Surrealistic-like Image Generation with Vision-Language Models

Elif Ayten[1][0009−0005−4954−0475], Shuai Wang[1][0000−0002−1261−9930], and
Hjalmar Snoep[2][0000−0002−6072−4073]

[1] Vrije Universiteit Amsterdam, Amsterdam, the Netherlands
e.ayten@student.vu.nl, shuai.wang@vu.nl
[2] Snoep Studio, Apeldoorn, the Netherlands
snoepstudio@gmail.com

Abstract. Recent advances in generative AI make it convenient to create different types of content, including text, images, and code. In this paper, we explore the generation of images in the style of paintings in the surrealism movement using vision-language generative models, including DALL-E, Deep Dream Generator, and DreamStudio. Our investigation starts with the generation of images under various image generation settings and different models. The primary objective is to identify the most suitable model and settings for producing such images. Additionally, we aim to understand the impact of using edited base images on the generated resulting images. Through these experiments, we evaluate the performance of selected models and gain valuable insights into their capabilities in generating such images. Our analysis shows that Dall-E 2 performs the best when using the generated prompt by ChatGPT.

Keywords: Image generation · generative AI · surrealism · vision-language models.

1 Introduction

The ability of Artificial Intelligence (AI) to generate images from text has been an active field of research over the past few years. For example, OpenAI's DALL·E [1] is a generative model that produces high-quality images from textual descriptions (i.e. prompts). As technology develops, more and more tools are available. DreamStudio [2], Deep Dream Generator [3], MidJourney [4], and various other models are getting more and more powerful in this domain. These models typically require users to input text, in the form of words or longer sentences, to generate images. Alternatively, some models can also take a base image along with the text input to facilitate the generation process. Such progress shows a tendency towards realism, rather than enhancing flexibility and novelty in artistic expression. Some tools make it possible to create some diverse and realistic visuals (e.g. using Generative adversarial networks (GANs) [5]), but at times, the generated objects in these images can exhibit a lack of logical coherence, both in their selection and how they are combined or the style of

assemblage. However, this feature can serve as a valuable source of inspiration for artists [6, 7].

Surrealism was an artistic movement that gained popularity between WWI and WWII. It emerged as an influential artistic movement following advances in modern psychology, especially influenced by the work on unconsciousness by Sigmund Freud [6]. The Surrealists sought to harness the creativity of the unconscious mind. Some of the previous studies claim that these generative models can be considered getting better at generating art that emulates surrealism and fantasy [7]. However, this claim was not justified with experimental evaluation. The generated images can exhibit creative potential in unconventional arrangements by being used as a basis for supernatural imagery or potential ideas to have objects *juxtaposed* (placing different things side by side for contrasting effect) [8]. Our cursory examination shows that some of these generated images can stand alone as thought-provoking pieces, showcasing their artistic value. We aim to explore how generated images can serve as thought-provoking creations, forging connections with the depths of our unconscious mind.
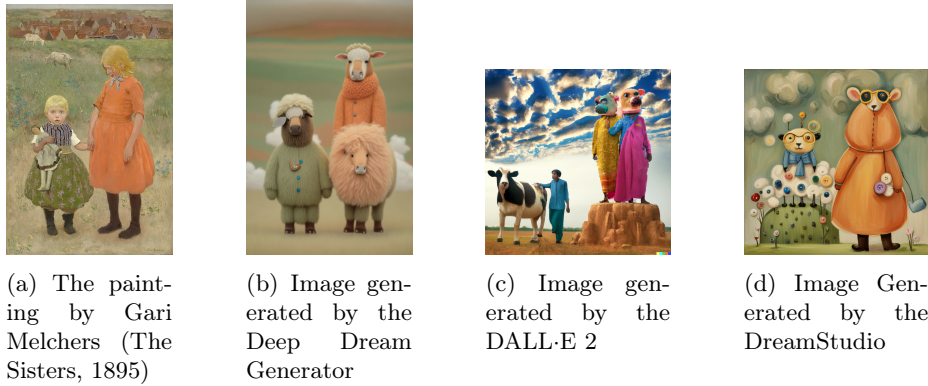

(a) The painting by Gari Melchers (The Sisters, 1895)

(b) Image generated by the Deep Dream Generator

(c) Image generated by the DALL·E 2

(d) Image Generated by the DreamStudio

Fig. 1: Generated images, diverse styles of image generation using different models and corresponding text. Each model's distinct approach yields varied styles of output for image captions.

The main objective of this research paper is to study the generation of images by using vision-language models. We focus on the text-to-image generation and the transformation of realistic images to images in the style of paintings in the surrealism movement. The term 'surrealistic painting' is often used to refer to paintings in the surrealism movement. However, this can be confusing and misleading since being surrealistic is not what paintings in the surrealism movement is all about. Given the lack of a well-established definition, in this paper, we let the term *surrealistic-like* refer to objects that exhibit the following features. These features were extracted from a survey of literature and the official websites of well-known museums and specified galleries.

– **dreamlike atmosphere**: Art pieces that seek to evoke a dreamlike or hallucinatory quality, blurring the boundaries between reality and imagination [9].
– **unexpected juxtapositions**: Art pieces that feature unexpected combinations of unrelated objects, images, or ideas, creating a sense of surprise and mystery [10].
– **symbolism and metaphor**: Art pieces that employ rich symbolism and metaphor to convey hidden meanings and explore psychological and philosophical concepts [11].
– **subversion of logic**: Art pieces that challenge conventional logic and rationality, embracing absurdity and irrationality as a means of expressing the subconscious [12].
– **unconscious exploration** : Art pieces that explore the realm of the unconscious mind, tapping into dreams, fantasies, and irrational thoughts [13].

Our approach is to generate surrealistic-like images from realistic base images. Take the paintings and images in Figure 1 for example. Figure 1a is the selected base image: the original painting named The Sisters by Gari Melchers in 1895. To its right are images generated by different means. Our research aims to determine the optimal settings for generating surrealistic-like paintings. Since there are currently no established tools to definitively judge the quality of surrealistic-like images, we conducted a survey to gather opinions from artists. This approach allows us to evaluate and assess which settings produce the best surrealistic-like artwork. Thus, our main research questions of the paper are the following:

RQ: How can we create surrealistic-like paintings using generative vision-language models?

We take two approaches. The first approach takes texts and the selected base images as input. For comparison, we take only text as input. Our research question is focused on assessing how to optimize the use of existing image generation models effectively. To address our primary research question, we utilize DALL·E2, DreamStudio, and Deep Dream Generator, as they provide capabilities for text-to-image and text-image-to-image generations. We try to identify the most suitable approach and model for creating surrealistic-like paintings. Additionally, we explore the potential of large language models (e.g. ChatGPT) to enhance the performance of image-generative tools automatically. We answer this research question by studying the following sub-research questions (SRQ):

SRQ1: Which image generative approach is best suited for the generation of surrealistic-like images?

SRQ2: Can generative language models make meaningful prompts for the generation of surrealistic-like images?

SRQ3: How can we improve the generation of surrealistic-like images?

To address our research questions, we have an approach that involves conducting three experiments with a survey to evaluate the obtained results. The main contribution of this paper is as follows.

1. A dataset[1] of 235 images was generated using three image generators under different settings. In addition, we generated 12 downscaled images and 12 burred images during our experiments. Finally, we generated 16 images as a showcase for future work.
2. An evaluation of the generated images to determine their suitability for creating surrealistic-like artwork.
3. The paper also presents some primitive steps towards optimizing the settings for the generation of compelling surrealistic-like images.

This paper is organized as follows: Section 2 presents related work and introduces state-of-the-art image generation techniques. Section 3 presents the methodology of this paper, including the models, base images, setup, survey, and image generation. Following that, Section 4 analyzes the evaluation results. Section 5 encompasses an extensive analysis of the results and the discussion. Finally, the conclusion and future work are included in Section 6.

2   Related Work

2.1   Artistic Image Generation

Image generation has witnessed some remarkable progress due to the availability of a large number of datasets and the advances of various large-scale generative AI methods. Among the leading tools in this field are DALL-E [1], Midjourney [4], and Stable Diffusion [14]. DALL-E has a simple interface and it is a text-to-image model [15]. In comparison, Midjourney is more complex and requires its users to use a specified platform (without an API). The stable diffusion model which we use is DreamStudio and can be hit or a miss with the generation depending on the version [15]. These tools use various generative AI techniques, including Generative Adversarial Networks (GANs) [5], Transformers [16], and Variational Autoencoders (VAEs) [17] for the generation of images.

As for evaluation, the Inception Score and Fréchet Inception Distance (FID) are algorithms used to assess the performance of generative image models. The FID score evaluates the quality and diversity of generated images by comparing their distribution with real images. It quantifies the dissimilarity between the feature representations of these two distributions [18]. The Inception Score, conversely, focuses on assessing the recognizability of objects within the generated images [19]. Both algorithms provide valuable insights into the capabilities of generative image models.

Another approach involves human evaluation, incorporating data and results that rely on human perception and understanding. However, it is important to note that the methods used for human evaluation metrics can vary among researchers. The qualifications of assessors may not be specified, or the survey and evaluation processes themselves might be overly complex or overly simplistic [20]. This can lead to differences in the results.

---

[1] The dataset, the code, and the survey and its results are in the repository as open source `https://github.com/ElifAyten/ElifAytenThesis2023`.

3 Methodology

This section presents an overview of the models utilized in this study (Section 3.1) and the base images for the models (Section 3.2), the setup for the prompt generations(Section 3.3). Additionally, an explanation of the image generation processes for each part of the survey (Section 3.4). Finally, the survey design in Section 3.5.

3.1 Models

In our study, we utilize three models for image generation. DALL-E is limited to generating images solely from text input, while Deep Dream Generator and DreamStudio have the capability to generate images from both text and base image inputs. The **Deep Dream Generator** is a generative AI solution focused on image generation [3]. Deep Dream Generator utilizes deep learning techniques to generate visually appealing and imaginative images based on given inputs [3]. **DALL·E**, a deep learning model by OpenAI, is designed for generating images from textual prompts [1]. It uses a text-image pair dataset and is built on the GPT (Generative Pre-trained Transformer) architecture [1]. **DreamStudio** is an interface that utilizes the Stable Diffusion Model, a text-to-image model capable of generating images based on textual inputs [2]. The Stable Diffusion Model incorporates a forward and reverses diffusion process, where an image is transformed into a noisy version in the forward process and then restore the image to its pre-noise state in the reverse process [2].

3.2 Base Images

Our input images consist of a selection of six base images that are curated from renowned realist artists, including Gustave Courbet, Rosa Bonheur, and Meredith Frampton. We deliberately include various realistic styles such as animal realism, satirical realism, and portrait realism. Table 1 provides details about the painting names, the artists' styles, and the respective years of creation.

Painting	Painter	Style	Year
The Sisters	Gari Melchers	Portrait Realism	1895
Greyhounds of Comte de Choiseul	Gustave Courbet	Animal Realism	1866
The Horse Fair	Rosa Bohnear	Animal Realism	1853
Horse in the Sunlight	Guiseppe Abbati	Animal Realism	1866
The Aristocrat's Breakfast	Pavel Fedotov	Satirical Realism	1850
Sir Clive Forster-Cooper	George Vernon Meredith Frampton	Portrait Realism	1945

Table 1: Comparison of the base paintings: name, artist, style, and year of creation.

3.3   Experimental Setup

For our experiments, we use input prompts consisting of realistic base images. To label the objects within these images, we employ the YOLO algorithm [21], which is an object detection tool. YOLO is known for its speed and accuracy.

(a) "The Aristocrat's Breakfast" with YOLO labeled objects and their probabilities: 1 person (Probability: 0.95), 1 dog (0.93), 2 chairs (0.80 and 0.92), 1 couch (0.50), 1 dining table (0.55), 1 book (0.30), 1 vase (0.89).

(b) "Sisters" with its detected objects and their probabilities using YOLO: 1 person (0.95), 1 person (0.96), 1 teddy bear (0.83), 1 sheep (0.42), 1 sheep (0.59).

Fig. 2: YOLO labeled images

Once the labeling process is completed, we use ChatGPT-3, to create longer and more detailed prompts based on the extracted object labels. For example, for Figure 2, we have the input as "Generate a prompt for surrealistic image generation within 15 words that contains the following objects: 1 teddy bear,2 sheep and 2 person". The generated descriptions are as follows. *Prompts of 15 words*[2] each: Two persons encounter two sheep, a cow, and a teddy bear amidst a floating garden *Prompts of 50 words* each:Two persons, one adorned with a cow mask, guide two sheep, each wearing oversized spectacles, towards a towering teddy bear monument. The sky weaves patterns of wool, while fluffy clouds rain streams of multicolored buttons, creating a surreal landscape of whimsical companionship.

---

[2] All the prompts in this paper were generated on June 3rd, 2023.

3.4 Image Generation

We conduct our image generation study in three settings described in the three following parts.

Part 1: Comparative analysis of multiple generative models The initial part of our study aims to compare three image generation models: Deep Dream Generator, DreamStudio, and DALL-E (version 2). We use different inputs for each model. For DeepDream and DreamStudio, we try both base images and diverse prompts. As for DALL-E, we focus on various prompt settings due to its unique nature. Our goal is to evaluate their abilities in producing surrealistic images by adjusting prompt types and lengths. In the first part, we employ three key setups for image generation. First, we create images using YOLO labels as prompts—simple words describing objects in the images. Second, we compare the models using 15-word prompts generated by ChatGPT, a proficient language model. Finally, we explore ChatGPT prompts containing 15 words and the artist's name from the base image. This extra prompt component adds contextual information about the artist, potentially impacting the surrealistic qualities of the images.

Part 2: Comparing various prompt settings within each generative model We focus on testing each model individually, examining their performance with different prompt settings. We specifically explore the capabilities of the three models using the same base image under various prompt settings. The objective of the second part is to assess how each model handled different prompt lengths and determine which model performed better under specific settings. By varying the prompt lengths and analyzing the generated results, we aim to gain insights into the strengths and limitations of each model and identify the most effective prompt settings for surrealistic-like image generation. Different from the previous phase, we also include prompts of 50 words, enabling more details.

Part 3: Base image editing comparison for Deep Dream Generator and DreamStudio Finally, we compare the performance of the models when using edited base images versus non-edited base images. First, we apply a blurring effect to each base image, creating a set of blurred images. These blurred images, along with a 50-word long prompt, are used as input for both the Deep Dream Generator and DreamStudio models. Second, we downsize the base images to one-fourth of their original size, generating a set of downscaled images. These downscaled images, accompanied by a 50-word prompt, are compared to the results obtained using only a 50-word prompt without downsizing.

3.5 Survey

Given the absence of neural networks capable of accurately determining features in this domain, we developed a large survey that relies on subjective opinions

by artists. We had 18 participants after reaching out to art students and recent graduates of some art schools and universities, including the Gerrit Rietveld Academie, Vrije Universiteit Amsterdam, Hogeschool voor de Kunsten Utrecht, and the Eugeniusz Geppert Academy of Art and Design in Wroclaw. Since the personal information of the participants was anonymized, the authors did not track the exact affiliation of each. The questionnaire is divided into three parts, each corresponding to the image generation as described in the previous subsection.

In Part 1 of the survey, we conduct a comparative analysis of multiple generative models and settings. Our goal is to gain insights into how different settings such as YOLO labeling and prompts generated by ChatGPT, with lengths of 15 words and 15 words with artist's name prompts. In Part 2 of the survey, our focus is on comparing different prompt settings within each generative model. We examine three models and evaluate their performance under distinct prompt configurations, including YOLO labeling and prompts generated by ChatGPT, with lengths of 15 words and 50 words and simply the YOLO labeling. In Part 3 of the survey, we introduce base image downscaling and blurring. Participants are presented with variations of the base input image for Deep Dream generator and DreamStudio models, including downscaled versions and blurred images.

The survey comprises a total of 60 questions, with three sub-questions each:

(a) Which image(s) translates the subject matter in the original painting to something unexpected (object or element of surprise, etc.)? [multiple answers can be selected]
(b) Which generated image(s) has unexpected juxtapositions of objects? [multiple answers can be selected]
(c) Which one is the best surrealistic-like painting among the three paintings, in your opinion? [only one answer can be selected]

4 Evaluation

To ease the description of the analysis of the survey data, we introduce the following notation of prompt setting (PS) and (PIS) prompt-image setting: **PS1:** The YOLO algorithm provides information about the objects and their quantities, such as '2 dogs, 1 person', which is used in the image generation process. **PS2:** Given the PS1 and a description request, ChatGPT generates prompts with a word limit of 15. For instance, the input "The description of a surrealistic-like painting with 1 person, 2 dogs in 15 words" is used to generate images. **PS3:** Given the PS1 and a 15-word request for a description, along with the name of an artist, ChatGPT generates a prompt for image generation. For instance, the input "The description of a surrealistic-like painting with 1 person, 2 dogs in 15 words, Gari Melchers" is used to generate images. **PS4:** Similar to PS2 but with a 50-word request for a description. For instance, the input "The description of a surrealistic-like painting with 1 person, 2 dogs in 50 words" is used to generate images. **PIS5:** Similar to PS4 but with the base image also as input. Additionally, the base image will

| Question	DALL-E 2	DreamStudio	Deep Dream Generator
P1-SQA	50.8%	9.0%	40.1%
P1-SQB	50.8%	8.8%	40.4%
P1-SQC	35.6%	12.5%	51.9%
P2-SQA	34.6%	30.7%	34.6%
P2-SQB	34.2%	31.0%	34.8%
P2-SQC	33.2%	32.7%	34.2%
P3-SQA	N|A	47.3%	52.7%
P3-SQB	N|A	47.0%	53.0%
P3-SQC	N|A	49.6%	50.4%

Table 2: Participants' preferences for sub-question selection in each survey section, expressed as percentages, with the most chosen generator highlighted.

be downscaled to its $1/4$ size. **PIS6:** Similar to PS4 but with the base image also as input. Additionally, the base image will be blurred (Gaussian blur) by using the Python Imaging Library (PIL).

Table 2 presents various sections of the survey, denoted as P1, P2, and P3, along with distinct sub-questions labeled SQA, SQB, and SQC. During our analysis of the responses to these sub-questions, we segregate the survey parts due to their specific content. In Part 1 and 2, all three models are featured, whereas Part 3 is exclusively about DreamStudio and Deep Dream Generator. It is evident that, concerning DALL-E, sub-questions SQA and SQB are the most frequently selected, while for DreamStudio, sub-question SQC is favored. Similarly, for Deep Dream Generator, sub-question SQC emerges as the predominant choice among participants.

Figure 3a shows the distribution of participants' answers for all the model settings. For the DALL-E PS1, the median count of participants' answers is 4.50. In contrast, the Deep Dream Generator PS1 shows a higher median count of 13.50. For the DreamStudio PS1, the median count is 1.00, indicating that it receives the least preferred overall. It is apparent that Deep Dream Generator is the model that is selected most frequently among the participants. The analysis of Figure 3b reveals that the DALL-E model with a 15-word prompt generated by ChatGPT was the most favored choice among the participants. Similarly, from Figure 3c, DALL-E model outperforms the other two models with a 15-word prompt generated by ChatGPT and artist name setting. Little difference was observed between PS2 and PS3 for Dall E. In summary, the Deep Dream Generator is the most preferred model for YOLO labeling while DALL-E is the most preferred tool if given textual input by ChatGPT (PS2 and PS3).

In Figure 4, we can observe the performance of different models under various settings. Figure 4a reveals that DALL-E achieves its highest median performance 12.0 with a prompt setting PS4, indicating its superior results compared to other prompt settings. The first quartile (Q1) and third quartile (Q3) values further support this effectiveness. Comparatively, the DALL-E PS2 has a median count of 10.5 while the YOLO labeling model receives
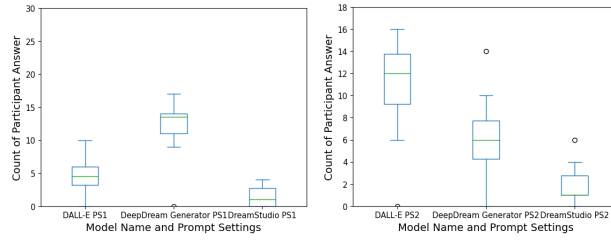
(a) Participants' choices of three models regarding PS1

(b) Participants' choices of three models regarding PS2
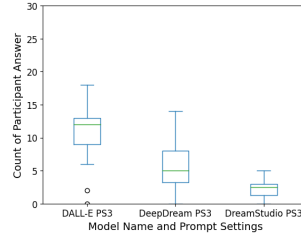
(c) Participants' choices of three models regarding PS3

Fig. 3: Survey Part 1 Results

the lowest median count 2.0. To summarize, DALL-E performs best with a prompt setting of 50 words.

In Figure 4b and Figure 4c, we observe that the Deep Dream Generator and DreamStudio exhibit different trends in comparison with Dall E. However, neither of them is comparable with Dall E. In Figure 4d and 4e, the effects of downscaling and blurring on the base image's impact on the models' performance are found to be limited. The outcomes indicate no significant enhancement compared to the original images. Notably, the downscaled iteration displays slightly superior performance compared to the blurred version, as indicated by our statistical analysis. For the Deep Dream Generator model, our analysis indicates that all three configurations exhibit similar distributions of participants' responses 4d . The median scores are 18.5 for Deep Dream Generator PS4, 9.0 for Deep Dream Generator Blurred, and 9.0 for Deep Dream Generator Downscaled. In conclusion, while downscaling and blurring the base images do not lead to a significant improvement in the models' performance, the downscaled version appears to have a slight advantage over the blurred version.

## 4.1   Aesthetics Analysis

Generated Images As the prompt size grows, Dall E generates more surrealistic-like images with good color combinations. The generated images exhibit a more frequent interplay of light and dark shades, which makes it
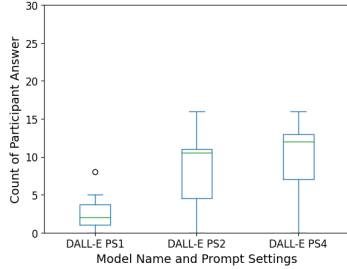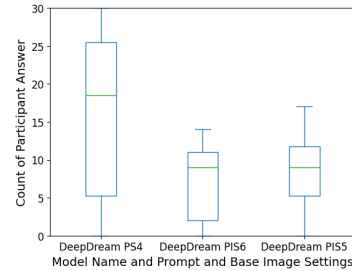
(a) A comparison between participants' choices of DALL-E regarding PS1, PS2, and PS4
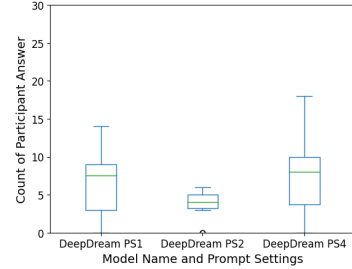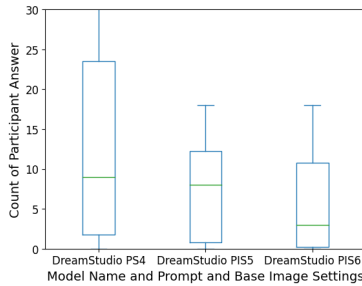

(b) A comparison between participants' choices of Deep Dream regarding PS1, PS2, and PS4
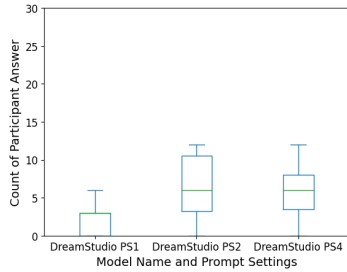
(c) A comparison between participants' choices of DreamStudio regarding PS1, PS2, and PS4

(d) A comparison between participants' choices of Deep Dream Generator regarding PS4, PIS5, and PIS6

(e) A comparison between participants' choices of DreamStudio Generator regarding PS4, PIS5, and PIS6

Fig. 4: Survey Part 2 and 3 Results

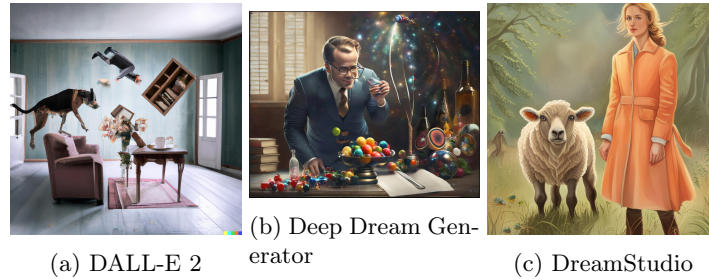(a) DALL-E 2   (b) Deep Dream Generator   (c) DreamStudio

Fig. 5: Examples for the comparison of aesthetics of generated images

easy to raise attention. We also noticed that the model grasps concepts of the input prompts well, though not always the provided numbers. DALL-E's image generation is somewhat limited due to its production of squared images. Some generated images are particularly interesting. For example, Figure 5a exhibits the idea of Dali's loss of gravity.

As for the second model, the Deep Dream Generator, we observe that it produces a wide range of colors in its generated images, e.g. Figure 5b. Deep Dream often shows the capacity to incorporate colors from a base image, resulting in outcomes with similar color combinations, likely due to the influence of the input base image (e.g. Figure 5b). It showcases a variety of unexpected object combinations and surrealistic-like elements, even when provided with simple word prompts through YOLO labeling.

DreamStudio, the final model in our analysis, exhibits certain limitations in its capabilities. While some of the generated images are surrealistic-like, there are instances where the images lack any surrealistic-like elements (e.g. Figure 5c). Despite this, DreamStudio excels at capturing the color, style, and positioning of objects from the base image when generating new images.

Generated Prompts Our experiments show that ChatGPT-generated prompts consistently feature YOLO labelings, as previously mentioned. The resulting prompts are remarkably clear and easily comprehensible, demonstrating the ability to produce surrealistic-like content implicitly, even without explicitly using the term "surrealistic." Notably, the prompts exhibited uniqueness and creativity, encompassing various words, actions, objects, and individuals. However, some generated prompts could be more creative.

5 Discussion

The statistical analysis and observations consistently support that DALL-E excels in surrealistic-like image generation, when provided with a 50-word prompt generated by ChatGPT in comparison to other settings. DALL-E's image generation capability gets better when longer prompts are provided since the model relies on textual input to produce images. The results indicate that making use of generative language models for surrealistic-like

prompts positively impacts the performance of the generative image model. The prompts generated by ChatGPT receive a high number of votes from the participants. This suggests that ChatGPT's natural language processing capabilities contribute to producing prompts that was considered surrealstic-like by our panel of artists, resulting in visually captivating and surrealistic-like images.

Furthermore, our analysis indicates a correlation between prompt length and participants' preferences. Longer prompts generated by ChatGPT received more votes, indicating that comprehensive and detailed textual guidance plays an important role in achieving superior surrealistic-like image outcomes. Longer prompts offer richer contextual information, enabling generative image models to produce more surrealistic-like results. The integration of language models in guiding image generation has the potential to result in even more sophisticated surrealistic-like artworks, inspiring artists and creators to explore novel artistic horizons. We observe that including the artist's name as a prompt did not significantly impact image generation for the chosen models. However, it is valuable that DALL-E, being a text-based generative model, exhibited a unique advantage in capturing the essence of the base image's artist. Since DALL-E does not rely on input images, incorporating the artist's name allowed it to interpret and incorporate the stylistic elements associated with the mentioned artist. As a result, DALL-E generated images with a higher resemblance to the specified artist's style, showcasing its adaptability in producing artistically influenced outputs based solely on textual descriptions. Despite not having a substantial impact on other models, this capability of DALL-E highlights its potential to leverage artistic references effectively in the image generation process.

The preference for downscaled base images as input to the model increased in votes across generations, when contrasted with the utilization of blurred base images as input. This indicates that the downsizing technique may outperform blurring in enhancing the quality of generated surrealistic artworks. By downsizing the base images to one-fourth of their original size and using a 50-word long prompt, we introduce a novel element that resonates with participants. It's also critical to recognize the possibility of bias created by both the choice of base images and their application in our study. The selection of the base images carries a subjectivity that may unintentionally impact the results that are produced. Due to the intrinsic artistic styles embodied in these base images, the generative process may unintentionally lean towards those particular aesthetics. Furthermore, the use of terminology like "surrealism" while creating prompts carries the risk of prejudice. The use of this phrase might generate a bias towards adhering to accepted ideas about surrealistic art, which would then direct the model's interpretation and creative output.

Our analysis shows that all three models can generate some images that are appealing to artists. This indicates that they can use such language-vision models to boost their creativity. The combination of human and AI

creativity can spark new ideas and inspire artworks. Knowing which model and settings to use can save time for artists.

The outcomes from the particular sub-questions may not provide straight-forward conclusions due to the exclusive focus on DreamStudio and Deep Dream Generator in Survey Part 3. Nevertheless, we can discern a notable trend: Deep Dream Generator garnered the highest number of votes in response to Part C questions, despite the constraint of allowing only a single choice. This contrast with Sub-questions A and B, which offered multiple selections, suggests that the clarity and simplicity of the Part C inquiries likely facilitated participants' choices. It can be inferred that the nature of the question, being comprehensible and straightforward with a single-response format, contributed to this outcome. Despite Deep Dream Generator receiving a significant number of votes in the sub-questions, DALL-E remains the premier choice as a surrealistic image generator. This distinction arises from DALL-E's exclusion in the final segment of the survey, where it particularly shines in specific configurations.

6   Conclusion and Future Work

In conclusion, our study evaluates the capabilities of three different generative models. We explore various settings for each model to determine the most suitable configurations, and to test each model effectively for optimal image generation. Addressing our first research question **SRQ1** (Which image generative models are best suited for the generation of surrealistic-like images), we present the participants' preferences when shown the images. The majority of participants selected DALL-E with the 50-word prompt setting(PS4) for producing the best surrealistic-like images. **This finding highlights DALL-E as the most proficient model among the three for generating surrealistic-like imagery, particularly when guided by ChatGPT-generated prompts.** Furthermore, when images are generated solely by DALL-E without input from ChatGPT, the resulting images tend to be less surrealistic-like. On the other hand, Deep Dream Generator, by default, tends to produce surrealistic-like images even without the need for external input or creative prompts solely from YOLO. This behavior is intrinsic to the Deep Dream Generator, which enhances patterns and details in an image in a way that often leads to surrealistic-like images.

For our second **SRQ2** (Can generative language models make meaningful suggestions by providing prompts), the results obtained indicate that the ChatGPT-generated prompts are highly preferred among participants compared to YOLO labelings alone. The preference for ChatGPT-generated prompts surpassed that of YOLO labelings, suggesting that generative language models can provide meaningful suggestions for image generation.

As for the final sub-research question **SRQ3** (How can we improve the generation of images), is addressed based on our results, revealing that settings involving a ChatGPT prompt of 50 words and a downscaled base image did not receive a high preference. However,the downscaled version is preferred

over the blurred version. Although downscaled base images show some improvement, they do not have a significant impact on DreamStudio and Deep Dream Generator. Additionally, to reaffirm our finding that PS4 with DALL-E is the best configuration, we decided to further explore DALL-E's capabilities by generating additional surrealistic-like images. For this purpose, we enriched the prompts with names of renowned surrealistic artists. Now, we present these new surrealistic-like creations, achieved by DALL-E with the 50-word prompt setting combined with the influence of surrealistic artists.

(a) An example using the name Max Ernst

(b) An example using the name Salvador Dalí
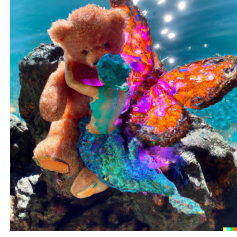
(c) An example using the name Juan Miro

Fig. 6: Images generated by Dall E 2 with the prompt of 50 words by ChatGPT including artists' names (see the supplementary material for the prompts)

In Figure 6, we observe that incorporating surrealistic artist names along with the ChatGPT generated prompt significantly enhances the surrealistic quality of the images. The presence of surrealistic artist names seems to stimulate the generation process, inspiring the model to produce artwork with a heightened level of surrealistic-like characteristics. The combination of the model's creative capabilities and the influence of renowned surreal artists creates better surrealistic-like images.

By combining their strengths, new generative models can be explored for use in the future work, which will lead to the exploration of more hybrid methodologies. Giving ChatGPT the assignment of creating artworks that reflect the distinctiveness of particular artists with their favorite color schemes and style components, for instance, will give the generative process more depth. The generations could become more surrealistic-like by experimenting with the already-present surrealistic-like components from the paintings. For instance, melting things, varied perspectives, and unusual scales. Even while some may believe these tools can be limited and can constrain artists, there is great potential for unique styles and novel forms of what we call surrealism. In the end, the development of surrealistic-like images can be enhanced and go further with the help of creative AI. This symbiotic relationship may allow artists to explore innovative art forms that go beyond traditional boundaries or find their distinct style in AI-generated art.

References

1. OpenAI. DALL-E: Creating Images from Text. *OpenAI*, 2021. Accessed: June 28, 2023.
2. Dreamstudio. `https://beta.dreamstudio.ai/generate`. Accessed: June 28, 2023.
3. Deep dream generator. `https://deepdreamgenerator.com/generator`. Accessed: June 28, 2023.
4. Midjourney. `https://www.midjourney.com/app/`. Accessed: July 12, 2023.
5. Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017.
6. Aaron H. Esman. Psychoanalysis and surrealism: André breton and sigmund freud. *Journal of the American Psychoanalytic Association*, 59(1):173–181, 2011.
7. James Gurney. What i've been thinking about generative art, Dec 2022.
8. Jose Antonio Lanz. Ai kryptonite: Why artificial intelligence can't handle hands, Apr 2023.
9. Tate. Surrealism. `https://www.tate.org.uk/art/art-terms/s/surrealism`. [Accessed 26-07-2023].
10. MoMA | Surrealist Objects and Assemblage. `https://www.moma.org/learn/moma_learning/themes/surrealism/surrealist-objects-and-assemblage/`. [Accessed 26-07-2023].
11. James Voorhies. Surrealism. `https://www.metmuseum.org/toah/hd/surr/hd_surr.htm`. [Accessed 26-07-2023].
12. https://www.facebook.com/thoughtcodotcom/. These Artists Thrived on Dreams - Discover Their Surreal World — thoughtco.com. `https://www.thoughtco.com/what-is-surrealism-183312`. [Accessed 26-07-2023].
13. GUGGENHEIM museum: Surrealism: Two Private Eyes. `http://pastexhibitions.guggenheim.org/surrealism/`. [Accessed 26-07-2023].
14. AI Image Generation Explained: Techniques, Applications, and Limitations — altexsoft.com. `https://www.altexsoft.com/blog/ai-image-generation/`. [Accessed 24-Jul-2023].
15. Craig Boehman. Midjourney vs. dall-e vs. stable diffusion: Which is better?, May 2023.
16. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017.
17. Diederik P Kingma and Max Welling. Auto-encoding variational bayes, 2022.
18. Chan-Ik Kim, Mingu Kim, Seojin Jung, and Euisin Hwang. Simplified fréchet distance for generative adversarial nets. *Sensors (Basel, Switzerland)*, 20(6):1548, 2020.
19. Shane Barratt and Rishi Sharma. A note on the inception score, 2018.
20. Mayu Otani, Riku Togashi, Yu Sawai, Ryosuke Ishigami, Yuta Nakashima, Esa Rahtu, Janne Heikkilä, and Shin'ichi Satoh. Toward verifiable and reproducible human evaluation for text-to-image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14277–14286, 2023.
21. Rohit Kundu. YOLO object detection. `https://www.v7labs.com/blog/yolo-object-detection`, Year. Accessed: 23-Aug-2023.