Integrating Artistic Metaphors into Text-to-Image Generation

Yunshan.gong@outlook.com
Beijing NO.101 School

ABSTRACT

Metaphors in artworks often serve as a bridge that helps viewers connect with the artist's true emotions, and many great pieces of art are filled with rich metaphors, which not only possess strong artistic qualities but also greatly enrich the manner in which content is expressed. In recent years, the advancement of text-to-image technology has enabled more people to create their own artworks merely by providing prompts. Nonetheless, integrating appropriate metaphors into the generated artworks remains a challenging task. This study extends the approach of automatically generating metaphors in text to the realm of text-to-image generation by constructing metaphorical triangles within prompts and introducing three distinct methods to achieve this. Two of these methods are unsupervised, while the third uses LambdaRank Method to integrate multiple features. In the experiments, these methods were applied to infuse 50 prompts spanning 10 different categories with common metaphorical objects extracted from artworks. The subsequent analysis evaluated the proposed methods based on the diversity and consistency of the results, showing that all three methods were effective. Additionally, the experiments revealed significant differences in the difficulty of constructing metaphors across prompt categories, which merits further exploration.

KEYWORDS

Metaphor generation, visual semantics, text-to-image generation, prompt engineering

1. INTRODUCTION

In art galleries, it is common to encounter a vast array of masterpieces with metaphors. These pictorial metaphors not only demonstrate profound artistic merit, but also significantly augment the expressive possibilities of the artwork, aiding viewers in comprehending the artist's creative intentions. A prime example is Mexican painter Frida Kahlo's "The Broken Column," a self-portrait recognized as one of her most iconic works [1], as shown in Figure 1. In the painting, Frida's spine is depicted as a broken column, symbolizing the spinal injury she suffered from a car accident and her prolonged physical pain. Her body is pierced with iron nails, which represent the agony she endured during her recovery and the lifelong torment of her illness. By visually depicting personal suffering, her work resonates with a profound emotional intensity.

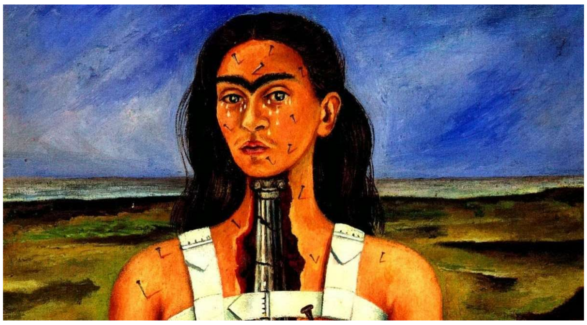


Figure 1: Both the broken column and the nails carry metaphorical meanings in "The Broken Column"

In recent years, with the development of text-to-image technology, ordinary people can engage in artistic creation by using prompts to generate impressive images through artificial intelligence, even mimicking the style of certain artists. However, incorporating appropriate metaphors into generated works remains a challenging task, requiring the creator to have a rich imagination and the ability to blend these metaphors seamlessly into the work. On the other hand, scholars specializing in art have accumulated extensive research on metaphors in painting. Integrating this prior knowledge into text-to-image technology can provide new tools for AI-based artists, helping them enrich the content of their images and make their works more diverse.

2. RELATED WORK

Research on the integration of metaphors into text-to-image generation builds upon three primary domains of prior work.

Research on Metaphor in Art. There's a wealth of research on metaphors in artworks. S. Malaguzzi and B. Phillips, in monographs "Food and Feasting in Art" [2] and "Jewelry, Stones and Precious Objects,"[3] catalogs frequently appearing food, tableware, and jewelry in art, analyzing their symbolic meanings. L. Impelluso and S. Sartarelli, in "Gardens in Art," [4] dissects the constituent elements of gardens and their hidden symbolism. Additionally, in "Gods and Heroes in Art," [5] They lists legendary figures from classical antiquity, along with their stories and special significance. Petrenko, V.F. provides illustrative analyses of various forms of metaphors in paintings, such as metonymy, hyperbole, irony, and oxymoron [6].

Metaphor construction in text. The second area focuses on metaphor construction in linguistic studies. Mengshi Ge et al. have comprehensively summarized various techniques for metaphor identification, interpretation, and generation in recent years [7], and listed datasets that support these techniques. Among these, D. Zheng et al. designed a metaphor generation model based on connectivity scores, which was successfully applied in Microsoft's chatbot, demonstrating the model's strong practicality [8].

Prompt optimization in text-to-image generation. Research on enhancing text-to-image quality through prompt optimization has proven to be highly fruitful. W. Mo et al. proposed a dynamic prompt optimization method, employing an online reinforcement learning strategy to explore the weight and injection time step of each word, resulting in dynamically fine-grained control prompts [9]. This method simultaneously considers aesthetic scores, semantic consistency, and user preferences, effectively improving the original prompts. Y. Hao et al. introduced a prompt adaptation method capable of automatically adjusting the original user input into prompts more aligned with the model's preferences. This method preserves the user's original intent while generating more aesthetically pleasing images [10].

In summary, the research findings across these three domains above provide valuable insights into the incorporation of metaphors within text-to-image generation. However, these studies have been conducted independently. This research aims to synthesize these successful investigations and discovering “chemical reactions” within this intersectional field.

3. PROBLEM FORMULATION

The metaphor generation model proposed by D. Zheng et al. has achieved excellent practical results [8]. They describe the process of constructing metaphors using a triangular structure, where the three vertices represent the source, target, and connection of the metaphor. The values of the three edges are calculated with the semantic similarities between the vertices using word2vector tools [14]. For example, the triangular structure of the metaphor "love is as lucky as lottery" is shown as below.

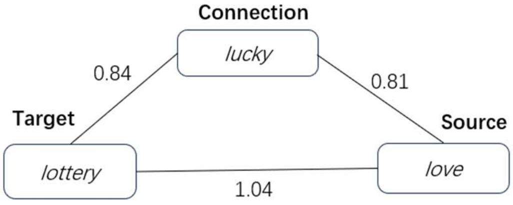


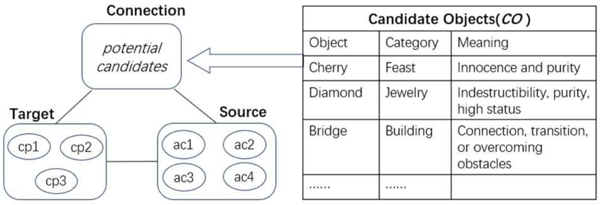
Figure 2: Triangular Structure of the Metaphor in Text

This triangular structure can be adapted to establish metaphors in the text-to-image generation process. Prompts often contain multiple concrete phrases (*CP*) and imply various abstract concepts (*AC*). For instance, in the prompt "*Design an intricate high fantasy digital art piece, focusing on a sprawling Elven city nestled in an ancient forest, bathed in the ethereal twilight glow*," the concrete phrases and abstract concepts contained within are listed in the table below. The extraction of these elements can be accomplished by calling the API interfaces of large language models such as ChatGPT or Gemini.

Concrete Phrases(*CP*)	
cp1	Sprawling Elven city
cp2	Nestled in an ancient forest
cp3	Bathed in ethereal twilight glow

Abstract Concepts(*AC*)	
ac1	Harmony with Nature
ac2	Timelessness
ac3	Magic and Enchantment
ac4	Tranquility and Peace

The concrete content can be mapped to the target of the metaphor, while the abstract concepts correspond to the connection in the metaphor triangle. Objects that have frequently appeared as sources of metaphors in artworks, such as jewelry, food, and so on, can all serve as potential candidates to be added to this new metaphor triangle. The candidate object can be defined as a 3-tuple containing a name, its category, and its metaphorical meaning in artworks.



Candidate Objects(*CO*)		
Object	Category	Meaning
Cherry	Feast	Innocence and purity
Diamond	Jewelry	Indestructibility, purity, high status
Bridge	Building	Connection, transition, or overcoming obstacles
......	......	

Figure 3: Triangular Structure of the Metaphor in Text-to-Image Generation Systems

From above, the Prompts, Concrete Phrases (*CP*), Abstract Concepts (*AC*) and Candidate Object (*CO*) are defined as follows:

$$Prompt=(CP, AC)$$

$$CP=（cp_1,...cp_m）$$

$$AC=(ac_1,...ac_n)$$

$$CO=(co_1,...co_t)$$

The problem can be defined as follows: given a *Prompt*（*CP*，*AC*）, identify the suitable candidate objects(*CO*) from a given set which is drew from artworks, such that the objects and the *Prompt* together form valid metaphors.

4. PROPOSED METHOD

In previous studies, the method of using connecting scores to select the target(*T*) and source(*S*) through Connection(*C*) has been proven effective [8], and it is defined as follows:

$$Connecting\ Score(C|T,S)\ =$$
$$Dist(S,C)\ +\ Dist(T,C) + log(|Dist(S,C)\ -\ Dist(T,C)| +\ \beta)$$

This score is utilized to identify the most suitable metaphor triangles, and two methods for calculating this score are proposed in the text-to-image generation process.

4.1 Maximum Connecting Score Model

In many artworks, the source of a metaphor is often applied to a single subject in the composition, conveying the meaning associated with that subject. For instance, in Frida Kahlo's "*Self-Portrait with Thorn Necklace and Hummingbird*", the thorn necklace forms a metaphor with the girl depicted in the painting, while its connections to other elements, such as the butterflies, monkey, cat, and plants in the background are weaker.

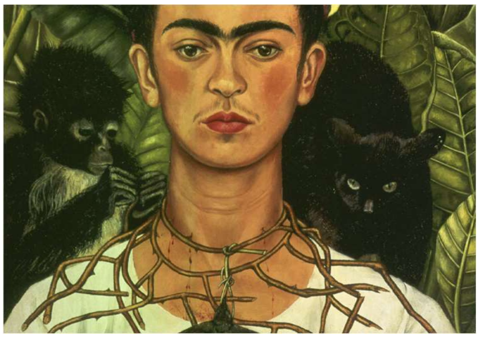


Figure 4: The thorn necklace pierces Frida's skin, drawing blood, symbolizing her physical pain and emotional trauma

Based on this idea, the construction of the metaphor triangle is refined to focus on each concrete phrase (cp_i) and abstract concept (ac_j), as illustrated in the graph.

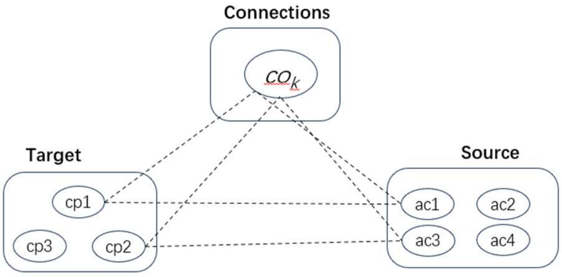


Figure 5: Multiple Potential Metaphor Triangulars in Artwork

There are multiple potential metaphor triangles as shown above. Define the Connecting Score of each triangle as below:

$$CS\ (co_k\ /\ cp_i,\ ac_j) = Dist(co_k,\ cp_j) + Dist(co_k,\ ac_j)+$$
$$log(|Dist(co_k,\ cp_j) - Dist(co_k,\ ac_j)|+\beta)$$

For each candidate object co_k, the max connecting score among of all the potential metaphor triangle is selected as the connecting score for candidate object co_k.

$$CS(CO_k|CP,AC)^{*}_{k} = \mathop{arg\,max}_{cp_i \in CP, ac_j \in AC} CS(co_k|cp_i, ac_j)$$

4.2 Average Connecting Score Model

In some artworks, the source of a metaphor relates to multiple subjects within the composition, collectively enriching the depth and meaning of the artwork. For instance, in Gustave Moreau's painting "*Oedipus and the Sphinx*" [11], the metaphorical themes of the conflict between good and evil are intricately depicted through the interplay of multiple elements, including the corpse in the foreground, the human figures, butterflies, a viper, and a fig tree.

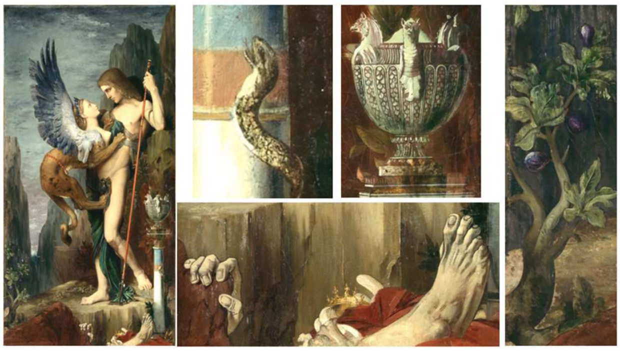


Figure 6: Metaphorical themes are intricately depicted through the interplay of multiple objects, including the corpse in the foreground, the human figures, a viper, and a fig tree

Inspired by this thought, the average value of each edge of all potential metaphor triangles is adopted as the distance between nodes to calculate the connecting score, shown as below.

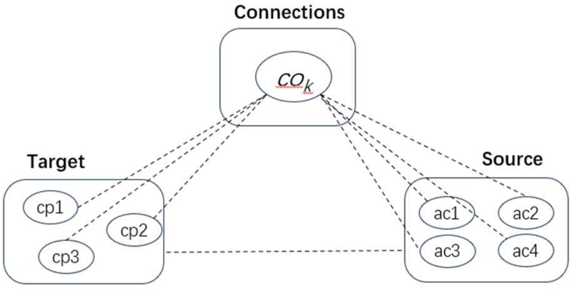


Figure 7: Multiple elements in the work interact with each other to form a metaphor

For each candidate object co_k, the average connecting score with cp_i, is used to measure the distance of between CO_k and CP, and the average connecting score with ac_j is used to measure the distance of between CO_k and AC,

$$Dist_{avg}(CO_k, CP) = \sum_{cp_i \in CP} Dist(co_k, cp_i)/m$$

$$Dist_{avg}(CO_k, AC) = \sum_{cp_i \in CP, ac_j \in AC} Dist(co_k, ac_j)/n$$

$$CS_{avg}\ (CO_k|CP,AC) = Dist_{avg}(CO_k,\ CP) + Dist_{avg}\ (CO_k,AC) + log(|Dist_{avg}\ (CO_k,\ CP) - Dist_{avg}\ (CO_k,\ AC)| + \beta)$$

4.3 Combine More Features by Learning

Machine learning methods can be employed to integrate the two models mentioned above. By further including the intrinsic features of the CO_k and CP, a total of 6 features, as outlined in the table below:

Table 1: Summary of all features

Features	Description
$Dist_{max}(CO_k, CP)$	Left edge in Maximum Connecting Score Model.
$Dist_{max}(CO_k, AC)$	Right edge in Maximum Connecting Score Model.
$Dist_{avg}(CO_k, CP)$	Left edge in Average Connecting Score Model.
$Dist_{avg}(CO_k, AC)$	Left edge in Average Connecting Score Model.
Density of CP	The average distance between all concrete phrases in this CP.
Candidate Category	The category of the candidate objects as show in Figure 3.

The LamdaRank [16] method can integrate the 6 features to rank candidate objects, which will be discussed in the experimental section.

5. EXPERIMENTAL RESULTS

In this section, a prompt dataset and a metaphorical object corpus that support the experiment are introduced, and two metrics for evaluating the quality of metaphors in text-to-image generation follow thereafter. Next, the performance of the proposed methods and the comparison are presented. Finally, some qualitative case studies demonstrate the effectiveness of the proposed methods.

5.1 Experimental Setup

The prompts dataset was derived from a blog by a company specializing in printing services, which provides guidance to its customers on crafting high-quality prompts [12]. This dataset comprises 10 categories, each containing 5 distinct prompts, as detailed in the table below:

Table 2: Categories of the Prompt Dataset

ID	Category Name	Abbreviate
1	Fantasy worlds and characters	Fantasy
2	Science fiction and futurism	Science
3	Historical and cultural imagery	History
4	Nature and wildlife	Nature
5	Abstract and surreal art	Abstract
6	Portraits and character studies	Portrait
7	Architectural wonders	Arch
8	Cyberpunk and neon aesthetics	Cyber
9	Cross-genre mashups	Mashup
10	Seasonal themes and events	Season

The metaphorical object corpus is compiled by organizing metaphors found in artworks. This corpus contains 16 categories, each with 20 objects frequently appearing in artworks, as shown in the **Table 3**. Each object is annotated with its name and the meaning it represents in artworks. For example, the object "bridge" often expresses the meaning of connection, transition, or overcoming obstacles in artworks.

Table 3: Categories of the Objects with Metaphor Meaning

id	Categories Name	Objects
1	Foods	Apple, Pear, Grapes, Fish...
2	Body parts	Eyes, Hands, Heart, Feet...
3	Clothing	Crown, Veil, Robes, Armor...
4	Postures	Folded Arms, Hands Clasped, Head Bowed...
5	Jewelry	Crown, Tiara, Necklace, Amulet...
6	Accessories	Hourglass, Skull, Mirror, Fruit...
7	Furniture	Chair, Table, Bed, Mirror...
8	Trees	Oak, Willow, Birch, Cypress...
9	Scenery	Sunrise, Sunset, Stormy Seas, Calm Waters...
10	Animals	Lion, Eagle, Dove, Snake...
11	Flowers	Rose, Lily, Sunflower, Tulip...
12	Buildings	Castle, Tower, Ruins, Church...
13	Weapons	Sword, Dagger, Bow and Arrow, Spear...
14	Weathers	Sunshine, Storm, Rain, Snow...
15	Mythological characters	Pandora, Medusa, Narcissus, Icarus...
16	Shapes	Circle, Triangle, Square, Rectangle...

The proposed methods are employed to identify top K appropriate metaphorical objects for each prompt. Subsequently, *N* images are then generated for each new prompt by leveraging the interfaces provided by existing text-to-image systems [13]. Finally, manual evaluations are performed on these *K*T* generated images, which serves as the ground truth for the study. Evaluation Metrics, and these evaluation jobs are conducted by professionals who have studied art-related courses.

Evaluation Metrics. When incorporating metaphorical objects into the text-to-image process, two aspects need to be deliberately considered. First, whether the metaphorical objects enrich the visual content, making the text-to-image output more diverse. Second, whether the outputs remain consistent with the original prompts and do not distort the original author's intent. Therefore, during the manual evaluation, each output image is assigned two tags: a Diversity Tag and a Consistency Tag, show as below table.

Table 4: Evaluation Metrics for Each Output Image

Tag	Yes	NO
Diversity Tag	1	0
Consistency Tag	1	0

In this experimental setup, the three models above were employed to introduce objects imbued with artistic metaphorical significance into 50 prompts. Subsequently, each modified prompt was utilized to generate three images via the text-to-image API provided by OpenAI. The final stage involved soliciting evaluations from four art professionals, who assessed the generated images based on the criteria of diversity and consistency. In this experiment, there 11,879 images generated using 50 prompts infused with 320 metaphorical objects.

By introducing the measures in IR field, such as Precision, Recall, and F-measure, the results marked with Diversity Tag and Consistency Tag can be measured using Diversity-Precision, Diversity-Recall, Consistency-Precision, etc. In this research, precision received more attention due to the excessive number of generated images.

5.2 Comparison of Two Score Models

The two charts below illustrate the average Diversity-Precision and Consistency-Precision for 50 prompts using the two score models: Maximum Connecting score Model (**MCM**) and Average Connecting score Model (**ACM**).

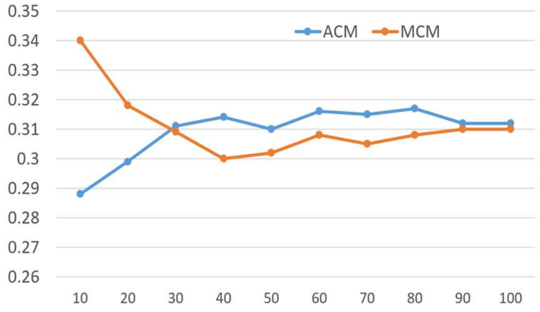


Figure 8: Diversity-Precision for 50 prompts using the two models varies with the number of outputs

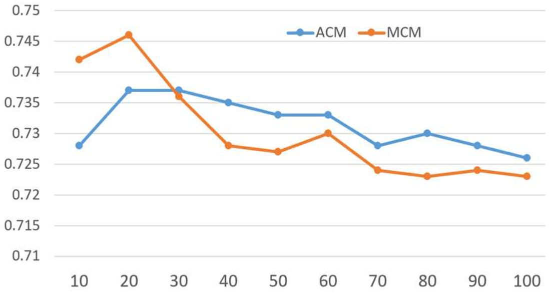


Figure 9: Consistency-Precision for 50 prompts using the two models varies with the number of outputs

From the chart above, it is evident that the two models are capable of generating images with metaphorical meaning. Notably, the ACM method demonstrates superior performance in terms of both diversity and consistency when the output number is below 20. When the output quantity exceeds 20, the

ACM demonstrates strong stability, which is attributed to ACM's consideration of a greater number of concrete phrases in the score calculation process.

5.3 Enhancements through Learning

With the 6 features in Table 1, the LamdaRank method (**LRM**) is employed to improve the diversity precisions. In the experiment, we generated 11,879 images using 50 prompts across 10 categories and conducted a manual evaluation. Due to the relatively small size of the dataset, we trained the model using the labeled results from 9 categories each time, leaving the remaining category for testing. The results are shown in the Figure 10, and comparing with results of MCM.

The average diversity precision of LRM across the 10 categories is 0.36, compared to 0.34 of MCM. Notably, LRM demonstrated a clear advantage in 6 out of the 10 categories. This indicates that the LRM method is effective and enhances the performance of generating metaphorical images.

Moreover, new findings have also emerged. The precisions of the generated metaphorical images show significant variations across different categories. Particularly, the "Architectural" and “Seasons” categories perform the poorest. This can be attributed to the highly specific nature of the prompts in this category, many of which refer to real-world entities, leaving little room for metaphorical interpretation. In contrast, categories like "Portrait," "Abstract," and "Science," which are inherently imaginative, become even richer when metaphors are incorporated.

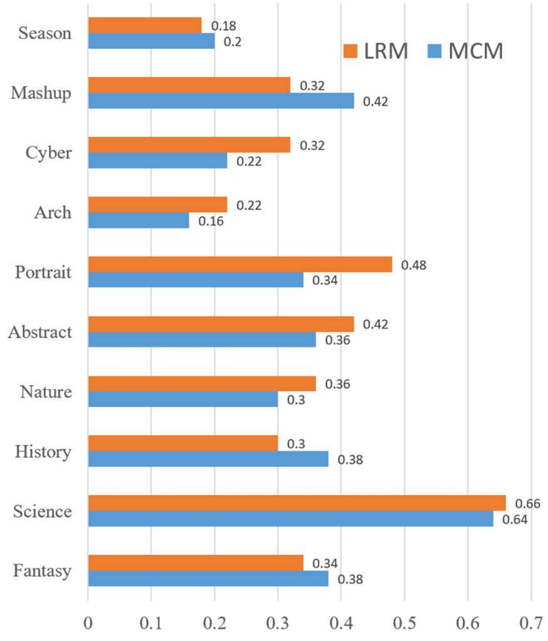
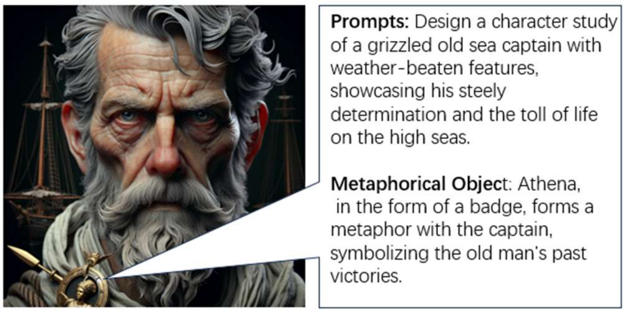


Figure 10. The Diversity-Precisions of MCM and LRM illustrate significant variations across different categories, and the output number is 10.

5.4 Demonstration

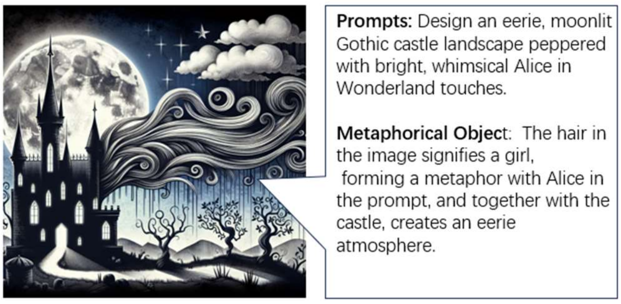
A large number of highly imaginative images were generated in this experiment, which also validates the feasibility of the proposed method, and below are a few examples.



Prompts: Design a character study of a grizzled old sea captain with weather-beaten features, showcasing his steely determination and the toll of life on the high seas.

Metaphorical Object: Athena, in the form of a badge, forms a metaphor with the captain, symbolizing the old man's past victories.

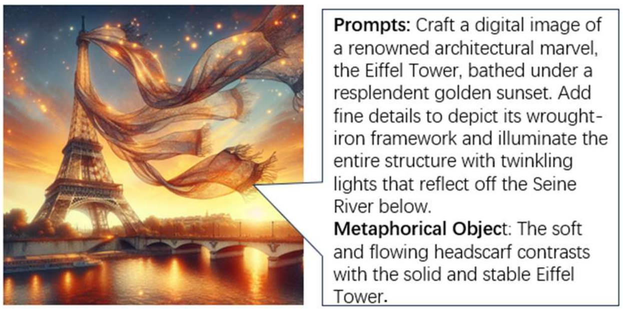
Figure 11: A grizzled old sea captain



Prompts: Design an eerie, moonlit Gothic castle landscape peppered with bright, whimsical Alice in Wonderland touches.

Metaphorical Object: The hair in the image signifies a girl, forming a metaphor with Alice in the prompt, and together with the castle, creates an eerie atmosphere.

Figure 12: An eerie, moonlit Gothic castle landscape



Prompts: Craft a digital image of a renowned architectural marvel, the Eiffel Tower, bathed under a resplendent golden sunset. Add fine details to depict its wrought-iron framework and illuminate the entire structure with twinkling lights that reflect off the Seine River below.

Metaphorical Object: The soft and flowing headscarf contrasts with the solid and stable Eiffel Tower.

Figure 13: The Eiffel Tower bathed under a golden sunset

6. CONCLUSION

In this paper, we integrated and innovatively applied research findings from three fields: constructing an art metaphor corpus based on studies of metaphors in the art domain, transferring methods for generating metaphors from text to text-to-image generation, and employing machine learning techniques to improve precisions. Experimental results demonstrate the effectiveness.

Additionally, the experiments revealed significant differences in the difficulty of constructing metaphors across prompt categories, which merits further exploration.

REFERENCES

[1] A. Kettenmann, Frida Kahlo, 1907-1954: pain and passion. Köln: Taschen, 2013.

[2] S. Malaguzzi and B. Phillips, Food and feasting in art. Los Angeles: J. Paul Getty Museum, 2008.

[3] S. Malaguzzi and C. Mulkai, Bijoux, pierres et objets précieux. Paris: Hazan, Dl, 2008.

[4] L. Impelluso and S. Sartarelli, Gardens in art. Los Angeles: J. Paul Getty Museum, 2007.

[5] L. Impelluso, Thomas Michael Hartmann, and Stefano Zuffi, Gods and heroes in art. Los Angeles: The J. Paul Getty Museum, 2003.

[6] Petrenko V.F and Korotchenko E.A, “Metaphor as a basic mechanism of art (painting),” Psychology in Russia: State of the art, vol. 5, 2014, Accessed: Sep. 02, 2024. [Online]. Available: https://cyberleninka.ru/article/n/metaphor-as-a-basic-mechanism-of-art-painting-1

[7] M. Ge, R. Mao, and E. Cambria, “A survey on computational metaphor processing techniques: from identification, interpretation, generation to application,” Artificial Intelligence Review, vol. 56, no. S2, pp. 1829–1895, Aug. 2023, doi: https://doi.org/10.1007/s10462-023-10564-7.

[8] D. Zheng, R. Song, T. Hu, H. Fu, and J. Zhou, “‘Love Is as Complex as Math’: Metaphor Generation System for Social Chatbot,” Lecture notes in computer science, pp. 337–347, Jan. 2020, doi: https://doi.org/10.1007/978-3-030-38189-9_36.

[9] W. Mo, T. Zhang, Y. Bai, B. Su, J.-R. Wen, and Q. Yang, “Dynamic Prompt Optimizing for Text-to-Image Generation,” Thecvf.com, pp. 26627–26636, 2024, Accessed: Sep. 02, 2024. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2024/html/Mo_Dynamic_Prompt_Optimizing_for_Text-to-Image_Generation_CVPR_2024_paper.html

[10] Y. Hao, Z. Chi, L. Dong, and F. Wei, “Optimizing Prompts for Text-to-Image Generation.” Accessed: Sep. 02, 2024. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2023/file/d346d91999074dd8d6073d4c3b13733b-Paper-Conference.pdf

[11] The Met, “Oedipus and the Sphinx,” Metmuseum.org, 2019. https://www.metmuseum.org/art/collection/search/437153

[12] Gelato, “How to Ceate AI Art Prompts (+ 50 Top Examples),” Gelato, 2024. https://www.gelato.com/blog/ai-art-prompts (accessed Sep. 02, 2024).

[13] “OpenAI API,” platform.openai.com. https://platform.openai.com/ docs/guides /imagesY.

[14] google, “Google Code Archive - Long-term storage for Google Code Project Hosting.,” *Google.com*, 2019. https://code.google.com/archive/p/word2vec/

[15] G. Lakoff and M. Johnson, *Metaphors we live by*. Chicago: University of Chicago Press, 2003.

[16] Christopher J.C. Burges, “From RankNet to LambdaRank to LambdaMART: An Overview,” Jun. 2010, Accessed: Aug. 20, 2024. [Online]. Available: https://www.researchgate.net/profile/Christopher-Burges/publication/228936665_From_ranknet_to_lambdarank_to_lambdamart_An_overview/links/00b49518c11a416a3b000000/From-ranknet-to-lambdarank-to-lambdamart-An-overview.pdf