An Introduction to Stable Diffusion's img2img

Jonah Henriksson

If you have been on the bleeding-edge of AI innovations, you may have heard of the advancements of AI-generated imagery. For those unaware, there are AI tools (DALL-E 2, Stable Diffusion, Midsummer, etc.) that can be used to generate new images from a text prompt. However, Stable Diffusion stands out, for its accessibility (the code is open-source, and the model is distributed under a permissive license), but also for its img2img mode, which allows for generating images from prompts, while retaining some aspects of the original image. It is easier to understand from example, so Figure 1 has a few internet memes modified with img2img (See Figure 2 for the original images).

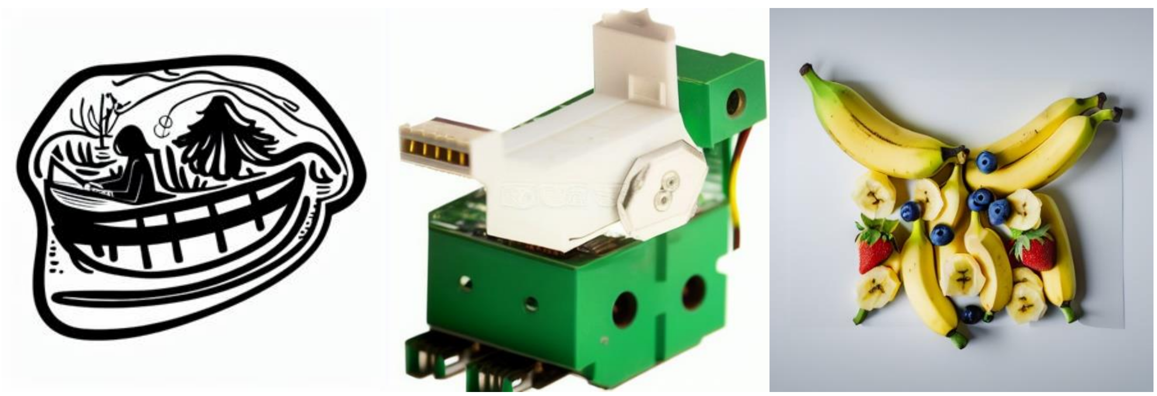


Figure 1: Prompts (shortened for brevity): "a woman sitting on a boat on a lake", "a pile of wires and PCB boards", and "a pile of bananas, strawberries and blueberries", respectively.

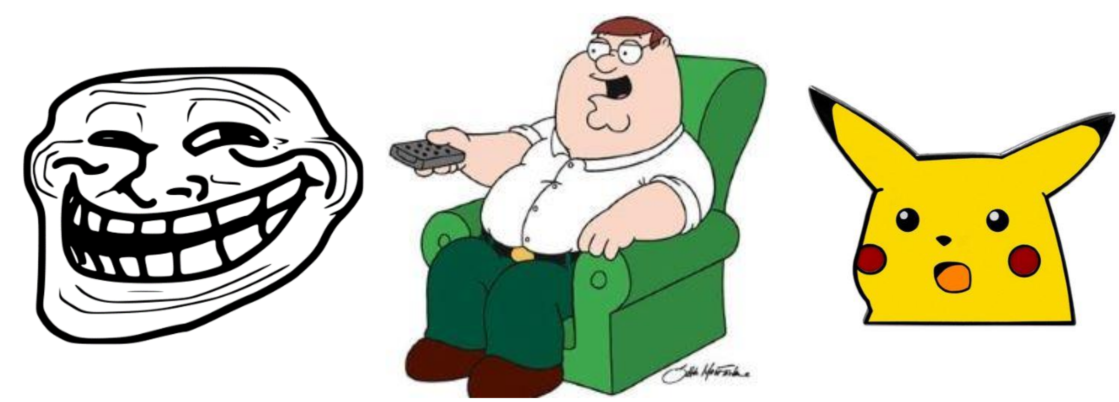


Figure 2: "Trollface" (a cartoon of a grinning man), "Ah Sweet!" (Family Guy character Peter Griffin in an armchair holding a remote) and "Surprised Pikachu" (Pokémon character Pikachu with a surprised expression), respectively.

In this tutorial, we'll go over the process of creating such images, as well as how to install and use Stable Diffusion. It is assumed that you are proficient with Windows and computers (you should know how to open/edit files, run commands and understand terms like VRAM).

Prerequisites

You'll need a powerful PC with 12GB of disk space and an NVIDIA GPU (with 8GB VRAM) running Windows to use this AI. It is possible to run on other operating systems or with other GPUs, but it overcomplicates something that is already complicated as it is.

The user-friendly tool for Stable Diffusion is written in the Python programming language, so you'll have to install it if you don't have it. Stable Diffusion uses version "3.10" (direct download here, run this installer and follow its instructions). You can verify your installation via `python --version` in your console (open "PowerShell" or "Command Prompt" and run this command, it should print the python version, which should be "3.10").

Installation

The Stable Diffusion tool can be downloaded from the GitHub repository, AUTOMATIC1111/stable-diffusion-webui (direct download here, extract to an accessible directory). Before you can run it, you need to download the AI model. This is a large file (>4GB), so it's not distributed with the rest of the code. You can get it from the official source here (you will need to make an account and agree to their terms before downloading "sd-v1-4.ckpt"). You can also download it directly from an unofficial source here (at your own risk!). Either way, put it in the "models" folder of the tool's folder and rename it to "model.ckpt" ("stable-diffusion-webui/models/model.ckpt").

Once you have the tool and model setup, simply run "webui-user.bat" (it should open a console) to run Stable Diffusion! It will take a while for it to start up, but eventually you should have a screen like Figure 3 (it may look be different, since I already installed Stable Diffusion).

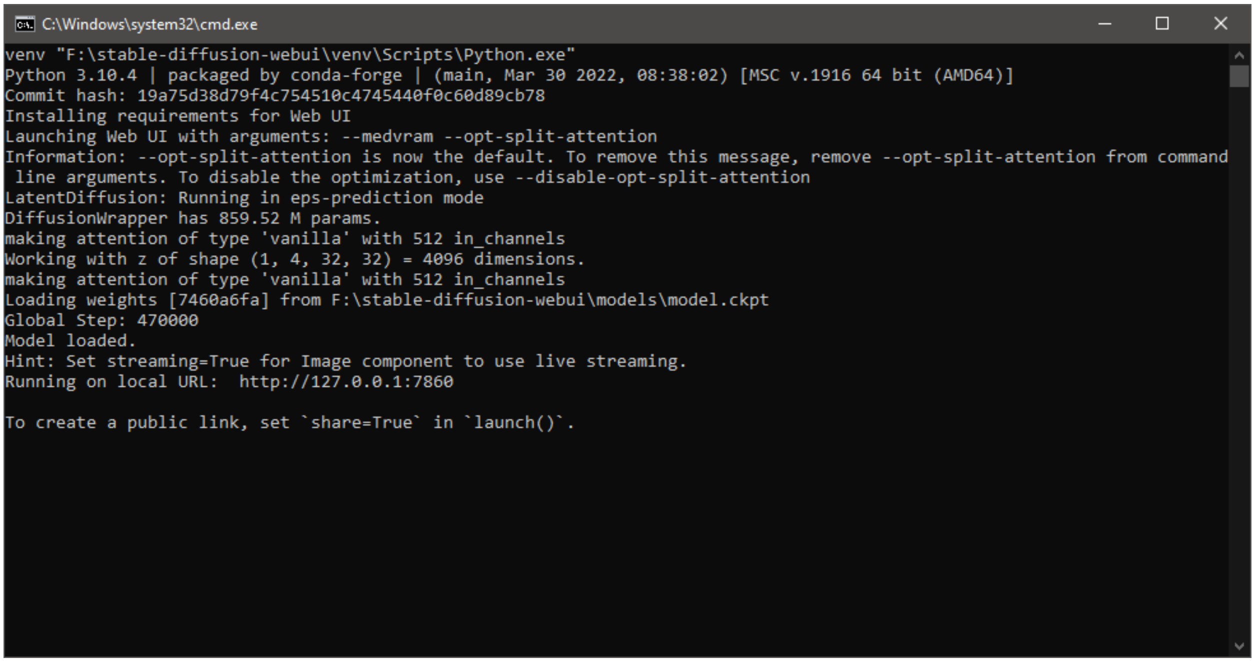


Figure 3: Stable Diffusion ready to run!

Usage

You can then open the URL printed to the screen in your browser (http://127.0.0.1:7860) to open the tool. You should see the screen in Figure 4 in your browser.

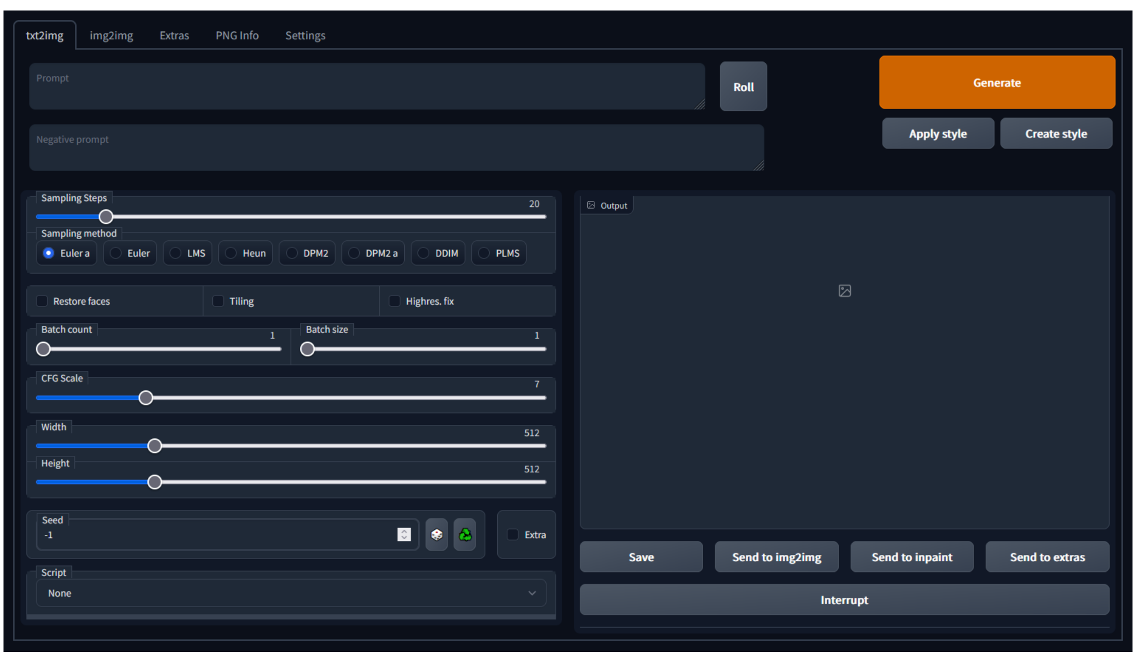


Figure 4: Stable Diffusion WebUI on the txt2img tab.

It is currently on the txt2img tab, which is used to create images from scratch using a given prompt. Let's first try it out! In Figure 5, I type in the prompt "a red car" and click "Generate". You might notice that if you try it that you get a different result, that's because "Seed" (bottom left) is set to "-1" which means it uses a random seed as the starting point. In the bottom right, you can see the seed it generates is "3758558171". You can try to put in the same seed to get the same result (also needs the same exact settings and might not be deterministic).

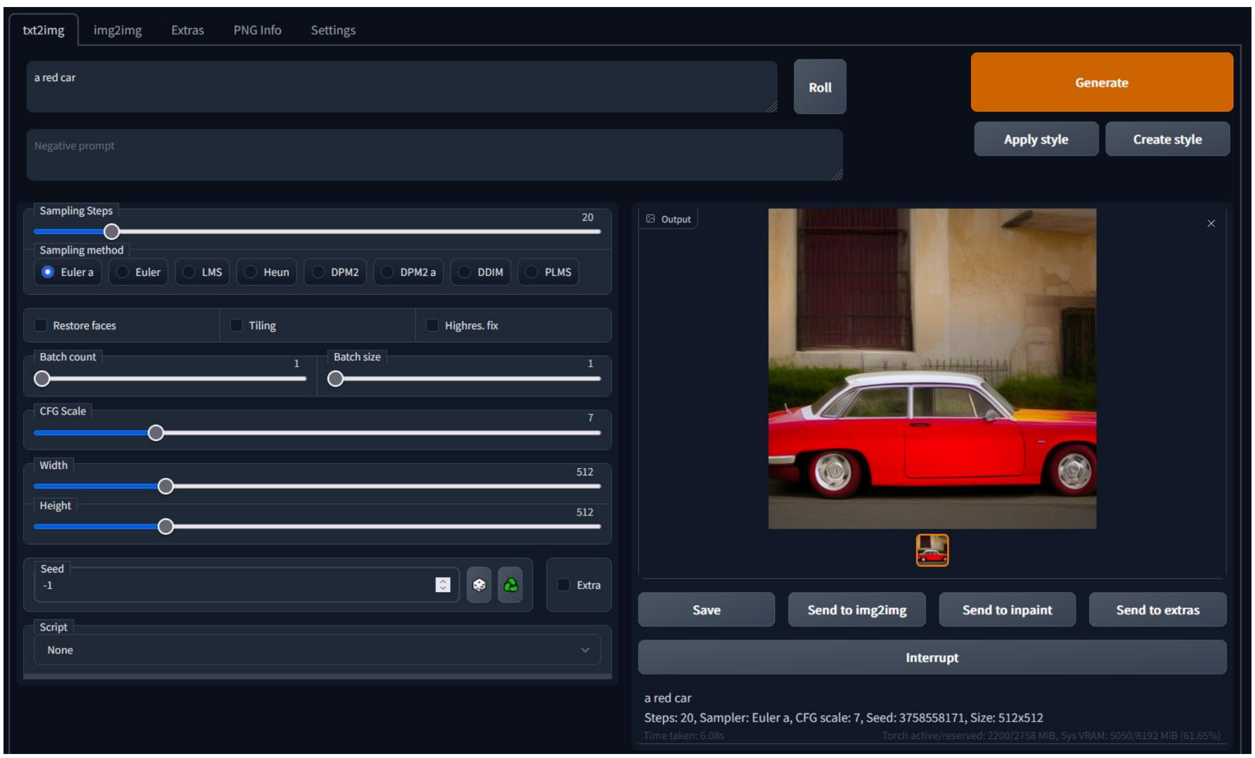


Figure 5: a red car

Anyways, let's get onto img2img. For this, just click on the "img2img" tab on the top. You will then see the screen in Figure 6 (zoomed out).

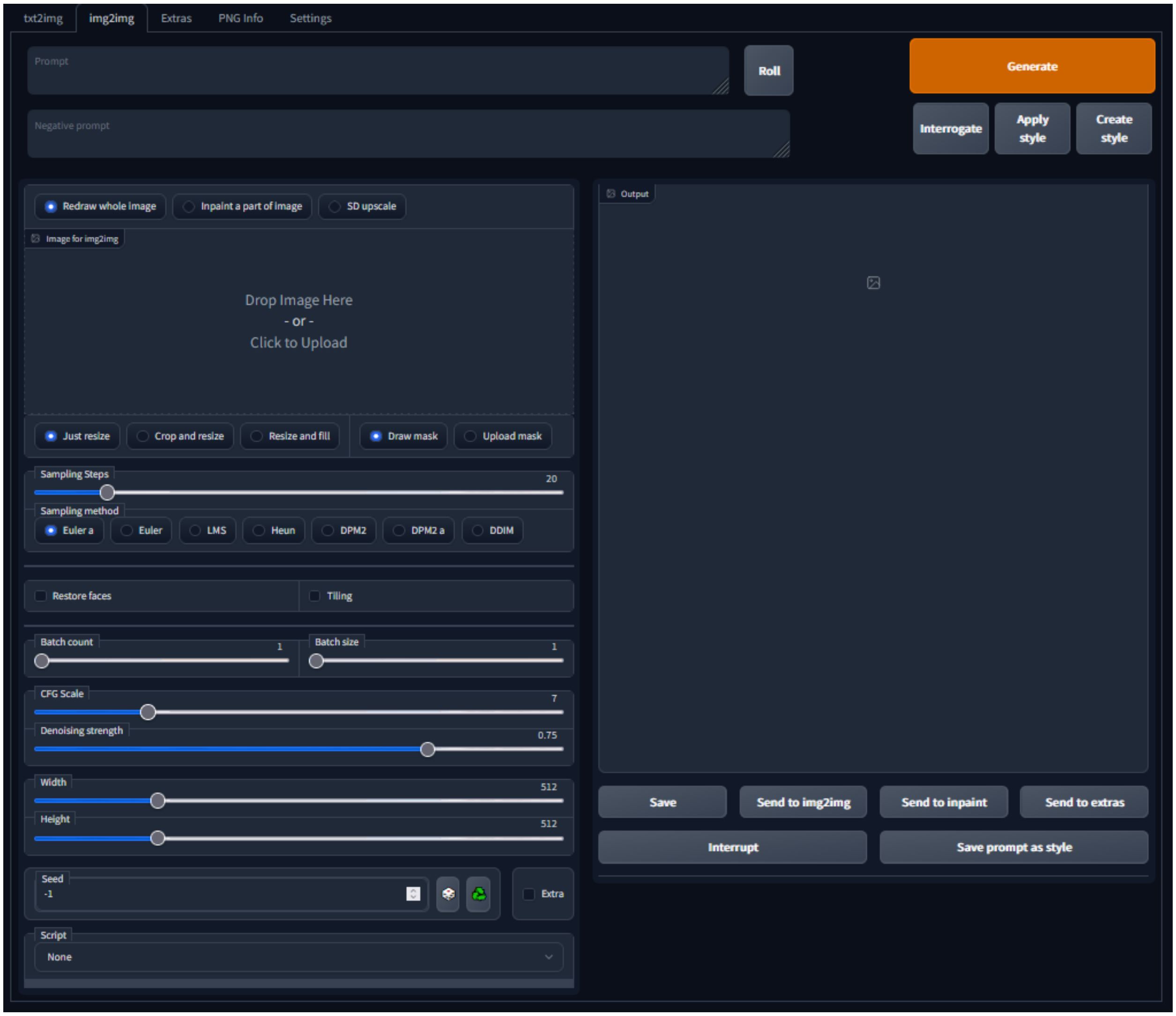


Figure 6: Stable Diffusion's img2img

Before you can use img2img, you need an image to use as the base. When making images that retain abstract aspects of the original, memes work well, since they stay recognizable. It's also good if the image is visually simple, since it allows for the AI to make changes without losing a ton of information. For this tutorial, let's use the "Surprised Pikachu" meme in Figure 2.

Save the image (if you can't, you can find it online, I just cropped this one to a square) and open it up in img2img. In the prompt, we need to figure out a prompt and settings to get a good result. This takes time and practice to learn and get right, so don’t worry if your first results are bad (it may help to test your prompt out in txt2img to get an idea of how the AI interprets it).

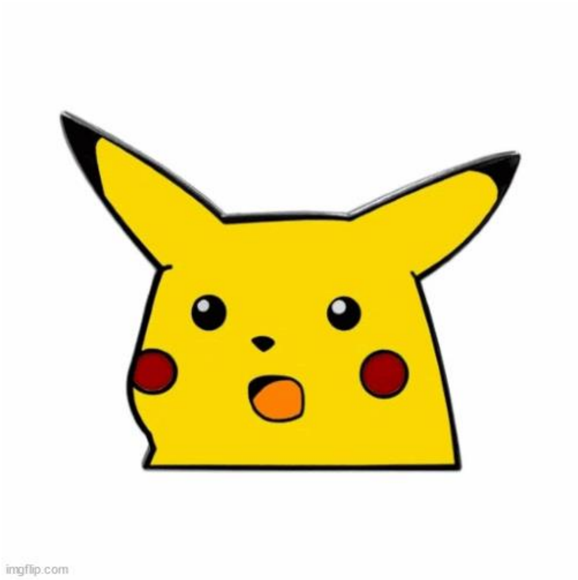
There are a few things we can do:

1. Set “loopback” mode: Under the “Script” dropdown menu, select “Loopback”. This feeds the resulting image back to img2img (in a “loopback”). This allows for making incremental changes to the image until we get something we like, which helps when you want abstract images. I set the “Loops” to 24 (this also means it takes 8 times as long).
2. Lower the “Denoising strength”: This controls how much of the original image is kept. “0.0” will give the exact same image, while “1.0” gives a completely different image. I tried setting it to “0.39”, so only 39% of the image changes per loop.
3. Raise the “Sampling Steps”: The AI produces better images with a higher step count, but also takes more time. I set this to 64 steps.
4. Mess with the “CFG Scale”: This determines the AI’s creativity; lower values make the AI more creative, but higher values produce results more accurate to the prompt. I tried setting this to “12”.

I finally decided on the prompt, “A pile of yellow (bananas), red strawberries and [blueberries] on a white background, professional (((photograph))), studio lighting” (the parentheses tell the AI to give more focus to that part of the prompt, while brackets take away focus). It also has the negative prompt “smile”, this tells the AI to avoid smiles in the generated image. The seed I got was “4042391893”.

The result I got from this was this cute picture of Pikachu as a pile of fruit:

I was able to also generate this small timelapse from the iterations:



If you need more information, check out the documentation at:

https://github.com/AUTOMATIC1111/stable-diffusion-webui/wiki/Features