Stable Diffusion AMD Linux ROCm

This repository contains instructions on how to host your own AI for image generation using stable diffusion with an 8GB VRAM AMD GPU on Linux. This is an affordable and efficient alternative to using Google Colab, which can be quite expensive.

Watch this YouTube video to learn how to install stable diffusion and make it work on your AMD GPU using ROCm. Please note that each GPU is unique, and the launch parameters required may vary. However, the launch parameters used in the video are as follows:

```
--no-half --always-batch-cond-uncond --opt-sub-quad-attention --medvram --disable-nan-check
```

For a complete list of launch parameters, check out the Optimizations wiki.

If you want to download my VAE, you can.

Prerequisites

To get started, you'll need the following:

- Linux OS (a Debian-based distro is recommended)
- AMD GPU (8GB or more) (You may try with less VRAM)
- Python 3

Installation

1. Download the driver for your AMD GPU from AMD's website.
2. Add yourself to the render and video groups using the following commands:

   ```
   sudo usermod -a -G render yourusername
   sudo usermod -a -G video yourusername
   ```

3. Confirm that you have Python 3 installed by typing the following command into the terminal:

```
python3 --version
```

4. Install ROCm by running the following command:

```
sudo amdgpu-install --usecase=rocm --no-dkms
```

5. Reboot your system using the following command:

```
sudo reboot
```

6. After rebooting, confirm that your GPU is recognized by running the following command:

```
rocminfo
```

7. Clone the stable diffusion GUI repository:

```
sudo apt-get install git
git clone https://github.com/AUTOMATIC1111/stable-diffusion-webui
cd stable-diffusion-webui
```

8. If you have Python 3.8 installed, make sure you have VENV capabilities by running the following command (replace with your Python version if necessary):

```
apt install python3.8-venv
```

9. Install pip3 and wheel, and update them using the following commands:

```
sudo apt install python3-pip
python -m pip install --upgrade pip wheel
```

10. Download any stable diffusion model you like and put it in the `models/Stable-diffusion` folder. You can find models at CIViTAI, which is also a great source of prompts.
11. For better performance, upgrade to the latest stable kernel by running the following commands:

```
sudo apt-get update
sudo apt-get dist-upgrade
```

12. Reboot your system again using the following command:

```
sudo reboot
```

13. Go to the virtual env of SD:

```
cd stable-diffusion-webui
python -m venv venv
source venv/bin/activate
```

14. Install the PyTorch machine learning library for AMD:

```
pip3 install torch torchvision torchaudio --extra-index-url
https://download.pytorch.org/whl/rocm5.4.2
```

Note: This is to be installed in the VENV, not on the OS!

15. After installation, check your version numbers with the command:

```
pip list | grep 'torch'
```

The output should show Torch, torchvision, and torchaudio version numbers with ROCM tagged at the end. image

16. Optimize VRAM usage with `--medvram` and `--lowvram` launch arguments. Use `--always-batch-cond-uncond` with `--lowvram` and `--medvram` options to prevent bad quality. If your results turn out to be black images, your card probably does not support float16, so use `--precision full` (at the cost of more VRAM).
17. Benchmark different options. The options that enabled generating nice images (1024x1024) upscaled to 4K were:

```
--no-half --always-batch-cond-uncond --opt-sub-quad-attention --medvram --disable-nan-check
```

Launch the command:

```
python launch.py --opt-sub-quad-attention --medvram --disable-nan-check --always-batch-cond-
uncond --no-half
```

Note: These options may differ for other graphics card models.

The time taken to generate 1024x1024 img2img was 1m 16.33s, and 1024x1024 hires fix was 1m 39s. Generating base images (512x512px) takes 10-20s.

19. If something does not work, check the Torch, torchvision, and torchaudio version numbers with the command:

```
pip list | grep 'torch'
```

The version numbers should have ROCM tagged at the end.

Usage after install

Every time you want to launch stable diffusion, go back to the venv where all dependencies are installed with the following commands:

```
cd stable-diffusion-webui
python -m venv venv
source venv/bin/activate
python launch.py --opt-sub-quad-attention --medvram --disable-nan-check --always-batch-cond-uncond --no-half
```

Watch out for VRAM usage and system temps with the following commands:

```
sudo radeontop
watch -n 1 sensors
```

Adjust your fan curve if your temps are too high (70C).

In this youtube video, I show the process of generating images with high details and a 4K end size using the same GPU (8GB).

Revision #4
Created 2022-09-12 20:15:02 UTC by tinfoil-hat
Updated 2023-06-21 09:11:00 UTC by tinfoil-hat