Conservation of AI-Based Artworks
Assignment #6 Experiments with Generative Models
Hee-Eun Kim
Report on Kandinsky 2.1 Diffusion Model
Text-to-Image Generation & Text Guided Image-to-Image Generation
https://huggingface.co/kandinsky-community/kandinsky-2-1

- A brief description of the chosen model and its original purpose:

Kandinsky 2.1 is a text-conditional diffusion model that combines elements of unCLIP and latent diffusion, composed of a transformer-based image prior model, a U-Net diffusion model, and a decoder. This architecture enables high-quality image generation with an alignment to textual descriptions and input image(s).

This model supports three different generation modes: Text-to-Image Generation, Text-Guided Image-to-Image Generation, and Interpolation. The Text-to-Image mode creates images based on user-provided prompts, translating textual descriptions into visual outputs. The Text-Guided Image-to-Image mode enhances an existing image based on a textual prompt and an input image, which can be provided via a URL or an uploaded file. The Interpolation mode blends multiple input images along with text conditions, allowing users to assign specific weights to each input to achieve the most expected results.

- Testing methodology and metrics of success:

For this assignment, I will evaluate two types of modes: Text-to-Image Generation and Text-Guided Image-to-Image Generation. First, I will provide the same text prompt describing a specific artwork but with variations in artistic mediums (e.g., tapestry, ceramics, oil painting, watercolors, engraving, etc.). I will compare the outputs from both modes to determine which one more accurately represents the intended material/medium. For the Text-Guided Image-to-Image Generation mode, I will input an actual image of the artwork via URL and examine how well the mode enhances the image to match the described medium. To further assess the Text-Guided Image-to-Image model, I will test its response to different styles of image inputs (e.g., black-and-white photography, color photography, and sketch drawing) while keeping the text prompt constant. By comparing the outputs, I will evaluate which type of input image best enhances the result and aligns most closely with the textual description.
Lastly, I will test the Text-Guided Image-to-Image Generation mode's ability to restore or reconstruct incomplete or deteriorated artwork. I will provide a fragment or damaged remnant of a work of art as an input image, along with a descriptive text prompt aimed at restoring it to its original state and context. This will help assess whether the mode can fill in missing details, maintain stylistic consistency, and adhere to the historical or artistic intent of the original artwork. This comparative analysis will help evaluate the strengths and limitations of each mode in terms of prompt adherence, material accuracy, visual quality, and overall image coherence.

- Analysis of when the model works best and when it doesn't.

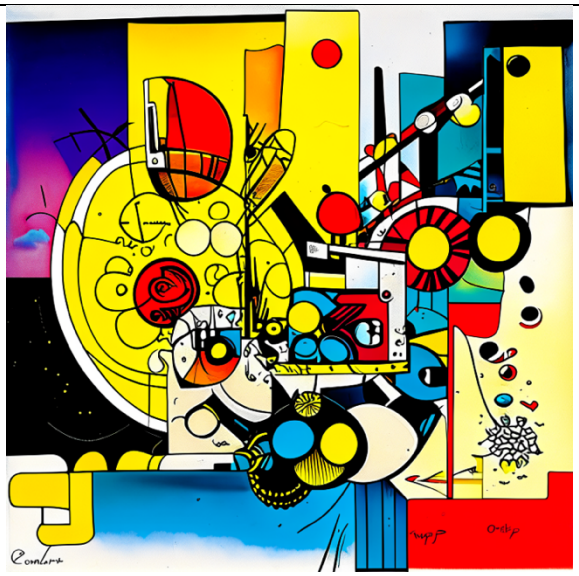
  - Text-to-Image Generation Mode: Accuracy in Representing Different Artistic Mediums

I tested the Text-to-Image generation mode by prompting it to create the same artwork in a distinctive style across 15 different artistic mediums, including kinetic sculpture, oil painting,

digital painting, tapestry, watercolor painting, collage, engraving, fresco, glass, pastel pencils, ceramics, electronic, film, performance, new media. While the model successfully generated 2D representations of traditional mediums such as oil painting, digital painting, and watercolor, it struggled with mediums that inherently exist in three-dimensional space, such as tapestry, new media, and film. Kinetic sculpture, which is the original medium of the artwork was the most accurately illustrated among all the outputs.

Thus, "good" answers to this model would be those where the model produces outputs that logically align with the specific artwork's original medium, especially in 2D format. "Bad" answers would be when the model misinterprets the medium, particularly for 3D, time-based, or interactive art forms in real space.

Some of the examples of the outcomes are as below:

<table>
<tr><td>  </td><td>  </td></tr>
<tr><td>Jean Tinguely's Homage to New York
Medium: Tapestry</td><td>Jean Tinguely's Homage to New York
Medium: New Media</td></tr>
<tr><td>  </td><td>  </td></tr>
<tr><td>Jean Tinguely's Homage to New York
Medium: Kinetic Sculpture</td><td>Jean Tinguely's Homage to New York
Medium: Film</td></tr>
</table>

Conservation of AI-Based Artworks
Assignment #6 Experiments with Generative Models
Hee-Eun Kim

- Text-Guided Image-to-Image Generation Mode: Accuracy in Representing Different Artistic Mediums

I tested the Text-Guided Image-to-Image Generation mode by keeping the text prompt constant while providing an image input of the specific artwork being described. Compared to the Text-to-Image Generation version, the results were more accurate and contextually aligned with the artwork, *Homage to New York*. This mode demonstrated a better ability to generate sculptures in real space, producing outputs with a more three-dimensional appearance, whereas the Text-to-Image model struggled to manifest. However, despite its improved spatial accuracy, the Text-Guided Image-to-Image model still had difficulty rendering certain materials such as Tapestry and Film. Textiles and moving images were misrepresented, appearing as oil paintings or posters rather than capturing their distinct material qualities. While these outputs were still not perfect, they more closely aligned with my intent of describing the artwork in real space compared to the purely text-driven generation mode.

<table>
<tr><td>  </td><td>  </td></tr>
<tr><td>Jean Tinguely's Homage to New York
Medium: Tapestry</td><td>Jean Tinguely's Homage to New York
Medium: New Media</td></tr>
<tr><td>  </td><td>  </td></tr>
<tr><td>Jean Tinguely's Homage to New York
Medium: Kinetic Sculpture</td><td>Jean Tinguely's Homage to New York
Medium: Film</td></tr>
</table>

Conservation of AI-Based Artworks
Assignment #6 Experiments with Generative Models
Hee-Eun Kim

- Same text inputs with Different Styles of Image inputs

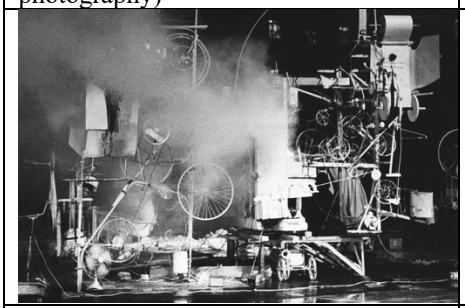
To evaluate the Text-Guided Image-to-Image Generation mode further, I tested how different styles of image inputs affected the results while keeping the text prompt constant. The input images included a schematic drawing of the artwork, a black-and-white documentary-style photograph, and a color photograph of the same artwork. The generated outputs closely followed the tone and manner of the original input image, demonstrating the model's tendency to preserve stylistic elements. Among the tested inputs, the black-and-white documentary-style image produced the most contextually accurate results, best reflecting the original intent of the artwork, which was a performance piece from 1960. However, a key limitation of the model was its tendency to generate a pedestal beneath the sculpture, despite it not being part of the description. This suggests that the model could be biased toward certain artistic conventions, particularly associating sculptures with pedestals or staged exhibition settings. This experiment highlights both the model's strength in maintaining the stylistic characteristics of the input image and its limitations in accurately interpreting unconventional artistic contexts.

<table>
<tr><td>  </td><td>  </td></tr>
<tr><td>Input Image: Schematic drawing</td><td>With an input image of Black and White documentary type photography</td></tr>
<tr><td>  </td><td>  </td></tr>
<tr><td>With an input image of Color Photography</td><td>With an input image of Black and White documentary type photography</td></tr>
</table>

Conservation of AI-Based Artworks
Assignment #6 Experiments with Generative Models
Hee-Eun Kim

- Assessing consistency and variability in repeated outputs

I also tested the consistency of the model's outputs by repeatedly generating images using the same text and image inputs. The goal was to examine how much variation would occur between iterations. The results showed minor differences between outputs, but there was no significant change in artistic style or image background across iterations. This suggests that while the model introduces some randomness, it maintains a high degree of consistency in preserving the overall composition and aesthetic of the input.
To illustrate this, I included two iterated results for each of the three different image inputs, demonstrating the model's tendency to produce similar stylistic interpretations with only subtle variations.

		
Original image input (sketch drawing)	Iteration I	Iteration II
		
Original image input (color photography)	Iteration I	Iteration II
		
Original image input (Black and white documentary photograph)	Iteration I	Iteration II

Conservation of AI-Based Artworks
Assignment #6 Experiments with Generative Models
Hee-Eun Kim

- Restoration of an artwork to its original state

I tested whether the model could restore fragmented artworks or reconstruct a one-time performance by providing an image input of an incomplete piece and a text prompt for restoration. For Jean Tinguely's *Homage to New York*, I initially prompted the model to recreate the March 17, 1960 performance at MoMA's Sculpture Garden. However, the result was incorrectly placed the artwork indoors. After modifying the prompt to explicitly mention "outside" and specifying the site, the output aligned more accurately with the original context. For *Laocoön and His Sons*, I tested whether the model could restore the missing right arm. The result was inaccurate, with the arm randomly positioned and distorted, showing that the model lacks historical awareness and struggles with accurate restorations.
These findings highlight the model's ability to adjust spatial context with refined prompts but also its limitations in restoring missing elements with historical accuracy.

		
Image Input: A Remnant of the performance	Text Prompt: Jean Tinguely's Homage to New York, reconstruct the performance in its original context and state in March 17, 1960 at the Sculpture Garden at MoMa, as Realistic as possible	Text Prompt: Reconstruct the performance, Homage to New York by Jean Tinguely, in its original context and state in March 17, 1960 at the Sculpture Garden (outside) at The Museum of Modern Art, New York City, as closely as possible

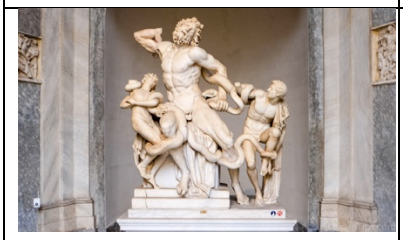
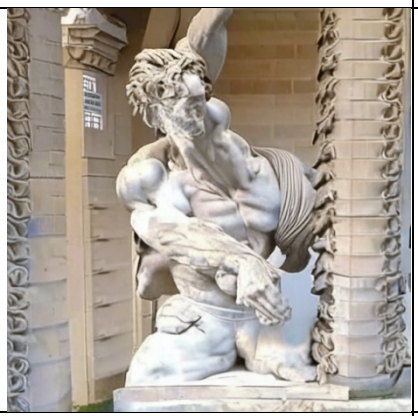

	

Image Input: Fragmented Laocoön and His Sons	Text Input: Restore the full right arm of Laocoon found in Rome at the end of the 15th century into its original state.	Text Input: Restore the damaged statue found in Rome at the end of the 15th century into its original complete state.

- Model provenance: Training and Development of Kandinsky 2.1

The Kandinsky model 2.1 model was developed by Arseniy Shakhmatov, Anton Razzhigaev, Aleksandr Nikolich, Igor Pavlov, Andrey Kuznetsov, and Denis Dimitrov. The training process involved multiple datasets and fine-tuning stages to enhance the model's ability to generate high-quality images based on text prompts. The image prior training was conducted using the LAION Improved Aesthetics dataset, which helped the model refine its understanding of artistic quality. The main Text-to-Image diffusion model was trained on 170 million text-image pairs from the LAION HighRes dataset, with a focus on images of at least 768x768 resolution to ensure high-detail outputs. At the fine-tuning stage, an additional dataset of 2 million high-quality, high-resolution images with detailed descriptions was incorporated. This dataset was compiled from COYO, anime, landmarks_russia, and other publicly available open sources. These curated datasets helped the model improve artistic representation, image-text alignment, and high-resolution image synthesis.