Guide to convert .ckpt models to .safetensors directly with Voldy (AUTOMATIC1111)'s UI

What are .safetensors in the first place?

You can read more about them in full detail here and further down the guide:
https://github.com/huggingface/safetensors

Basically, like their name implies, they're a much safer file format for distributing models. .ckpt are pretty much .zip files with Python code inside, exploiting them is as easy as unzipping them, inserting totallynotavirus.py or adding a totallynotavirus line inside the existing code and zipping them back. Meanwhile .safetensors just contain the necessary data for generation and are harder to exploit.

Why were .ckpt even used if they're so unsafe? Because it was researcher only tech until some months ago, only meant to be used in their controlled labs and not for wider distribution at first, also lolPython is faster to get working projects going on, regardless of security concerns.

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.

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

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

That's it, once everything is downloaded just launch the UI as usual. If you still get errors:

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

This will download/update the required dependencies inside the Virtual Environment (venv) which can fix issues.

Where do I place .safetensors? Do I need to rename them?

They go in the same folder all regular models go: stable-diffusion-webui\models\Stable-diffusion.

Just like you don't need to rename a .rar into a .zip, you don't need to rename them. Just use them as normal.

Note before converting .ckpt to .safetensors

In order to convert .ckpt to .safetensors, the data inside the .ckpt needs to be read and loaded first, which means potential bad pickles (malicious code) are also loaded. To prevent bad pickles, it's better to use conversion methods that go through the UI, since the built in pickle checker should catch any bad pickles before converting them. You should probably still scan sketchy .ckpt models with dedicated pickle checkers before converting them, just in case.

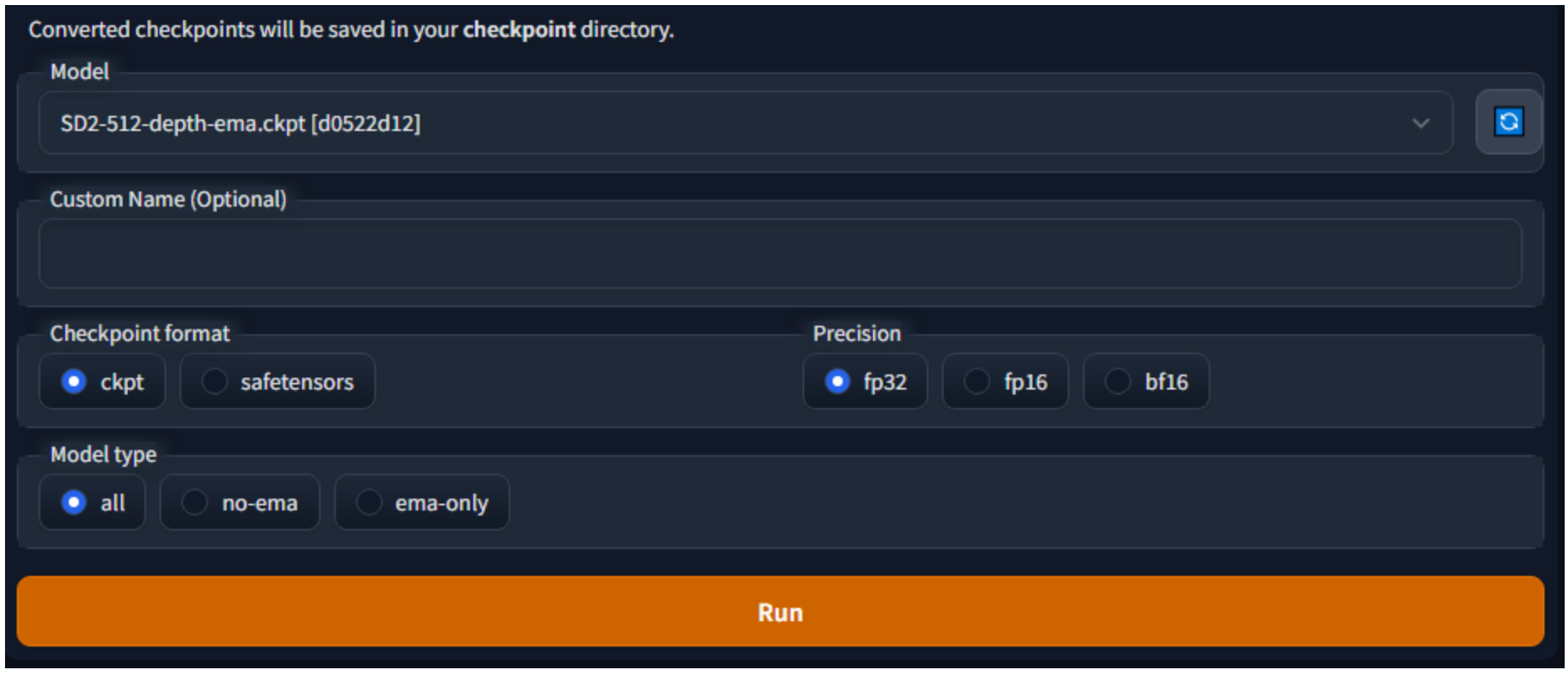
https://github.com/lopho/pickle_inspector
https://github.com/zxix/stable-diffusion-pickle-scanner

How to convert your .ckpt model to .safetensors using the Model Converter extension

Recently a new extension came out that simplifies the conversion process even more. You can find it in the extensions tab under Model Converter and also in their github link:

https://github.com/Akegarasu/sd-webui-model-converter

The extension is self explanatory



1. Pick your model from the dropdown
2. Give it a custom name if you want. If you don't, it'll use the name of the model + extra details like FP32, subfolder, etc.
3. Select your model format, you can convert back and forth between .safetensors and .ckpt
4. Select your precision. FP32 is full precision, FP16/B16 is lower precision which introduces some variations in generations but has the benefit of lower filesize (2GB when pruned). Note that some cards can't handle FP16, also the UI converts all models on VRAM as FP16 when loading them, unless you add --no-half to the launch args in the .bat.

- tl;dr I recommend FP32 as it just werks, FP16 if you don't mind variations and want lower filesizes.

5. Select your model type, I recommend leaving it on All. If you know what non-EMA and EMA-only mean, you know.
6. Press Run and wait for the model to be saved

Models are saved in the base root of the models folder: stable-diffusion-webui\models\Stable-diffusion.

Like all extensions, especially an early one like this one at the time of writing this, some options might change in the future, but the process should remain mostly the same.

Can you prune .safetensors?

Currently .safetensors can't be pruned, if you want to prune them and reduce their filesize (from 8GB to 4GB if using FP32, 2GB if using FP16), you'll have to do it before converting them to .safetensors. If your models are already in .safetensors, you'll have to convert them to .ckpt first to prune them, then back to .safetensors. I recommend this pruning script:

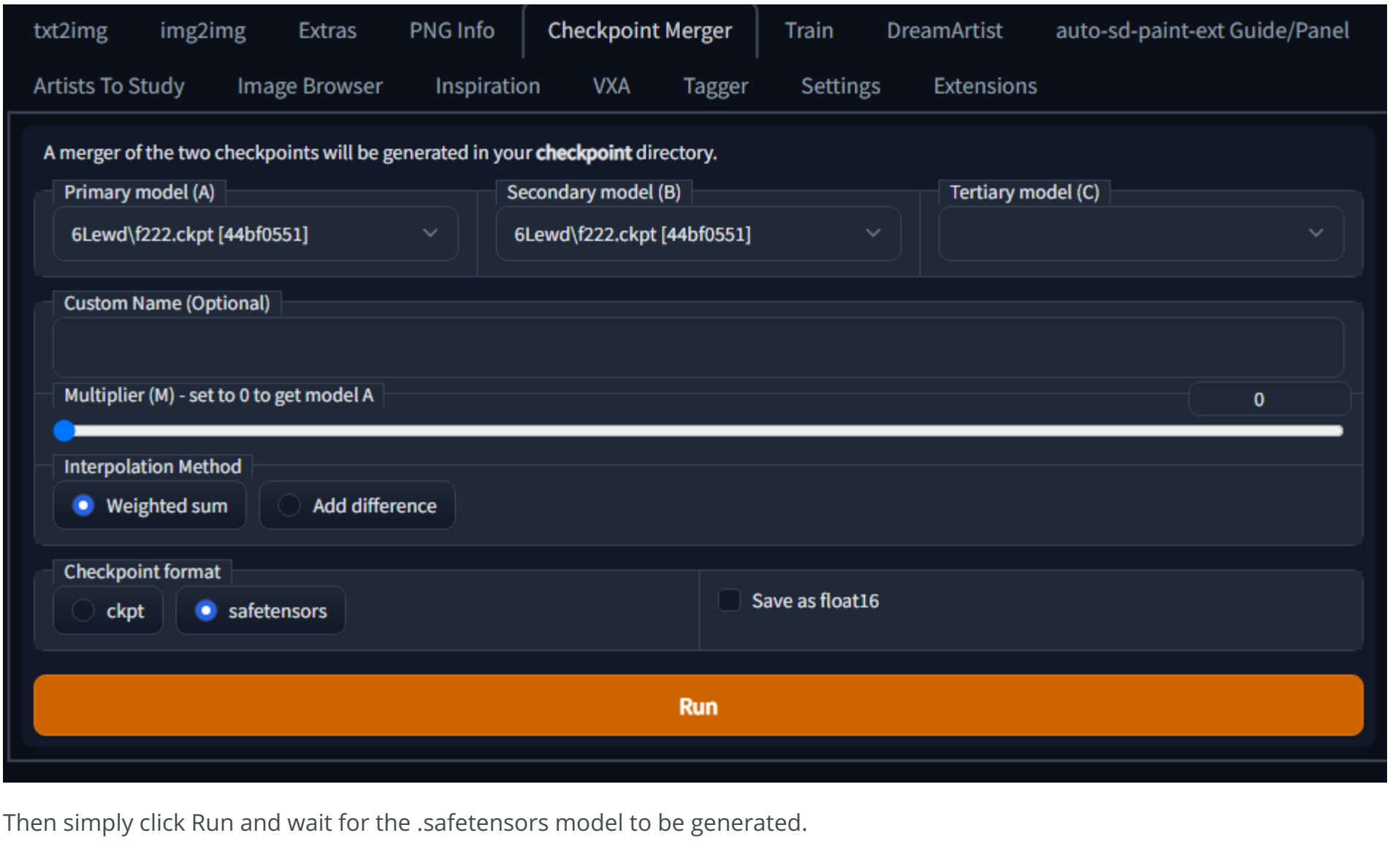
https://github.com/lopho/stable-diffusion-prune

What's pruning? It removes weights that are 0 or almost 0 and have no effect on generations. If the peepoo weight = 0.000001 it basically does nothing and is safe to be removed, same for any weight not being used.

How to convert your .ckpt model to .safetensors using the Checkpoint Merger

1. Go to the "Checkpoint Merger" tab
2. Put the .ckpt model you want to convert in slot A
3. Put the same .ckpt model in slot B (technically it doesn't matter but just in case)
4. Put in a custom name (also doesn't matter, but note that if you leave it blank, the name will contain the name of both models used plus the difference, same as when merging any model, it's just a name anyway)
5. Put the "Multiplier" slider exactly at 0
6. Keep "Weighted Sum" selected, otherwise it'll error out because there's no C model selected
7. Finally in "Checkpoint format", select "safetensors"

Your Checkpoint Merger window should look like this



Then simply click Run and wait for the .safetensors model to be generated.

That's it, you can load the converted model and test to see if everything went right, then start using it normally as you would any other model. Note that you should still have the original .ckpt model, you can keep it as a backup or delete it later once you've confirmed everything works perfectly, up to you.

Optionally, you can also convert to float 16 (FP16) if you really want to. Remember that FP16 can somewhat change outputs and some cards can only use FP32.

Can I merge .safetensors models with other .safetensors and .ckpt ?

Yes, you can merge both .safetensors with each other and even a .safetensors with a .ckpt, just be sure to save as .safetensors.

Why convert .ckpt to .safetensors?

No more fear of "pickles", AKA malicious Python code inserted into the models and no more need to scan for pickles. All current models contain pickles since they're a Python standard, but the word "pickle" became tied specifically to malicious code.

With .safetensors only the weights and specific data needed for generations are included. No additional unrelated and potentially malicious Python code can be included and run when loading .safetensors, like it can be done with .ckpt models.

Converting should be considered when distributing new models and merges going forward, as it avoids the minor paranoia when downloading a new model.

Is there a visual difference between .ckpt and .safetensors?

Nope, they output 100% the same images. Any difference in output is only caused by performance tweaks like --xformers, or also converting them to FP16 during the conversion process. Here's an example, one was generated with the .ckpt model, the other with the .safetensors model, corporate needs you to find the difference between these two pictures.

Do I need to convert all my older models?

Not really, only if you're (re)distributing them. If you're pointing someone to a download of Anything v3 for example, it's probably better to point them to the .safetensors version than the .ckpt version. If you're redistributing your own mix or new Dreambooth model, it's also better to share it as .safetensors.

Is there another advantage to .safetensors?

A bit faster load times, more noticeable in recent builds.

Are .safetensors themselves really safe?

Much safer than .ckpt at least. Do keep in mind that almost all file formats in history have been exploited in one way or another. The current advantage of .safetensors is that malicious arbitrary Python code can no longer be inserted directly and easily into the models, so another type of more advanced exploit would have to be found.

Can I mass convert?

There's a script in the comments for the .safetensors pull request: https://github.com/AUTOMATIC1111/stable-diffusion-webui/pull/4930

Script:
https://gist.github.com/xrpgame/8f756f99b00b02697edcd5eec5202c59

Keep strongly in mind that converting outside of the UI means you don't get the included pickle protection the UI provides, so you need to have scanned everything you want to convert with an external pickle checker beforehand.

Can I train on .safetensors?

Yes, you can train embeddings and hypernetworks normally. Support for .safetensors in the Dreambooth extension is still coming soon ™, it might be even there by the time you read this, hopefully.

My .safetensors load slower than .ckpt

Recently, there was a commit to address slow load times. If you still experience them, add `set SAFETENSORS_FAST_GPU=1` in your `webui-user.bat`, below set `COMMANDLINE_ARGS=` is fine.

If your .safetensors load times are still significantly slower than .ckpt load times, also add `--lowram` to `COMMANDLINE_ARGS=`, this will force all models to load via GPU which will also make them load faster, it shouldn't (?) impact anything else.

tl;dr?

Fuck pickles. Use .safetensors for distributing models, they're safer, make them common use.