Stable Diffusion web UI img2img generation with automatic masking

Sergey Egorov[1, 2*, 3**]

[1]Artifcial Intelligence Research Institute, Moscow, Russia
[2]Skolkovo Institute of Science and Technology, Moscow, Russia
[3]Moscow Institute of Physics and Technology, Moscow, Russia

*sergey.egorov@skoltech.ru **egorov.sa@phystech.edu

Introduction

The img2img generation using Stable Diffusion often faces a problem where the output image, when prompted with new elements, results in a completely different image but retains the original style. However, the goal in many tasks is to retain the original image while incorporating specific changes or new elements based on the prompt. This project addresses this issue trying to apply automatic masking techniques that allow precise modifications to the image, ensuring that only desired areas are altered while maintaining the integrity of the original image.

Purposes

The primary goal is to automate image preprocessing and editing using advanced models and diffusion techniques. This includes applying automatic masking to refine the areas of images that undergo transformation, improving visual quality and user control.

Methods & Models

The project employs the **AUTOMATIC1111** aka **Stable Diffusion Web UI** for img2img generation, providing a comprehensive interface for advanced image manipulation. Key methods and models include:

txt2img and img2img Civitai models

The Web UI integrates with **Civitai**, offering access to a vast collection of open-source pretrained networks. These models enable extensive customization and allow users to select specific styles and functionalities tailored to their needs.

Upscaling

The **SwinIR_4x** model is utilized for upscaling, improving image resolution by enhancing details and sharpness. This model is essential for maintaining high-quality output after modifications.

Captioning

The project uses the **BLIP** model for generating textual descriptions from images, facilitating a deeper semantic understanding. This captioning capability supports further text-based modifications and analysis.

Sampling Methods

The **DPM++** sampling method is employed to optimize the balance between creativity and coherence during image generation. This method is particularly effective for complex modifications, ensuring that the generated images remain true to the original prompt while being visually appealing.

Experiments & Results

1'st Block — Base Image | Style | Face — SwinIR_4x — BLIP img2txt prompt — Additional prompt → Redacted merged prompt

i'th Block — New Base Image | Style | Face — SwinIR_4x — BLIP img2txt prompt — Additional prompt → Redacted merged prompt

Final Image

Code available at