Larger resolutions with Stable Diffusion

Problem

Stable Diffusion 1.4 was trained on 512×512 images. That means generating pictures in any other dimensions is going to mess your result up. When straying too far away from 512×512 and 1:1 aspect ratio, you'll get twin heads on characters, long necks, broken composition, tiled repetition and plenty of unwanted results in general. Here's a comparison.

512×512, *John Berkey Sci-Fi*	1024×1024, *John Berkey Sci-Fi* (resized back to 512)

512×512, *photo of a man in the park*	320×896, *photo of a man in the park* (resized back to 512)

Any picture with a clearly defined subject is going to end up like this. Some pictures like landscapes, backgrounds and other similar scenes will actually benefit from repetition, to an extent.

However even those will get garbled once you take it too far.

So, how do you make pictures with larger resolutions?

SD Upscale / GoBIG

If you are using this Web UI, you have a feature called SD upscale (on the img2img tab). It's probably available in other wrappers for Stable Diffusion as well, but I will focus on this one. It will upscale your picture 2x (512×512 will become 1024×1024), using SD itself to invent more details. It can be repeated to make images of larger resolutions. It doesn't take up more memory, just requires proportionally more time. It can yield arbitrarily detailed pictures from mere 512×512, and these would be not fake but "real" details. It can even fix some faces and hands as they tend to be drawn better at larger sizes.

The algorithm works like this:

1. Upscale your image 2x by normal means
2. Divide the 2x image into a bunch of tiles, with some overlap
3. Run img2img on every tile, with respect to your prompt and settings.
4. Combine the tiles to even out the seam.

The settings page for it looks like this:

lower layers of dense jungle with huge trees, cold twilight volumetric mist, earthporn, wide panorama by Raphael Lacoste

Negative prompt

Generate

○ Redraw whole image　○ Inpaint a part of image　○ Loopback　● SD upscale

Image for img2img

● Just resize　○ Crop and resize　○ Resize and fill

Sampling Steps　　30

Sampling method
○ Euler a　● Euler　○ LMS　○ Heun　○ DPM2　○ DPM2 a　○ DDIM

□ Restore faces　　□ Tiling

Tile overlap　256

Upscaler
○ None　○ Lanczos　● 4x_foolhardy_Remacri　○ 4x_foolhardy_Remacri_ExtraSmoother
○ lollypop　○ Real-ESRGAN 4x plus　○ Real-ESRGAN 4x plus anime 6B
○ Real-ESRGAN 2x plus

Batch size　1

CFG Scale　15

Denoising strength　0.45

Height　576

Width　1024

Seed
-1　　□ Extra

Script
None

Output

Save　　Send to img2img　　Send to inpaint　　Send to extras

Interrupt

lower layers of dense jungle with huge trees, cold twilight volumetric mist, earthporn, wide panorama by Raphael Lacoste
Steps: 30, Sampler: Euler, CFG scale: 15, Seed: 1214640615, Size: 1024x576, Denoising strength: 0.45, Denoising strength change factor: 1
Time taken: 496.37s

Your upscaler for the step 1 ("prescaler")

Width and height of ONE TILE this is not the final resolution! final result is always 2x input

Optimal settings

Tile size is best kept the same as original because at different dimensions img2img will generate a completely diffrent picture and your result is going to be different from the original. However, if you don't care about it, you can make tiles larger or smaller to fit the content better. Read the Limitations section below and think how tiles will be layed out in respect to the underlying content.

Keep in mind that if you give the tiles too little overlap, the result might considerably differ in different tiles. If you give too much overlap, you'll waste performance and may get double seams in extreme cases.

Prompt can either be the same, describing the entire image (but also see the Limitation section), or just your styling vectors or something average if your content is too diverse across the tiles, or something entirely different if you want to get creative with adding details.

If you don't want the result deviating too much from the original, keep seed the same.

The main setting is Denoising strength, it works the same as in img2img, as it just runs img2img on each tile. The higher denoising is, the closer the result to the prompt and CFG. The lower denoising, the closer the result to the input picture. So:

- more denoising -> more details induced by prompt and settings, but prone to unexpected hallucinations, visible seams, difference between tiles
- less denoising -> less details, but safer

Usually, denoising > 0.45 gives undesired effects. (depends on the picture, though)

CFG scale works exactly like it does in img2img, again because SD upscale is just tiled img2img.

Sampler is irrelevant but keep in mind that ancestral samplers (Euler_a and DPM2_a) mutate the picture every step, so if you change the

number of steps or switch from ancestral to non-ancestral, they will produce different pictures. Avoid ancestral samplers in your pipeline (including txt2img) if you don't fully understand the implications.

If you want the picture to deviate as little as possible from the original (just add details), keep all settings except denoising the same, including the tile size (same as picture size), prompt, seed etc. If you still want to set the different tile size, try playing with the seed resize feature but I was unable to make it work reliably:

- tick Extra
- set seed the same as original
- set little W and H sliders to the size of the original

Prescaler

The first step in the algorithm (prescaling) is crucial. Here's a crop of two pictures upscaled with SD, with two different pre-scaling algorithms used in the process.

Lanczos	ESRGAN Remacri

Why does this happen? Lanczos is purely algorithmic, ESRGAN Remacri is a neural upscaler which is tuned for crispness and detail preservation. While neither of them is even remotely close to SD upscale, remacri keeps more detail for SD upscale to latch on when hallucinating new details.

Two custom finetuned models for ESRGAN were found to work particularly good with SD upscale: remacri (works better for backgrounds, also tends to amplify brush strokes somewhat), and lollypop (works better for more or less realistic people). You can experiment with other specialized ESRGAN models listed in Upscale Wiki.

Note: Real-ESRGAN is **not** ESRGAN. The naming is confusing, but Real-ESRGAN is a newer, different model which doesn't seem to have finetuned variants. Don't use it, it's better in theory but is shit for this particular purpose.

For the AUTOMATIC1111's wrapper we are talking about, drop your ESRGAN models into the ESRGAN folder, they will be available in SD upscale then.

Limitations

SD upscale has a considerable limitation: the prompt is the same for all tiles, and you can't manually lay tiles out. The more tiles you need to cover the image, the worse the issue is. Think small tiles on larger resolutions.

See the problem? Each tile covers very different content yet they are described by a single prompt. This can lead to SD upscale suddenly dreaming in a face in the grass, or invent another detail unrelated to the particular part of the picture.
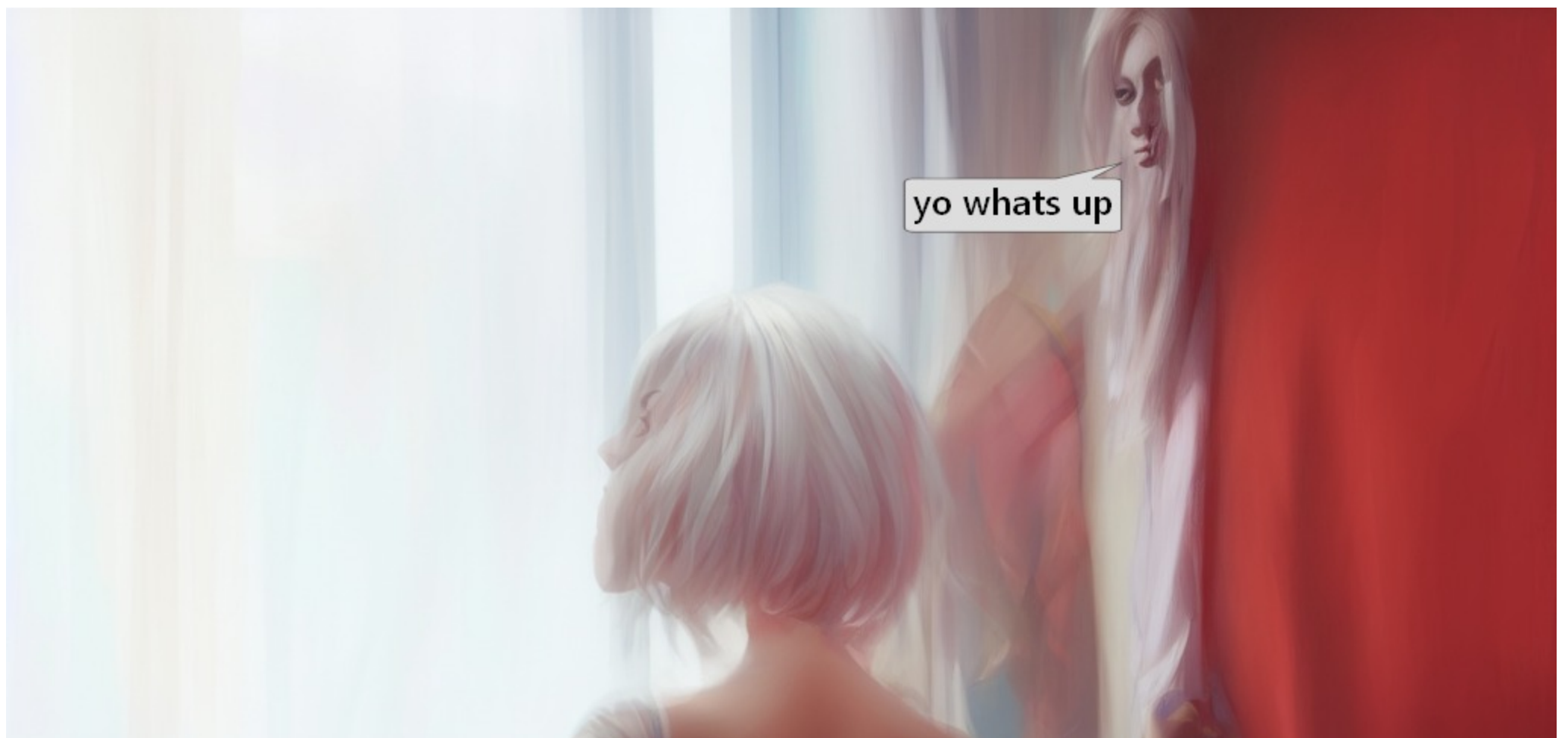
There are several workarounds for this:

- Use larger tiles, or just run the entire prescaled picture through img2img manually, as one piece, if you can fit it in your VRAM. Since img2img is guided by the underlying pre-scaled picture, larger tiles won't give repetition. You will inevitably deviate from your original picture as your tile size will be different from the original size, though. Another person wrote a separate guide for this: https://rentry.org/b7vcb
- Don't set denoise too high on such images. More denoise = more chances for unexpected hallucinations.
- Only use styling vectors in your prompt, no descriptions of the content. Or even no prompt at all. The downside is that details won't be that relevant or good, especially with subsequent upscaling.

- Use manual compositing in Krita or Photoshop. You can prescale anything, and then manually detail it with img2img with any layout, prompt and settings you want.