 $Q1$=4.00, $Q3$=7.00	Mdn=6.00, Mo=6.00 $Q1$=5.00, $Q3$=6.00	Mdn=6.00, Mo=6.00 $Q1$=4.00, $Q3$=6.00	Mdn=5.00, Mo=4.00 $Q1$=6.00, $Q3$=6.00
Significance	W=10098.00, p=0.314		W=10834.50, p=0.962		W=9163.50, p=0.021*	
Fig16-(b)	**Professionals**					
	Outcome Satisfaction		Workflow Satisfaction			
	Baseline	LumiMood	Baseline	LumiMood		
Statistics	Mdn=5.00, Mo=6.00 $Q1$=4.00, $Q3$=6.00	Mdn=6.00, Mo=6.00 $Q1$=5.00, $Q3$=6.00	Mdn=5.00, Mo=5.00 $Q1$=4.00, $Q3$=6.00	Mdn=6.00, Mo=6.00 $Q1$=4.75, $Q3$=6.00		
Significance	Z=2.50, p=0.011*		Z=2.56, p=0.010*			
Fig16-(b)	**Novices**					
	Outcome Satisfaction		Workflow Satisfaction			
	Baseline	LumiMood	Baseline	LumiMood		
Statistics	Mdn=5.00, Mo=6.00 $Q1$=4.00, $Q3$=6.00	Mdn=6.00, Mo=6.00 $Q1$=5.00, $Q3$=6.00	Mdn=5.00, Mo=5.00 $Q1$=4.00, $Q3$=6.00	Mdn=6.00, Mo=7.00 $Q1$=5.00, $Q3$=7.00		
Significance	Z=4.41, p<0.001***		Z=4.24, p<0.001***			
Fig17	**Outcome-template difference**		**Number of Recaller used**			
	Professionals with LumiMood	Novices with LumiMood	Professionals with LumiMood	Novices with LumiMood		
Statistics	M=93.37 SD=5.37	M=89.97 SD=4.86	M=18.05 SD=11.28	M=9.50 SD=13.89		
Significance	t(38)=-2.10, p=0.042*		t(38)=-2.14, p=0.039*			

Table 8: Correlation between the perceived quality of Generator, Creator, and Recaller and workflow satisfaction.

	Professionals		Novices	
	Pearson's r	p	Pearson's r	p
Generator	0.459	<0.001	0.450	<0.001
Creator	0.605	<0.001	0.556	<0.001
Recaller	0.453	<0.001	0.153	0.062

their own design from an intermediate step. This offers the freedom to either use the final template design from the Creator or customize it according to their needs. We used the independent t-test to examine the interaction counts of the designers with the intermediate steps and found that novices had significantly fewer interaction counts than that of professionals. This suggests that professionals tend to customize the template scene by examining the changed properties, whereas novices tend to pay less attention to customizing the template scene.

6 DISCUSSION

This section reports the findings and discusses the results of the user study. Based on the results of the user study, we conducted interviews to find the implications for designing AI-driven CSTs.

6.1 Implications of the results of T1 and T2

In this study, we conducted T1 and T2 to examine whether incorporating LumiMood into the designers' workflows benefits them in creating the mood of a scene. In T1, all participants were tasked with replicating the same target scene. The purpose of this task was to determine if LumiMood assists designers in addressing Challenge2 and Challenge3. In T2, the participants experienced the full LumiMood workflow, which included mood ideation using the Generator, to assess how designers manage Challenge1 and observe how LumiMood had affected their entire workflow satisfaction.

6.1.1 Resolving Challenge1: Bridging the abstract concepts of mood and concrete visual representations. We conducted T2 to assess if LumiMood effectively addressed Challenge1, which arises from the abstractness of mood concepts. The results of T2 indicated that the participants in the user study were satisfied with the reference image and their workflows. In the interview, both professionals and novices mentioned that they found inspiration for mood designs from the reference image. Specifically, professionals relied on the prompting capability to guide the synthesized results according to their intentions, whereas novices focused on the recommendation system itself; *"By prompting with keywords, I was able to see various mood designs based on the current scene. This inspired me significantly" (Professional8)*, *"Before starting the actual design, I had to think of specific designs. LumiMood inspired me by recommending images" (Novice1)*. Previous studies have also noted that providing example images for designing tasks enhances the performance of various tasks, such as interface designing [39] or 3D modeling [40]. Based on these studies, we concluded that the reference images generated by the Generator inspired the designers by bridging the gap between abstract mood concepts and concrete visual representations.

6.1.2 Resolving Challenge2: Connecting visual representations to property values. Challenge2 arose when designers found it difficult to translate the visual representations into specific lighting and post-processing property values. This often led to a trial-and-error approach, as the property values could not be directly deduced from the imagery. The results indicated that LumiMood streamlined the design process, enhancing precision and satisfaction while reducing the design time and interaction counts. Interviews revealed that professionals appreciated the streamlined workflow, whereas novices particularly valued the reduced difficulty; *"The designing was much faster when using LumiMood, which was the most satisfactory aspect" (Professional3)*, *"It was easier to design with LumiMood, as it provided the template designs that I wanted" (Novice17)*. Based on these observations, we conclude that using LumiMood can help designers achieve improved precision in lesser time and with fewer interactions, thus streamlining the design process and resolving Challenge2.

6.1.3 Resolving Challenge3: Understanding the effects of each property on the design outcome for professionals. In the formative study, designers noted that understanding the properties of lighting and post-processing posed difficulties, which were denoted as Challenge3. The results indicate that professionals using LumiMood gained a better understanding of the property settings in game scenes, whereas there was no notable difference in the understanding of the property settings between the baseline workflow and LumiMood among novices. LumiMood enabled the professional designers to broaden their knowledge of properties, thereby diversifying their mood design techniques; *"I was able to understand split toning by recalling the property changes" (Professional 12)*, *"I think using LumiMood would be great for practicing lighting and post-processing" (Professional 13)*, and *"During the experiment, I was able to adjust some properties with confidence" (Professional17)*. Through the interview results, we confirmed a difference in the reaction to the Recaller, a tool designed to enhance the understanding of the properties, between professionals and novices. Professionals reported that Recaller helped them understand the specific properties that were changed to achieve the current design. However, novices tended to use Recaller less frequently than the professionals did; *"I simply accepted the template created by the system. I really did not use the Recaller often to see how the properties have changed" (Novice7)*, *"I was satisfied with the template results; so, I did not feel the need to recall" (Novice11)*. As novices used the Recaller less frequently than professionals did, the resolution effect for Challenge3 was less significant for them.

6.2 Designing with templates

In this study, we designed LumiMood to provide design template scenes for designers to improve their workflow efficiency. The results indicate that using the templates reduces the design time and interaction counts while increasing the design precision and mood intention accuracy. To identify the component—Generator, Creator, or Recaller—that had the most effect on the user's satisfaction with the entire workflow, we examined the correlation between the quality of each component and workflow satisfaction in T2. The results indicated that for both professionals and novices, the Creator exhibited the most significant correlation with workflow satisfaction when compared with those of the Generator and Recaller. Hence, we explored how initiating the design process from the template scene benefited the designers.

6.2.1 Advantages of template based designing. After conducting the interviews, we examined the participants' opinions on using the template scenes. While highlighting the streamlining effect of template-based designing, the novices noted that they were better able to evaluate their designs after using the templates. In comparison with the baseline, where designers have to design from scratch, templates enable the designers to start from the base designs that are closer to the desired outcome. As a result, the novices reported that evaluating the designs when iterating with properties became easier; *"When designing using a template, I was able to better understand the strengths and weaknesses of my design" (Novice12)*. This possibly increased the satisfaction with the entire workflow. The professionals also reported that using the templates simplified their work by embodying the concepts of abstract moods that are difficult to imagine. *"The templates suggested by the Creator simplified the entire process significantly by turning the abstract concepts into actual property values" (Professional3)*. Thus, providing base designs allows designers to create their scenes with lesser difficulty.

6.2.2 Implications for template generation. Although the participants expressed satisfaction with the template scenes, we observed that AI-driven CSTs should be able to synthesize a wide variety of references for inspiration. In an interview after T2, the participants noted that the synthesized reference images should be more flexible to accommodate various designer prompts and generate diverse template scenes; *"Overall, I like the idea of AI designing for me, but I am concerned that excessive reliance on AI will result in similar designs by different designers" (Professional20)*. Although providing reference images can improve both design satisfaction and performance, previous studies have pointed out that design fixation, where designers over-rely on a single satisfactory design, may also occur [11]. This fixation can limit the creativity of designers; hence, preventing it is essential for ensuring the quality of the design outcome. To address fixations with CST, Liu et al. proposed the use of diffusion models, which can be combined with prompting capabilities to solve the fixation problem [40] by synthesizing various images that can inspire designers. This study utilized a diffusion model to provide references for 3D modeling and identified that designers were able to overcome fixation with DALL-E, a well-generalized diffusion model. Therefore, we suggest that the generalizability of the AI model for synthesizing desired images is a key factor in building CSTs for inspiring the designers' creativity. Because designers are sensitive to fixation issues [56], they should be provided with a wide variety of synthesized results when designing the CSTs. A biased AI model may force the designers to focus on a single image. Hence, training an AI model with diverse image synthesis capabilities is required for AI-driven CSTs.

6.3 Professionals prefer control, novices prefer automation

We conducted a user study to assess whether our system can assist both professional and novice game designers. To gain further insights into the use of a CST across both groups, we examined the interview results to analyze the interactions of professional and novice designers with our system, along with differences between their performances.

During the interview, the designers were asked to discuss their experience with LumiMood and rank the functionalities of the Generator, Creator, and Recaller. The novice designers found the template generation functionality, which automatically sets the lighting and post-processing properties, particularly impressive. They found it to be the most advantageous feature of LumiMood, as it allowed them to bypass the initial design stage; *"It was convenient and time-saving to start with templates generated by the system" (Novice5), "As a beginner, I liked the part where AI provides basic directions for designing" (Novice16).* In contrast, professionals preferred to retain control over the automation. The interview revealed that professionals enjoyed the process of creating a scene that reflected their imagination; *"Setting up a lighting property was the most enjoyable part of scene designing. Controlling the properties and visualizing the intermediate steps were the features that I liked." (Professional20).*

Previous studies have shown that guaranteeing user control over the tasks increases their preference for and use of CSTs, suggesting that an end-to-end system can hinder usability by excluding the user"s role in the entire design workflow [77]. However, the existing studies have typically gathered insights from participants with a single level of expertise. Therefore, we provide guidance on leveraging automation based on the designers' experience levels. As our system supports end-to-end automation of the entire design process, the participants were allowed to carry out their design task according to their preference, with flexible operation ranging from fully automated to fully controlled. We found that novices relied more on automation, whereas professionals followed a more involved process of designing rather than relying solely on automation. Thus, the preference for the balance between automation and control may differ among designers, especially depending on their experience levels. CSTs can offer a high level of control to experienced users, whereas they should automate the workflow and may even offer an end-to-end workflow to users with low experience levels.

7 LIMITATIONS AND IMPLICATIONS FOR BETTER AI-DRIVEN CST DESIGN

LumiMood supports designers' workflows by suggesting lighting and post-processing designs. Although the results indicate that LumiMood enhances satisfaction with the design workflow, several improvements can be made. In this section, we suggest exploring the directions for designing better CSTs for future works.

7.1 Improvements for image generation

In this study, a diffusion model was employed to create reference images that inspire designers during the conceptualization phase of their design. However, the use of the diffusion model presents several areas for improvement in future research. First, the diffusion model used in our study is a "black box" [1], which makes it difficult to explain how the Generator's reference affective images are synthesized. Some participants noted that they wanted direct explanations on how the diffusion model in the Generator created the reference affective image; *"I am curious about the aspect of the reference image that the AI believed people would feel sad about." (Novice4).* In future studies, explainable AI (XAI) [22, 70] features could be integrated into LumiMood to provide information about the AI's decision-making process. One possible solution is to introduce a classifier concatenated with a class activation map [62] to visualize which pixels in the reference image had the most significant effect on the AI's decision. By incorporating XAI capabilities into AI-driven CSTs, user experience can be enhanced by providing rich information about the system's decision-making.

Second, the diffusion model employed in this study generates the reference images based on a single image captured by a virtual camera. Consequently, in the user study, we evaluated the performance of LumiMood by having participants design a single scene. However, commercial games typically comprise multiple scenes, each requiring the design of a suitable mood. If a designer aims to achieve similar mood designs across different scenes, the Generator needs to be extended to support image generation that ensures consistent styles across various images. This enhancement can be realized by ensuring consistency within the diffusion model [69, 73]. By considering the spatial and temporal similarities between the images, these diffusion models aim to preserve the similarity in styles over multiple synthesis requests. The incorporation of such mechanisms into the image generation process of the Generator can support multiple reference image generations with style consistency.

7.2 Extending the automation to other visual effects

LumiMood focuses on low-level elements, such as color distribution, contrast, and brightness, to design the mood of a scene. A formative study has confirmed that these elements, along with lighting and post-processing, play a dominant role in determining a scene's mood. Lighting and post-processing significantly influence a scene's color palette, but they are part of a wider array of contributing factors. For instance, shaders are essential for simulating the interaction of light with material surfaces, which affects properties such as reflection, refraction, and transparency [63]. Beyond their technical functionality, shaders can also be used to create artistic styles, such as cell shading [66]. This is particularly important when attempting to replicate artistic styles in reference images that are challenging to achieve with realistic light physics alone. Because we found that matching the appearance of a template scene with a reference image is challenging, utilizing shaders can improve the overall design process.

In the user study, some designers recommended the mesh editing functionality, as a greater diversity of designs could be achieved if designers could change the semantics; *"I would like to change the 3D models inside the scene in addition to lighting and post-processing" (Professional14).* For example, in T2, a scene of a mountain with few trees and rocks was used. The designers would have had greater freedom to design the scene if they were allowed to change the height of the mountains or add more trees. Therefore, high-level features, such as the semantic meanings of objects, can be incorporated by introducing additional 3D generative models such as Unity Muse, SceneSeer [8], or GET3D [19]. This extension would provide designers with greater freedom to design and evaluate their work. In future research on the development of AI-driven CSTs for 3D designing, both high-level semantics and low-level color features should be considered to facilitate the realization of the scene conceptualized by the designers.

8 CONCLUSION

We present LumiMood, an AI-driven CST that supports the designers in creating the mood of their 3D scenes by automatically adjusting the lighting and post-processing properties. Through the formative study, we observed that designers are facing challenges of realizing the abstract mood concepts, labor-intensive iterative designing process, and confusion in selecting the properties of lighting and post-processing. The user study involving professionals and novice designers showed that LumiMood resolves these challenges with increased design precision, mood intention accuracy, and workflow satisfaction. During our user study, we explored the differences between professional and novice designers, with particular focus on their preference regarding control. While the design automation has to be cared to possess variety of references for all designers, the balance between automation and control has to be set based on the target users' experience levels. Hence, this study proposes the level of expertise as an important factor for future CST tools.

ACKNOWLEDGMENTS

This work was supported by the GIST-MIT Research Collaboration grant funded by the GIST in 2024.

REFERENCES

[1] Amina Adadi and Mohammed Berrada. 2018. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE access* 6 (2018), 52138–52160.
[2] Saeema Ahmed, Ken M Wallace, and Lucienne T Blessing. 2003. Understanding the differences between how novice and experienced designers approach design tasks. *Research in engineering design* 14 (2003), 1–11.
[3] Marie Altenburg. 2016. The Adobe Photoshop Lightroom CC Book for Digital Photographers. *PSA Journal* 82, 4 (2016), 8–9.
[4] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–13.
[5] Chris Bateman. 2011. *Imaginary games*. John Hunt Publishing.
[6] Jeremy Birn. 2014. *Digital lighting & rendering*. Pearson Education.
[7] Bruce A Block. 2001. The visual story: seeing the structure of film, TV, and new media. (2001).
[8] Angel X Chang, Mihail Eric, Manolis Savva, and Christopher D Manning. 2017. SceneSeer: 3D scene design with natural language. *arXiv preprint arXiv:1703.00050* (2017).
[9] Siddhartha Chaudhuri, Evangelos Kalogerakis, Stephen Giguere, and Thomas Funkhouser. 2013. Attribit: content creation with semantic attributes. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*. 193–202.
[10] Victoria Clarke and Virginia Braun. 2021. Thematic analysis: a practical guide. *Thematic Analysis* (2021), 1–100.
[11] Nathan Crilly. 2015. Fixation and creativity in concept development: The attitudes and practices of expert designers. *Design studies* 38 (2015), 54–91.
[12] Nigel Cross. 2004. Expertise in design: an overview. *Design studies* 25, 5 (2004), 427–441.
[13] Tim Edensor. 2015. Light design and atmosphere. *Visual Communication* 14, 3 (2015), 331–350.
[14] Paul Ekman. 1992. An argument for basic emotions. *Cognition & emotion* 6, 3-4 (1992), 169–200.
[15] Magy Seif El-Nasr and Ian Horswill. 2004. Automating lighting design for interactive entertainment. *Computers in Entertainment (CIE)* 2, 2 (2004), 15–15.
[16] Andrew J Elliot and Markus A Maier. 2014. Color psychology: Effects of perceiving color on psychological functioning in humans. *Annual review of psychology* 65 (2014), 95–120.
[17] Anna Felnhofer, Oswald D Kothgassner, Mareike Schmidt, Anna-Katharina Heinzle, Leon Beutl, Helmut Hlavacs, and Ilse Kryspin-Exner. 2015. Is virtual reality emotionally arousing? Investigating five emotion inducing virtual park scenarios. *International journal of human-computer studies* 82 (2015), 48–56.
[18] Kevin Frans, Lisa Soros, and Olaf Witkowski. 2022. Clipdraw: Exploring text-to-drawing synthesis through language-image encoders. *Advances in Neural Information Processing Systems* 35 (2022), 5207–5218.
[19] Jun Gao, Tianchang Shen, Zian Wang, Wenzheng Chen, Kangxue Yin, Daiqing Li, Or Litany, Zan Gojcic, and Sanja Fidler. 2022. Get3d: A generative model of high quality 3d textured shapes learned from images. *Advances In Neural Information Processing Systems* 35 (2022), 31841–31854.
[20] Sarra Graja, Phil Lopes, and Guillaume Chanel. 2020. Impact of visual and sound orchestration on physiological arousal and tension in a horror game. *IEEE Transactions on Games* 13, 3 (2020), 287–299.
[21] Elif Güneş and Nilgün Olguntürk. 2020. Color-emotion associations in interiors. *Color Research & Application* 45, 1 (2020), 129–141.
[22] David Gunning, Mark Stefik, Jaesik Choi, Timothy Miller, Simone Stumpf, and Guang-Zhong Yang. 2019. XAI—Explainable artificial intelligence. *Science robotics* 4, 37 (2019), eaay7120.
[23] Negin Hamzeheinejad, Daniel Roth, Daniel Götz, Franz Weilbach, and Marc Erich Latoschik. 2019. Physiological effectivity and user experience of immersive gait rehabilitation. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 1421–1429.
[24] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems* 33 (2020), 6840–6851.
[25] Ziqi Huang, Kelvin CK Chan, Yuming Jiang, and Ziwei Liu. 2023. Collaborative diffusion for multi-modal face generation and editing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6080–6090.
[26] Youngseung Jeon, Seungwan Jin, Patrick C Shih, and Kyungsik Han. 2021. 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*. 1–18.
[27] Eui Jun Jeong, Frank A Biocca, and Corey J Bohil. 2012. Sensory realism and mediated aggression in video games. *Computers in human behavior* 28, 5 (2012), 1840–1848.
[28] Ellen Jiang, Kristen Olson, Edwin Toh, Alejandra Molina, Aaron Donsbach, Michael Terry, and Carrie J Cai. 2022. Promptmaker: Prompt-based prototyping with large language models. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–8.
[29] Ondrej Karlik, Martin Ruzicka, Václav Gassenbauer, Fabio Pellacini, and Jaroslav Křivánek. 2013. Toward Evaluating the Usefulness of GlobalIllumination for Novices in Lighting Design Tasks. *IEEE Transactions on Visualization and Computer Graphics* 20, 6 (2013), 944–954.
[30] Tero Karras, Samuli Laine, and Timo Aila. 2019. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 4401–4410.
[31] Florian Kern, Carla Winter, Dominik Gall, Ivo Käthner, Paul Pauli, and Marc Erich Latoschik. 2019. Immersive virtual reality and gamification within procedurally generated environments to increase motivation during gait rehabilitation. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 500–509.
[32] William B Kerr and Fabio Pellacini. 2009. Toward evaluating lighting design interface paradigms for novice users. *ACM Transactions on Graphics (TOG)* 28, 3 (2009), 1–9.
[33] Aelee Kim, Minha Chang, Yeseul Choi, Sohyeon Jeon, and Kyoungmin Lee. 2018. The effect of immersion on emotional responses to film viewing in a virtual environment. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 601–602.
[34] Rex B Kline. 2023. *Principles and practice of structural equation modeling*. Guilford publications.
[35] Irene Kotsia, Stefanos Zafeiriou, and Spiros Fotopoulos. 2013. Affective gaming: A comprehensive survey. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 663–670.
[36] Grzegorz Krawczyk, Karol Myszkowski, and Hans-Peter Seidel. 2005. Perceptual effects in real-time tone mapping. In *Proceedings of the 21st spring conference on Computer graphics*. 195–202.
[37] Vincent Laganier and Jasmine Van der Pol. 2011. Light and emotions: exploring lighting cultures: conversations with lighting designers. (2011).
[38] Peter J Lang, Margaret M Bradley, Bruce N Cuthbert, et al. 1997. International affective picture system (IAPS): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention* 1, 39-58 (1997), 3.

[39] Brian Lee, Savil Srivastava, Ranjitha Kumar, Ronen Brafman, and Scott R Klemmer. 2010. Designing with interactive example galleries. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2257–2266.

[40] Vivian Liu, Jo Vermeulen, George Fitzmaurice, and Justin Matejka. 2023. 3DALL-E: Integrating text-to-image AI in 3D design workflows. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. 1955–1977.

[41] Joe Marks, Brad Andalman, Paul A Beardsley, William Freeman, Sarah Gibson, Jessica Hodgins, Thomas Kang, Brian Mirtich, Hanspeter Pfister, Wheeler Ruml, et al. 2023. Design galleries: A general approach to setting parameters for computer graphics and animation. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*. 73–84.

[42] Kent McQuilkin and Anne Powers. 2011. *Cinema 4D: The Artist's Project Sourcebook*. Taylor & Francis.

[43] Ravi Mehta and Rui Zhu. 2009. Blue or red? Exploring the effect of color on cognitive task performances. *Science* 323, 5918 (2009), 1226–1229.

[44] Joseph A Mikels, Barbara L Fredrickson, Gregory R Larkin, Casey M Lindberg, Sam J Maglio, and Patricia A Reuter-Lorenz. 2005. Emotional category data on images from the International Affective Picture System. *Behavior research methods* 37, 4 (2005), 626.

[45] Ron Mokady, Amir Hertz, and Amit H Bermano. 2021. Clipcap: Clip prefix for image captioning. *arXiv preprint arXiv:2111.09734* (2021).

[46] Mohammad Amin Mozaffari, Xinyuan Zhang, Jinghui Cheng, and Jin LC Guo. 2022. GANSpiration: Balancing Targeted and Serendipitous Inspiration in User Interface Design with Style-Based Generative Adversarial Network. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–15.

[47] KAYA Naz and H Helen. 2004. Color-emotion associations: Past experience and personal preference. In *AIC 2004 Color and Paints, Interim Meeting of the International Color Association, Proceedings*, Vol. 5. Jose Luis Caivano, 31.

[48] Changhoon Oh, Jungwoo Song, Jinhan Choi, Seonghyeon Kim, Sungwoo Lee, and Bongwon Suh. 2018. 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*. 1–13.

[49] Rameswar Panda, Jianming Zhang, Haoxiang Li, Joon-Young Lee, Xin Lu, and Amit K Roy-Chowdhury. 2018. Contemplating visual emotions: Understanding and overcoming dataset bias. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 579–595.

[50] Matt Pharr, Wenzel Jakob, and Greg Humphreys. 2023. *Physically based rendering: From theory to implementation*. MIT Press.

[51] Robert Plutchik. 1980. A general psychoevolutionary theory of emotion. In *Theories of emotion*. Elsevier, 3–33.

[52] Ben Poole, Ajay Jain, Jonathan T Barron, and Ben Mildenhall. 2022. Dreamfusion: Text-to-3d using 2d diffusion. *arXiv preprint arXiv:2209.14988* (2022).

[53] Sreeja PS and G Mahalakshmi. 2017. Emotion models: a review. *International Journal of Control Theory and Applications* 10, 8 (2017), 651–657.

[54] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*. PMLR, 8748–8763.

[55] Alec Radford and Jeffrey Wu. 2019. Rewon child, david luan, dario amodei, and ilya sutskever. 2019. *Language models are unsupervised multitask learners. OpenAI blog* 1, 8 (2019), 9.

[56] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125* 1, 2 (2022), 3.

[57] Samuel Rhys Cox, Yunlong Wang, Ashraf Abdul, Christian Von Der Weth, and Brian Y. Lim. 2021. Directed diversity: Leveraging language embedding distances for collective creativity in crowd ideation. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–35.

[58] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10684–10695.

[59] James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology* 39, 6 (1980), 1161.

[60] Aditya Sanghi, Rao Fu, Vivian Liu, Karl DD Willis, Hooman Shayani, Amir H Khasahmadi, Srinath Sridhar, and Daniel Ritchie. 2023. CLIP-Sculptor: Zero-Shot Generation of High-Fidelity and Diverse Shapes From Natural Language. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 18339–18348.

[61] Christoph Schuhmann, Romain Beaumont, Richard Vencu, Cade Gordon, Ross Wightman, Mehdi Cherti, Theo Coombes, Aarush Katta, Clayton Mullis, Mitchell Wortsman, et al. 2022. Laion-5b: An open large-scale dataset for training next generation image-text models. *Advances in Neural Information Processing Systems* 35 (2022), 25278–25294.

[62] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*. 618–626.

[63] Peter Shirley, Michael Ashikhmin, and Steve Marschner. 2009. *Fundamentals of computer graphics*. AK Peters/CRC Press.

[64] Rukshani Somarathna, Tomasz Bednarz, and Gelareh Mohammadi. 2022. Virtual reality for emotion elicitation–a review. *IEEE Transactions on Affective Computing* (2022).

[65] Mariska Stokkermans, IMLC Vogels, Yvonne de Kort, and I Heynderickx. 2018. Relation between the perceived atmosphere of a lit environment and perceptual attributes of light. *Lighting Research & Technology* 50, 8 (2018), 1164–1178.

[66] Thomas Strothotte and Stefan Schlechtweg. 2002. *Non-photorealistic computer graphics: modeling, rendering, and animation*. Morgan Kaufmann.

[67] Jiaming Sun, Zihao Wang, Siyu Zhang, Xingyi He, Hongcheng Zhao, Guofeng Zhang, and Xiaowei Zhou. 2022. Onepose: One-shot object pose estimation without cad models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6825–6834.

[68] Tina M Sutton and Jeanette Altarriba. 2016. Color associations to emotion and emotion-laden words: A collection of norms for stimulus construction and selection. *Behavior research methods* 48 (2016), 686–728.

[69] Ondřej Texler, David Futschik, Michal Kučera, Ondřej Jamriška, Šárka Sochorová, Menclei Chai, Sergey Tulyakov, and Daniel Sỳkora. 2020. Interactive video stylization using few-shot patch-based training. *ACM Transactions on Graphics (TOG)* 39, 4 (2020), 73–1.

[70] Erico Tjoa and Cuntai Guan. 2020. A survey on explainable artificial intelligence (xai): Toward medical xai. *IEEE transactions on neural networks and learning systems* 32, 11 (2020), 4793–4813.

[71] Alexis Van Hurkman. 2014. *Color correction handbook: professional techniques for video and cinema*. Pearson Education.

[72] Tengfei Wang, Bo Zhang, Ting Zhang, Shuyang Gu, Jianmin Bao, Tadas Baltrusaitis, Jingjing Shen, Dong Chen, Fang Wen, Qifeng Chen, et al. 2023. Rodin: A generative model for sculpting 3d digital avatars using diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4563–4573.

[73] Jay Zhangjie Wu, Yixiao Ge, Xintao Wang, Stan Weixian Lei, Yuchao Gu, Yufei Shi, Wynne Hsu, Ying Shan, Xiaohu Qie, and Mike Zheng Shou. 2023. Tune-a-video: One-shot tuning of image diffusion models for text-to-video generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 7623–7633.

[74] Tongshuang Wu, Michael Terry, and Carrie Jun Cai. 2022. Ai chains: Transparent and controllable human-ai interaction by chaining large language model prompts. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–22.

[75] Zhen Xing, Yijiang Chen, Zhixin Ling, Xiangdong Zhou, and Yu Xiang. 2022. Few-shot single-view 3d reconstruction with memory prior contrastive network. In *European Conference on Computer Vision*. Springer, 55–70.

[76] Liwen Xu, Zhengtao Wang, Bin Wu, and Simon Lui. 2022. Mdan: Multi-level dependent attention network for visual emotion analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9479–9488.

[77] Chuan Yan, John Joon Young Chung, Yoon Kiheon, Yotam Gingold, Eytan Adar, and Sungsoo Ray Hong. 2022. FlatMagic: Improving flat colorization through AI-driven design for digital comic professionals. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–17.

[78] Jingyuan Yang, Jie Li, Xiumei Wang, Yuxuan Ding, and Xinbo Gao. 2021. Stimuli-aware visual emotion analysis. *IEEE Transactions on Image Processing* 30 (2021), 7432–7445.

[79] Quanzeng You, Jiebo Luo, Hailin Jin, and Jianchao Yang. 2016. Building a large scale dataset for image emotion recognition: The fine print and the benchmark. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 30.

[80] Wangbo Yu, Yanbo Fan, Yong Zhang, Xuan Wang, Fei Yin, Yunpeng Bai, Yan-Pei Cao, Ying Shan, Yang Wu, Zhongqian Sun, et al. 2023. NOFA: NeRF-based One-shot Facial Avatar Reconstruction. In *ACM SIGGRAPH 2023 Conference Proceedings*. 1–12.

[81] Sicheng Zhao, Yue Gao, Xiaolei Jiang, Hongxun Yao, Tat-Seng Chua, and Xiaoshuai Sun. 2014. Exploring principles-of-art features for image emotion recognition. In *Proceedings of the 22nd ACM international conference on Multimedia*. 47–56.

[82] Sicheng Zhao, Zizhou Jia, Hui Chen, Leida Li, Guiguang Ding, and Kurt Keutzer. 2019. PDANet: Polarity-consistent deep attention network for fine-grained visual emotion regression. In *Proceedings of the 27th ACM international conference on multimedia*. 192–201.

[83] Sicheng Zhao, Xingxu Yao, Jufeng Yang, Guoli Jia, Guiguang Ding, Tat-Seng Chua, Bjoern W Schuller, and Kurt Keutzer. 2021. Affective image content analysis: Two decades review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 10 (2021), 6729–6751.