Dummy proof guide to get SD 2.x working (and other 2.x based models)

Quick tl;dr

1. Download the SD 2.x model
2. Download its corresponding .yaml.

- If it's a custom model it might already come with its own .yaml, download it from the same place you downloaded the custom model.
- If it's base SD 2.x or the custom model doesn't have a .yaml you can get it from this link:
- https://github.com/Stability-AI/stablediffusion/tree/main/configs/stable-diffusion
- v2-inference-v.yaml is for 768 models
- v2-inference.yaml is for 512 models
- **Don't just right click and save a .yaml or it will most likely save it as .html, which will break it.**
- In Hugging Face, simply click the tiny arrow next to the .yaml's filesize to download the .yaml.
- Alternatively, click the .yaml's link to open it, find and click an option that says RAW, then right click and download that as .yaml, or copypaste that text into Notepad and save that as .yaml. In Github you can download the repo as .zip and extract the .yaml
- If your .yaml is 2kb, it's good
- If your .yaml is 70kb , you fucked up and it saved as .html.

3. The model and the .yaml need to be in the same folder and have the same name:
   Model: peepoo.ckpt or peepoo.safetensors
   .yaml: peepoo.yaml
4. If you see the Terminal/CMD start downloading something when you load the 2.x model, you did everything right, it should be downloading the new CLIP and whatever else it needs for new 2.x based models. It should only need to do this once.
5. ??????
6. Prooompt

- Old SD 1.x based embeddings won't work with 2.x models and viceversa. As of updating this, the UI now loads and unloads embeddings as needed so you don't need to worry. Read the descriptions where you downloaded the embedding to find out if it's an SD 1 or SD 2 embedding, or just to see which ones are loaded and which ones are skipped in the Terminal/CMD.

Prerequisites

This guide assumes you already have a working install of the UI, there are better guides for that, like this one: https://rentry.org/voldy. This guide also assumes you're using Windows, if you're using Linux you probably already know what you're doing anyway. (yes I copypasted this from the safetensors guide).

If you already have a working install, but maybe haven't updated your UI in a while, getting 2.x support is a simple as:

Open cmd or the Terminal in your main "stable-diffusion-webui" folder
Type `git pull`
Then type `pip install -r requirements.txt`
Wait for stuff to download and install

If you still get errors even after doing everything in this guide, come back to this step and do this too:

Open cmd again in your main "stable-diffusion-webui" folder
Type `venv\Scripts\activate`
Then type `pip install -r requirements.txt`

This will download/update the required dependencies inside the Virtual Environment (venv) which can fix issues. Sometimes it's needed, sometimes it's not, it doesn't hurt just in case.

Download the models.

Assuming all past, current and future models get uploaded and kept on the same site, and the weird 1.5 release situation doesn't repeat itself,

models should be found on the main Stability AI Hugging Face page.

https://huggingface.co/stabilityai

This is so I can be lazy and not update this for every 2.x release.

EMA vs Non EMA:
"Some folks think non-EMA is for inference, while EMA is for fine-tuning. Use EMA weights for generation! You can use non-EMA for fine-tuning but probably fine to use EMA for fine-tuning too. EMA stands for exponentially moving average. It's a trick where the model weights are average over some window of multiple steps to "smooth" out some of the noise in model updates during training. Researchers find EMA models have better fidelity and less artifacts. For example a paper from 2020 demonstrates EMA models gave more diverse colors in the images"

https://twitter.com/iScienceLuvr/status/1601011140934664193

tl;dr use EMA

Pruned: means extra additional leftover unnecessary data that's not needed to generate images got removed. To elaborate, that data means if a weight for say peepeepoopoo after training equaled 0 or 0.0000000000000000000000000000000000000001, then it pretty much does nothing and was safe to remove. All 0 or basically 0 weights get removed when pruning, since they have no effect in prompts.

2.1 512 model links :
https://huggingface.co/stabilityai/stable-diffusion-2-1-base/tree/main
2.1 768 model links:
https://huggingface.co/stabilityai/stable-diffusion-2-1

Download their corresponding .yaml

The .yaml are all here:

https://github.com/Stability-AI/stablediffusion/tree/main/configs/stable-diffusion

v2-inference.yaml is for the 512 model.
v2-inference-v.yaml is for the 768 model.

The one with the "v" is for the 768 model.

For 2.1 they're the exact same .yaml as for 2.0. This is assuming all possible future .yaml files also get uploaded on the same repo. Check their timestamp in relation to the release of future 2.x models to determine if you'll need to download new corresponding .yaml files for future versions.

Note that if you right click and save, it'll save them as .html. Best way to make sure that doesn't happen is to just download the repo as a .zip, then only unzip the .yaml. Go here, click the big green Code button and select "Download ZIP"

https://github.com/Stability-AI/stablediffusion/

Alternatively, you can click on the .yaml in the Github, select Raw, then right click and save that.

Place models and .yaml in stable-diffusion-webui\models\Stable-diffusion

Place both models and .yaml in the usual models folder, they both have to be in the same folder.

Rename the .yaml to have **the same name as the corresponding .ckpt**. If you've done VAE renaming in the past back when it was needed, it's similar to that.

Example with default names:

v2-1_512-ema-pruned.ckpt is the current 2.1 512 model
Its .yaml by default is named **v2-inference.yaml**
Thus it needs to be renamed **v2-1_512-ema-pruned.yaml**

v2-1_768-ema-pruned.ckpt is the current 2.1 768 model
Its .yaml by default is named **v2-inference-v.yaml**
Thus it needs to be renamed **v2-1_768-ema-pruned.yaml**

Model name: **v2-1_512-ema-pruned.ckpt**
.yaml name: **v2-1_512-ema-pruned.yaml**

Model name: **v2-1_768-ema-pruned.ckpt**

.yaml name: **v2-1_768-ema-pruned.yaml**

You can rename them whatever you want, just make sure they have the same name with their corresponding extension:

SD2.1.ckpt = SD2.1.yaml
peepeepoopoo.ckpt = peepeepoopoo.yaml

Note on the 768 model: as of writing this, the 768 model seems to only output black images, unless you add `--no-half` to `set COMMANDLINE_ARGS=` in the webui-user.bat. Note that this increases VRAM usage and is slower in general.

However, using xformers (by adding `--xformers` to `set COMMANDLINE_ARGS=`) seems to also prevent black images, so it's the better option since xformers also increase performance, albeit with the trade off of very minor changes when trying to reproduce a seed.

For future releases I'd stick to the 512 model and just wait until a new update gets merged to fix it. 768 as a whole is different, since as the name implies, it was trained on a different resolution than the usual 512x512 that's currently the "standard", so it'll be fucky wucky until higher resolutions become standard.

Launch the UI and test them out

If everything worked out, they should load as normal like any other model. Note that embeddings made with previous models won't work with 2.x, if your prompt includes them, you'll get an error. You'll need embeddings made specifically with/for 2.x to use them with 2.x models, though those embeddings will likely not work for previous models either.

Depth model

Already working in the UI, there are better tutorials on Youtube about it. Same process to get it working, pair it with its .yaml.

https://huggingface.co/stabilityai/stable-diffusion-2-depth

Extra models that don't currently work in the UI:

4x upscaler
https://huggingface.co/stabilityai/stable-diffusion-x4-upscaler

Inpainting:
https://huggingface.co/stabilityai/stable-diffusion-2-inpainting