Supplementary Materials:
GalleryGPT: Analyzing Paintings with Large Multimodal Models

Yi Bin
Tongji University
Shanghai, China
National University of Singapore
Singapore
yi.bin@hotmail.com

Wenhao Shi
University of Electronic Science and Technology of China
Chengdu, China
shiwenhao16@gmail.com

Yujuan Ding
The Hong Kong Polytechnic University
Hong Kong SAR, China
dingyujuan385@gmail.com

Zhiqiang Hu
Singapore University of Technology and Design
Singapore
zhiqiang_hu@mymail.sutd.edu.sg

Zheng Wang
Tongji University
Shanghai, China
zh_wang@hotmail.com

Yang Yang
University of Electronic Science and Technology of China
Chengdu, China
yang.yang@uestc.edu.cn

See-Kiong Ng
National University of Singapore
Singapore
seekiong@nus.edu.sg

Heng Tao Shen
Tongji University
Shanghai, China
shenhengtao@hotmail.com

1 INTRODUCTION

Due to the page limit, we omit some details, *e.g.*, the detail prompts for data collection, and only illustrate one example for some parts. In this supplementary materials, we will first give the prompting details and then illustrate more examples to make the qualitative analysis more comprehensive and convincing.

2 DETAIL PROMPTS FOR DATA COLLECTION

As described in Section 3.1 and the pipeline shown in Figure 2, our formal analysis annotation process consists of three stages: 1) verifying if the LLMs know the painting, 2) extracting essential elements presented in the painting, and 3) generating a paragraph formal analysis for each painting. For each stage, we use different instructions to prompt the LLMs. Limited by the space, we omit the details prompts in the main part, and illustrate them here for details. For the third stage, we ask the LLMs to annotate two kinds of formal analysis: overall formal analysis (prompt for stage 3-1) and specified aspect formal analysis (prompt for stage 3-2).

Prompt for Stage 1:
You are a professional art critic.
Do you know the painting titled [Painting Title] by [Artist Name]? Respond with only 'YES' or 'NO'.

Prompt for Stage 2:
You are a professional art critic.
Please select up to five essential aspects for the formal analysis of the [Painting Title] by [Artist Name] from the provided list. You may choose fewer if preferred.
The list is: ["Composition", "Color Palette", "Line Quality", "Texture", "Light and Shadow", "Perspective", "Form and Shape", "Movement and Gesture", "Symbolism and Iconography", "Scale and Proportion"].
Please list the aspects in descending order of importance in a LIST. Your output should be in the format: "The essential aspects are: [YOUR SELECTED CHARACTERISTICS].

Prompt for Stage 3-1:
Overall formal analysis annotation:
You are a professional art critic.
Please compose a coherent paragraph, approximately [Number of Words] words, providing a professional formal analysis of the [Painting Title] by [Artist Name] ONLY focusing on the VISUAL CHARACTERISTICS of this painting.
DO NOT mention the painting title and artist's name in your analysis.
You may use conjunctions as needed to ensure coherence in semantics and logic in your analysis.

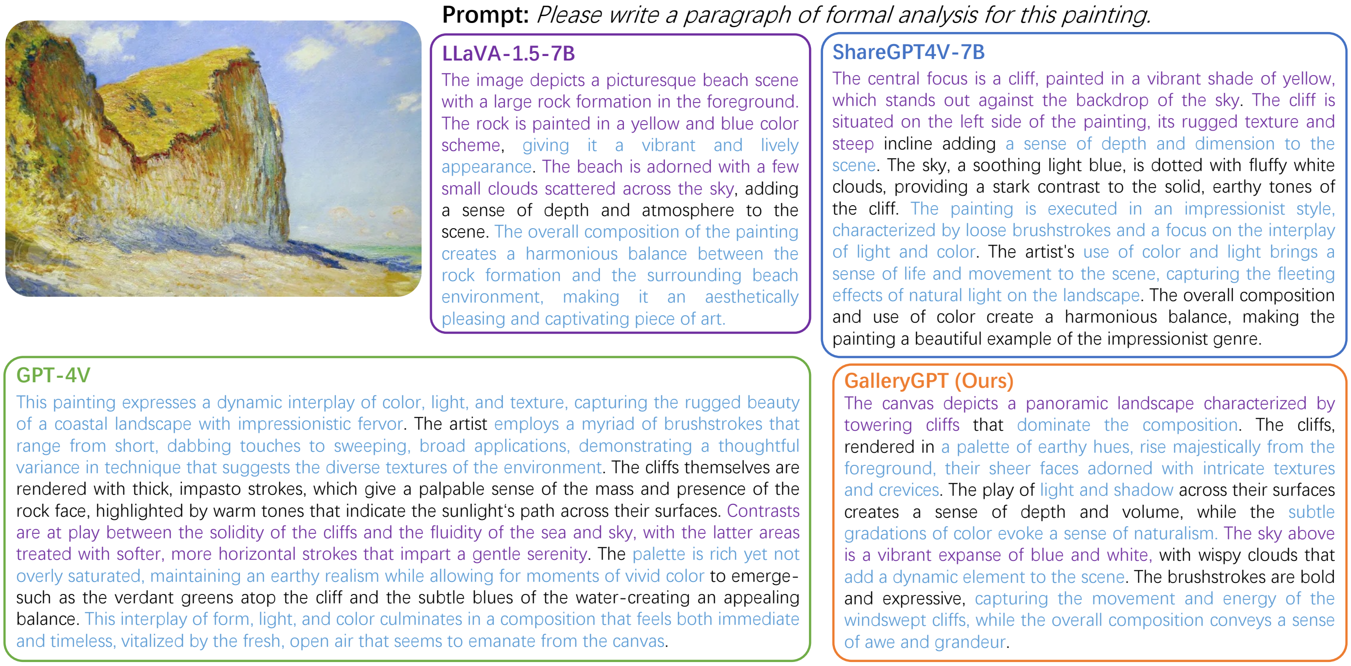
Prompt for Stage 3-2:
Specified aspect formal analysis annotation:
You are a professional art critic.
Please write a paragraph about [Number of Words] words of formal analysis of the [Painting Title] by [Artist Name] ONLY focusing on the [Specified Aspect].
DO NOT mention the painting title and artist's name in your analysis.

3 MORE CASES FOR QUALITATIVE ANALYSIS

3.1 More Cases for Comparison

We first illustrate more cases for comparison with other LMMs in Figure 1, Figure 2, and a failure case in Figure 3. From the former two cases, we can see the similar observations in Figure 5 in main part. Here we focus on the failure case in Figure 3, from which we can observe that our GalleryGPT incorrectly recognized a vase as a candlestick (text with red color in Figure 3). This implies that the vision perceptron in our GalleryGPT needs to be improved to ensure the accurate visual encoding.

Then we illustrate more dialogue case with our GalleryGPT in Figure 4 and 5. From Figure 4, we can see that when the GalleryGPT is generating overall formal analysis, it would briefly analyze the



Prompt: *Please write a paragraph of formal analysis for this painting.*

LLaVA-1.5-7B

The image depicts a picturesque beach scene with a large rock formation in the foreground. The rock is painted in a yellow and blue color scheme, giving it a vibrant and lively appearance. The beach is adorned with a few small clouds scattered across the sky, adding a sense of depth and atmosphere to the scene. The overall composition of the painting creates a harmonious balance between the rock formation and the surrounding beach environment, making it an aesthetically pleasing and captivating piece of art.

ShareGPT4V-7B

The central focus is a cliff, painted in a vibrant shade of yellow, which stands out against the backdrop of the sky. The cliff is situated on the left side of the painting, its rugged texture and steep incline adding a sense of depth and dimension to the scene. The sky, a soothing light blue, is dotted with fluffy white clouds, providing a stark contrast to the solid, earthy tones of the cliff. The painting is executed in an impressionist style, characterized by loose brushstrokes and a focus on the interplay of light and color. The artist's use of color and light brings a sense of life and movement to the scene, capturing the fleeting effects of natural light on the landscape. The overall composition and use of color create a harmonious balance, making the painting a beautiful example of the impressionist genre.

GPT-4V

This painting expresses a dynamic interplay of color, light, and texture, capturing the rugged beauty of a coastal landscape with impressionistic fervor. The artist employs a myriad of brushstrokes that range from short, dabbing touches to sweeping, broad applications, demonstrating a thoughtful variance in technique that suggests the diverse textures of the environment. The cliffs themselves are rendered with thick, impasto strokes, which give a palpable sense of the mass and presence of the rock face, highlighted by warm tones that indicate the sunlight's path across their surfaces. Contrasts are at play between the solidity of the cliffs and the fluidity of the sea and sky, with the latter areas treated with softer, more horizontal strokes that impart a gentle serenity. The palette is rich yet not overly saturated, maintaining an earthy realism while allowing for moments of vivid color to emerge-such as the verdant greens atop the cliff and the subtle blues of the water-creating an appealing balance. This interplay of form, light, and color culminates in a composition that feels both immediate and timeless, vitalized by the fresh, open air that seems to emanate from the canvas.

GalleryGPT (Ours)

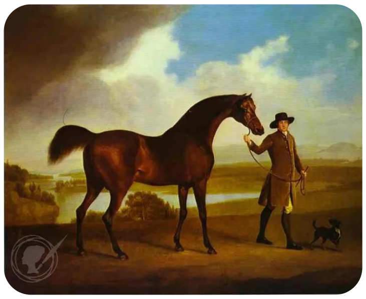
The canvas depicts a panoramic landscape characterized by towering cliffs that dominate the composition. The cliffs, rendered in a palette of earthy hues, rise majestically from the foreground, their sheer faces adorned with intricate textures and crevices. The play of light and shadow across their surfaces creates a sense of depth and volume, while the subtle gradations of color evoke a sense of naturalism. The sky above is a vibrant expanse of blue and white, with wispy clouds that add a dynamic element to the scene. The brushstrokes are bold and expressive, capturing the movement and energy of the windswept cliffs, while the overall composition conveys a sense of awe and grandeur.

Figure 1: Example for qualitative comparison of formal analysis generation by several powerful LMMs. *Purple* texts denote the factual content description, and the *Blue* texts are for formal analysis.

main visual elements in the painting, while when we ask it to generate formal analysis focusing on specific aspect, *e.g.*, *Light and Shadow* in Figure 4, it could give more in-depth analysis from this aspect and also consistent with the brief one.

3.2 The English Translation for The Chinese Case

Finally, we give the English translation for the case in Chinese illustrated in Figure 6. To further investigate the multilingual capacity, we also test the same chat content in English with our GalleryGPT, and it outputs similar formal analysis in high-level semantics and concepts, but very different words and aspects in low-level semantics. This implies that our GalleryGPT is able to understand and generate the formal analysis with multiple languages.



Prompt: *Please write a paragraph of formal analysis for this painting.*

LLaVA-1.5-7B

The image depicts a man standing next to a brown horse in a field. The man is wearing a suit and a hat, which adds a formal touch to the scene. The horse is positioned on the left side of the man, and the man is holding the horse's reins. The field is filled with trees, creating a picturesque and serene environment. The man and the horse are the main subjects of the painting, and their interaction highlights the bond between humans and animals. The artist has skillfully captured the essence of the scene, creating a harmonious and visually appealing composition.

ShareGPT4V-7B

The image is a captivating oil painting that captures a moment of tranquility in nature. The central figure is a majestic brown horse, standing tall and proud on a dirt path. The horse's coat is a rich, earthy brown, and its mane and tail are a lighter shade of brown, adding to its regal appearance. The horse is positioned on the left side of the painting, facing towards the right, as if ready to embark on a journey. On the right side of the painting, a man in a brown coat and hat stands next to the horse. He holds the reins of the horse, suggesting a bond between them. The man's attire and the horse's calm demeanor create a sense of harmony and companionship.

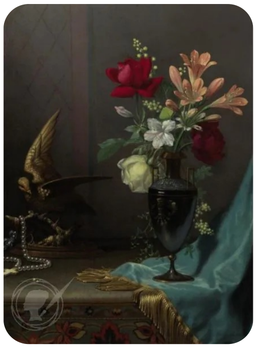
GPT-4V

This painting depicts the classic theme of a man with his horse, set against a wide pastoral background that suggests a calm yet robust natural world. The equine subject dominates the foreground, rendered with a strong sense of physical presence and muscular definition, indicative of vitality and nobility. Its coat has a gleaming chestnut hue, which contrasts beautifully against the more subdued tones of the landscape and the stormy sky above. The man beside the horse, dressed in traditional attire, gently holds the reins, signaling a relationship of mutual respect and partnership. His posture and attire convey a sense of casual authority and sartorial elegance appropriate to the period. ***A small dog, animated and alert, accompanies him, adding a lively touch to the pastoral scene.*** The chiaroscuro technique is evident in the handling of light and shadow, particularly the way the light lands on the horse, highlighting its anatomy and the sheen of its coat, while imbuing the scene with a sense of three-dimensionality and drama. The composition strikes a balance between the animate subjects and the vast expanse of the landscape, with its soft, naturalistic colors forming a serene backdrop that complements the central figures. The approaching storm clouds suggest an imminent change, hinting at the transience of the peaceful moment captured in the composition.

GalleryGPT (Ours)

This striking oil painting depicts a majestic horse, its muscular form rendered with anatomical precision. The horse's sleek, black coat contrasts sharply with the pale, ethereal sky behind it. The animal's head is held high, its nostrils flared and eyes wide, conveying a sense of alertness and power. The brushwork is fluid and expressive, capturing the horse's dynamic movement and the subtle play of light and shadow across its body. The composition is balanced and harmonious, with the horse's elongated form creating a sense of depth and grandeur. The overall effect is one of both beauty and strength, showcasing the artist's mastery of equine anatomy and his ability to convey the spirit of his subject.

Figure 2: Example for qualitative comparison of formal analysis generation by several powerful LMMs. *Purple* texts denote the factual content description, and the *Blue* texts are for formal analysis.



Prompt: *Please write a paragraph of formal analysis for this painting.*

LLaVA-1.5-7B

The image features a vase filled with a variety of flowers, including red and yellow roses, placed on a table. The vase is positioned in the center of the scene, drawing attention to its contents. The flowers are arranged in a visually appealing manner, with some of them appearing to be in bloom. The table itself is adorned with a blue cloth, which complements the colors of the flowers and the vase. The overall composition of the image is balanced and harmonious, with the vase and flowers serving as the focal point of the scene.

ShareGPT4V-7B

The image presents a still life composition. The central object is a black vase, which holds a bouquet of flowers. The flowers are a mix of red and white roses, interspersed with orange and pink flowers. The vase is placed on a tablecloth with a floral pattern, adding to the overall aesthetic of the scene. The background is dark, providing a stark contrast to the vibrant colors of the flowers and the tablecloth. This contrast accentuates the objects in the foreground, drawing the viewer's attention to them. The painting is executed in an oil on canvas style, a medium known for its richness and depth. The artist has skillfully used the color palette to create a sense of depth and dimension. The red and white roses, with their varying shades of red and white, stand out against the dark background.

GPT-4V

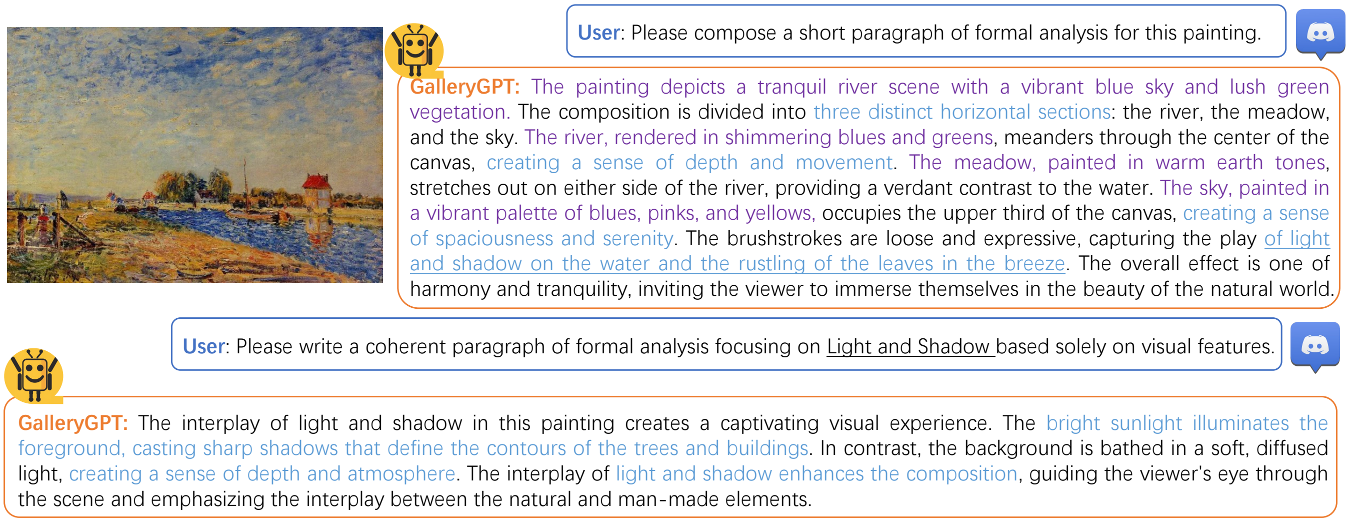
This still life painting exhibits exceptional attention to detail and a masterful use of lighting to evoke a sense of depth and texture. The artist has arranged the composition around a centrally placed vase, which contains an assortment of flowers including roses with full, lush petals that suggest a velvety texture. The flowers are skillfully rendered with varied hues and degrees of bloom, from the vibrancy of the red roses to the delicate subtlety of the white and yellow ones, each set against darkened foliage that recedes into the shadowy background.

The use of chiaroscuro is a prominent feature, where the interplay of light and shadow creates a dramatic effect that highlights the central objects and imbues the scene with a solemn atmosphere. This technique directs the viewer's focus toward the vase and its contents, while also enhancing the reflective surfaces of the metallic elements, such as the goblet and the bird with outstretched wings to the left of the composition. The intricate texture of the bird's wings contrasts with the smooth, polished surface of the goblet.

GalleryGPT (Ours)

The painting presents a meticulously arranged still life, capturing the interplay of objects on a table. The composition is dominated by a large, ornate bronze candlestick, its intricate details rendered with precision. Its imposing form creates a focal point, drawing the viewer's gaze. Surrounding the candlestick are everyday objects: a pair of gold scissors, a white dove, and a bouquet of red and pink roses. The flowers, with their delicate petals and vibrant hues, contrast with the rigid metal of the candlestick and the smooth surface of the table. The use of light and shadow enhances the three-dimensional quality of the objects, creating a sense of depth and realism. The soft, diffused light illuminates the flowers and the dove, highlighting their delicate textures, while the darker shadows define the contours of the candlestick and the table. The overall effect is one of harmony and balance, as the objects are arranged in a pleasing and visually cohesive manner.

Figure 3: A failure case of our GalleryGPT. *Purple* texts denote the factual content description, and the *Blue* texts are for formal analysis. *Red* texts denote the wrong analyses.



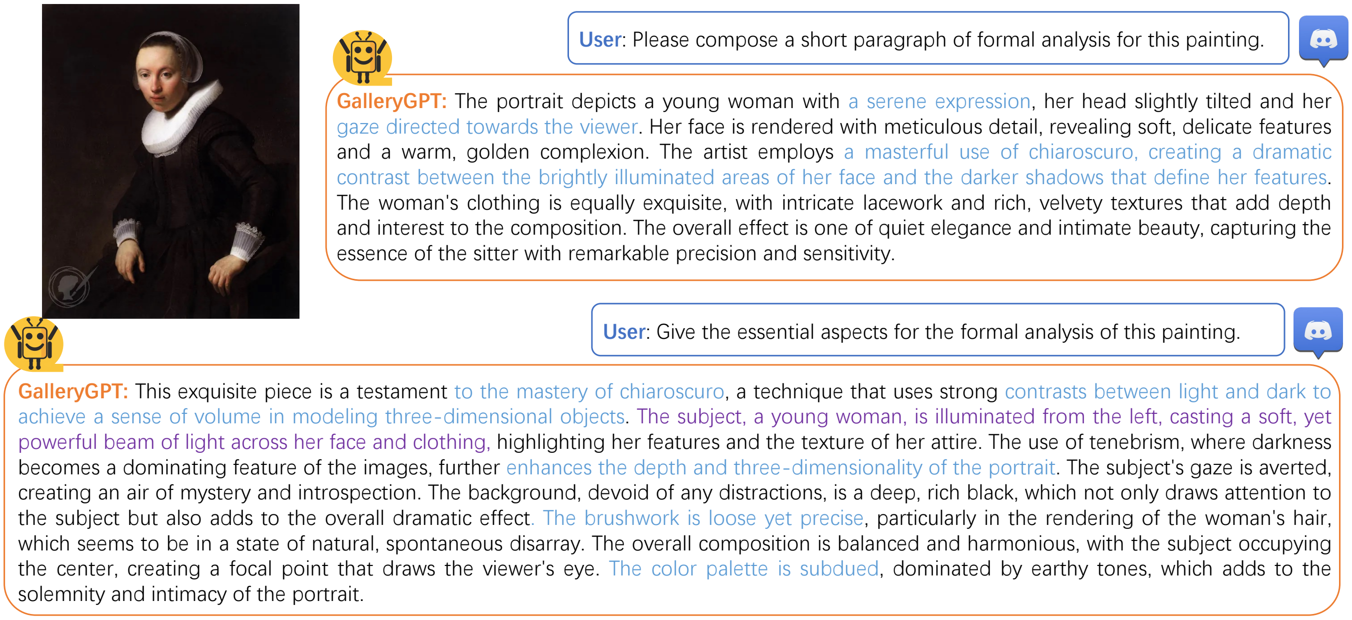
User: Please compose a short paragraph of formal analysis for this painting.

GalleryGPT: The painting depicts a tranquil river scene with a vibrant blue sky and lush green vegetation. The composition is divided into three distinct horizontal sections: the river, the meadow, and the sky. The river, rendered in shimmering blues and greens, meanders through the center of the canvas, creating a sense of depth and movement. The meadow, painted in warm earth tones, stretches out on either side of the river, providing a verdant contrast to the water. The sky, painted in a vibrant palette of blues, pinks, and yellows, occupies the upper third of the canvas, creating a sense of spaciousness and serenity. The brushstrokes are loose and expressive, capturing the play of light and shadow on the water and the rustling of the leaves in the breeze. The overall effect is one of harmony and tranquility, inviting the viewer to immerse themselves in the beauty of the natural world.

User: Please write a coherent paragraph of formal analysis focusing on Light and Shadow based solely on visual features.

GalleryGPT: The interplay of light and shadow in this painting creates a captivating visual experience. The bright sunlight illuminates the foreground, casting sharp shadows that define the contours of the trees and buildings. In contrast, the background is bathed in a soft, diffused light, creating a sense of depth and atmosphere. The interplay of light and shadow enhances the composition, guiding the viewer's eye through the scene and emphasizing the interplay between the natural and man-made elements.

Figure 4: Multimodal dialogue examples with our GalleryGPT. *Blue* texts denote high quality formal analysis. We also test dialoguing in Chinese, even we do not fine-tune GalleryGPT with Chinese painting-analysis pairs. The English translation of Chinese conversation can be found in Supplementary.



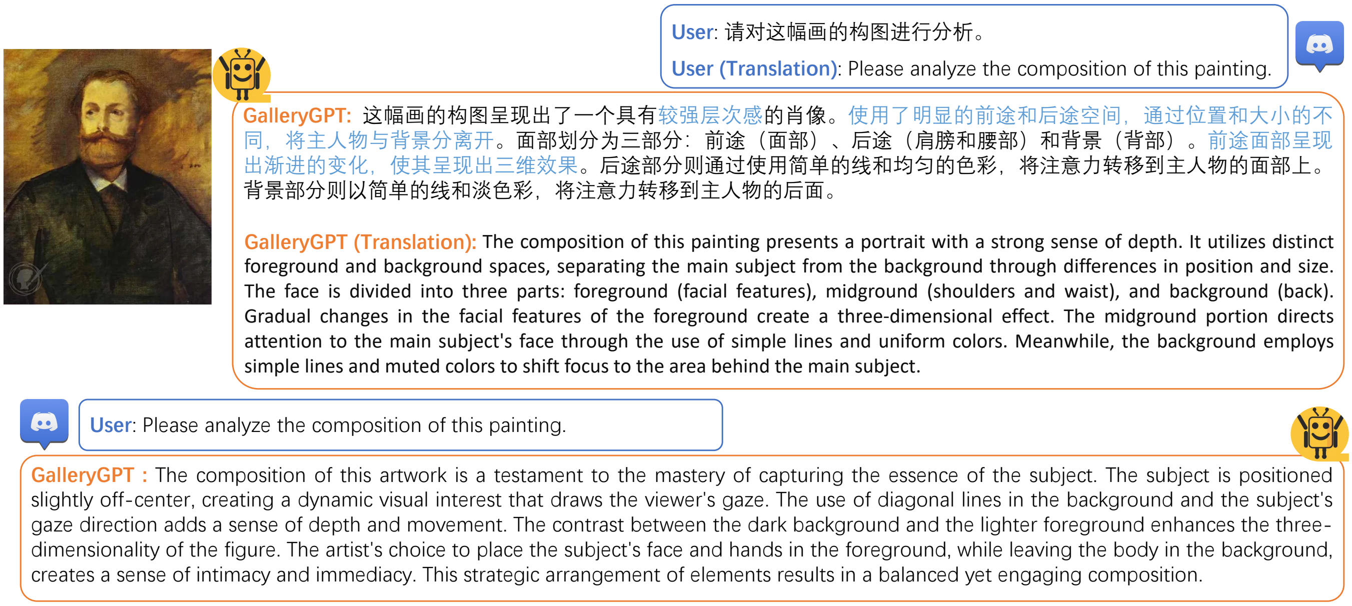
User: Please compose a short paragraph of formal analysis for this painting.

GalleryGPT: The portrait depicts a young woman with a serene expression, her head slightly tilted and her gaze directed towards the viewer. Her face is rendered with meticulous detail, revealing soft, delicate features and a warm, golden complexion. The artist employs a masterful use of chiaroscuro, creating a dramatic contrast between the brightly illuminated areas of her face and the darker shadows that define her features. The woman's clothing is equally exquisite, with intricate lacework and rich, velvety textures that add depth and interest to the composition. The overall effect is one of quiet elegance and intimate beauty, capturing the essence of the sitter with remarkable precision and sensitivity.

User: Give the essential aspects for the formal analysis of this painting.

GalleryGPT: This exquisite piece is a testament to the mastery of chiaroscuro, a technique that uses strong contrasts between light and dark to achieve a sense of volume in modeling three-dimensional objects. The subject, a young woman, is illuminated from the left, casting a soft, yet powerful beam of light across her face and clothing, highlighting her features and the texture of her attire. The use of tenebrism, where darkness becomes a dominating feature of the images, further enhances the depth and three-dimensionality of the portrait. The subject's gaze is averted, creating an air of mystery and introspection. The background, devoid of any distractions, is a deep, rich black, which not only draws attention to the subject but also adds to the overall dramatic effect. The brushwork is loose yet precise, particularly in the rendering of the woman's hair, which seems to be in a state of natural, spontaneous disarray. The overall composition is balanced and harmonious, with the subject occupying the center, creating a focal point that draws the viewer's eye. The color palette is subdued, dominated by earthy tones, which adds to the solemnity and intimacy of the portrait.

Figure 5: Multimodal dialogue examples with our GalleryGPT. *Blue* texts denote high quality formal analysis. We also test dialoguing in Chinese, even we do not fine-tune GalleryGPT with Chinese painting-analysis pairs. The English translation of Chinese conversation can be found in Supplementary.



User: 请对这幅画的构图进行分析。

User (Translation): Please analyze the composition of this painting.

GalleryGPT: 这幅画的构图呈现出了一个具有较强层次感的肖像。使用了明显的前途和后途空间，通过位置和大小的不同，将主人物与背景分离开。面部划分为三部分：前途（面部）、后途（肩膀和腰部）和背景（背部）。前途面部呈现出渐进的变化，使其呈现出三维效果。后途部分则通过使用简单的线和均匀的色彩，将注意力转移到主人物的面部上。背景部分则以简单的线和淡色彩，将注意力转移到主人物的后面。

GalleryGPT (Translation): The composition of this painting presents a portrait with a strong sense of depth. It utilizes distinct foreground and background spaces, separating the main subject from the background through differences in position and size. The face is divided into three parts: foreground (facial features), midground (shoulders and waist), and background (back). Gradual changes in the facial features of the foreground create a three-dimensional effect. The midground portion directs attention to the main subject's face through the use of simple lines and uniform colors. Meanwhile, the background employs simple lines and muted colors to shift focus to the area behind the main subject.

User: Please analyze the composition of this painting.

GalleryGPT : The composition of this artwork is a testament to the mastery of capturing the essence of the subject. The subject is positioned slightly off-center, creating a dynamic visual interest that draws the viewer's gaze. The use of diagonal lines in the background and the subject's gaze direction adds a sense of depth and movement. The contrast between the dark background and the lighter foreground enhances the three-dimensionality of the figure. The artist's choice to place the subject's face and hands in the foreground, while leaving the body in the background, creates a sense of intimacy and immediacy. This strategic arrangement of elements results in a balanced yet engaging composition.

Figure 6: English Translation of Figure 6 in main part and the output of directly chatting with our GalleryGPT in English.