

Conditional Generation for Inverse Problems and Class/Text-Based Conditioning

Adam Imdieke (imdie022@umn.edu)

I. Text to image Generation

A. (A)No Prompt

For the no-prompt generation, I used the same model as in part B, with the parameters $\omega CFG = 0.0$, num steps = 50, and $eta = 0.0$, with a fixed seed at 42. The resulting image is shown below.



Fig. 1. No Prompt Image Generation

B. (B-C) 5X3 prompts- Manual evaluation

I implemented a pipeline that will read from a set of prompts, and then use them to generate the images from the diffusers StableDiffusionPipeline module, with the $\omega CFG = 10$, num steps = 50, $eta = 0.0$, and a fixed seed at 42. The prompts for each image are included in the caption of the figures.

- Space Exploration

- Simple: Human: 5/10, CLIP: 28.49
- Medium: Human: 8/10, CLIP: 32.65
- Detailed: Human: 9/10, CLIP: 33.75

- Underwater World

- Simple: Human: 9/10, CLIP: 31.39
- Medium: Human: 6/10, CLIP: 32.93
- Detailed: Human: 5/10, CLIP: 30.34

- Ancient Architecture

- Simple: Human: 8/10, CLIP: 29.36
- Medium: Human: 7/10, CLIP: 31.98
- Detailed: Human: 9/10, CLIP: 30.38

- Futuristic City

- Simple: Human: 8/10, CLIP: 33.70
- Medium: Human: 8/10, CLIP: 34.95
- Detailed: Human: 7/10, CLIP: 29.07

- Fantasy Creatures

- Simple: Human: 2/10, CLIP: 25.97
- Medium: Human: 2/10, CLIP: 27.57
- Detailed: Human: 6/10, CLIP: 37.96



Fig. 2. Space Exploration. Left: Simple - "An astronaut in space" (Human: 5/10, CLIP: 28.49). Center: Medium - "An astronaut in a spacesuit floating above Earth" (Human: 8/10, CLIP: 32.65). Right: Detailed - "A single astronaut in a shiny white spacesuit drifting serenely against the stars in the sky. There is a silent planet below with swirling clouds and blue oceans, with their ship orbiting in the distance" (Human: 9/10, CLIP: 33.75).

In almost all cases, the most detailed prompts are the best looking images. While the short prompts also did well, the longer prompts aligned better with what I was expecting from the model. I think that in most cases, if you don't know the prompt, the short and long prompts produce images of similar fidelity, which is expected as the model is trained to approach the image manifold similarly, even without a prompt as seen in part A.

The CLIP and the human scores are pretty well aligned in terms of relative changes. While the scale of the measurements was not aligned well, when the human score



Fig. 3. Underwater World. Left: Simple - "A coral reef" (Human: 9/10, CLIP: 31.39). Center: Medium - "A colorful coral reef with tropical fish of and sunlight filtering through the water" (Human: 6/10, CLIP: 32.93). Right: Detailed - "An underwater coral reef that has all sorts of life, with many fish and sharks swimming around. It has bright corals of all colors and shapes, with sunlight filtering through the clear blue water from above." (Human: 5/10, CLIP: 30.34).

changes, the clip score will usually also have a similar change in score, at least in terms of magnitude.

C. (D) Negative Prompts

I chose the prompt: "An astronaut in a spacesuit floating above Earth"



Fig. 4. Ancient Architecture. Left: Simple - "A medieval castle" (Human: 8/10, CLIP: 29.36). Center: Medium - "A lively medieval castle surrounded by a moat and lush greenery" (Human: 7/10, CLIP: 31.98). Right: Detailed - "A beautiful german day, with a large castle made of stone, with a few vines climbing up the spires. The village around the castle is full of life, with people walking around the market." (Human: 9/10, CLIP: 30.38).



Fig. 5. Futuristic City. Left: Simple - "A cyberpunk city" (Human: 8/10, CLIP: 33.70). Center: Medium - "A distopian cyberpunk city, with neon lights and flying cars." (Human: 8/10, CLIP: 34.95). Right: Detailed - "A breathtaking cyberpunk megacity that has bustling streets filled with people and vendors. The skyline has many towering skyscrapers, and there are futuristic flying cars." (Human: 7/10, CLIP: 29.07).



Fig. 6. Fantasy Creatures. Left: Simple - "A cat with a bird body" (Human: 2/10, CLIP: 25.97). Center: Medium - "A chimera with the body of a cat, wings of a bird." (Human: 2/10, CLIP: 27.57). Right: Detailed - "A beautiful chimera creature that has the body of a maine coon cat, with large majestic wings of an eagle." (Human: 6/10, CLIP: 37.96).