Stable Diffusion

Stable Diffusion is an open-source diffusion model for generating images from textual descriptions. Note: as of writing there is rapid development both on the software and user side. Take everything you read here with a grain of salt.

Contents

How to Use

Usage instructions for both online and local use.

Getting started

- beta.dreamstudio.ai (https://beta.dreamstudio.ai): official web service, starts you off with 200 free generations (standard resolution).
- Official Github page (https://github.com/CompVis/stable-diffusion): CLI with the bare minimum of features.
- webui developed by AUTOMATIC1111, a.k.a Voldemort (https://github.com/AUTOMATIC1111/stable-diffusion-webui): gradio-based GUI that you access from your browser, use this if you can. It has a lot of cutting-edge features like negative prompts or attention via parentheses. Installation instructions that attempt to be more noob friendly than the Github page can be found here (https://rentry.org/voldy). A non-graphical frontend (https://github.com/JohannesGaessler/stable-diffusion-ipython-shell) is in early development. Despite the name the sd-webui a.k.a hlky fork (https://github.com/sd-webui/stable-diffusion-webui) was actually forked from AUTOMATIC1111.
- Waifu Diffusion finetune (https://github.com/harubaru/waifu-diffusion): finetune of Stable Diffusion on Danbooru tags, significantly improved visuals for anime-style images. A recipe for prompt design can be found below.
- Stable Diffusion Akashic Records (https://github.com/Maks-s/sd-akashic) large collection of links to useful SD-related resources. Be aware that the collection is barely curated so take anything linked there with a grain of salt.

Generation Parameters

The names for the input parameters when generating images differ slightly from interface to interface but they generally do the same thing. Below is a brief explanation of what the parameters do:

- Prompt: textual description of what you want to generate.
- Negative prompt: textual description of things that you don't want in the image.
- Sampling Steps: diffusion models work by making small steps from random Gaussian noise towards an image that fits the prompt. This is how many such steps should be done. More steps means smaller, more precise steps from noise to image. Increasing this directly increases the time needed to generate images. Diminishing returns, depends on sampler.
- Sampling method: which sampler to use. *Euler a* (short for ancestral) produces great variety with a low number of steps, but it's very difficult to make small adjustments. The non-ancestral samplers all produce mostly the same images as the number of steps increases, use LMS if you're unsure.
- ddim_eta: amount of randomness when using DDIM.
- Batch count/n_iter: how often to generate a set of images.
- Batch size: how many images to generate at the same time.**Increasing this value can improve performance but you also need more VRAM.**Total number of images is this multiplied with batch count.
- CFG Scale (classifier-free guidance scale): how strongly the images match your prompt. Increasing this value will result in images that resemble your prompt more closely (according to the model)

but it also degrades image quality after a certain point. Can be somewhat counteracted with more sampling steps.
- Width: width of individual images in pixel.**To increase this value you need more VRAM.**Image coherence on large scales (model was trained on 512x512) becomes worse as the resolution increases. Very small values (e.g. 256 pixels) also degrade image quality.
- Height: same as Width but for individual image height.
- Seed: starting point for RNG. Keep this the same to generate the same (or almost the same) images multiple times. There are no seeds that are inherently better than others but if you vary your input parameters only slightly seeds that produced good results previously will likely still produce good results.

Example Prompts

Some example prompts that highlight different ways of using Stable Diffusion (intended to be educational, not to show the absolute best prompts). Straightforward prompt for a portrait of a conventionally attractive woman:

> thick lips, black hair, fantasy background, cinematic lighting, highly detailed, sharp focus, digital painting, art by junji ito and WLOP, professional photoshoot, instagram

Prompt 1 (portrait), cherrypicked result (best of 9).

Cryptic prompt that inadvertently produces Albert Einstein cryptocurrency:

> HODL\0HODL\0E=MC2\01299d18a-1b5b-4c20-8580-8534af5e4995, 4K, 8K, award-winning

"4K, 8K, award-winning" are just generic buzzwords to make the images look better. Everything necessary to understand what's happening can be found on this wiki page.

Prompt to generate non-hideous shortstacks:

> photorealistic render of a plump halfling maiden cooking in the kitchen, pixar, disney, elf ears, pointed ears, big eyes, large cleavage, busty, cute, adorable, artists portrait, fantasy, highly detailed, digital painting, concept art, sharp focus, depth of field blur, illustration

Shortstacks are frequently associated with goblins but the training data contains a lot of examples for "goblin girl" that are just plain ugly. And because diffusion models are learning to create images that resemble the data, any prompt that uses "goblin girl" will produce similarly ugly-looking goblin girls. The trick of this prompt is to avoid the association with goblins and to instead go with "halfling" which has much more aesthetic-looking examples in the dataset. To make the resulting shortstacks look more like goblins the prompt can be modified to for example specify green skin.

Prompt 2 (Albert Einstein cryptocurrency), cherrypicked result (best of 9).

To generate anime-style visuals specifically the use of the Waifu Diffusion finetune is recommended. The TLDR is that it allows you to generate images from Danbooru tags, a generic anime girl can for example be generated with the following prompt:

> original 1girl bangs blue_eyes blush long_hair looking_at_viewer open_mouth smile solo

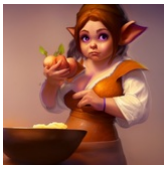
Prompt Design / General Tips

Guidelines for creating better prompts and getting better results.

What To Expect

The images you generate will not be perfect, even if you do everything right. The important thing to remember about the images that people post on the internet is that you don't see the process that went into the images. Very frequently people simply generate hundreds of images for a given prompt and only post the very best ones; why wouldn't you when you can just make your PC generate a bunch of images while you sleep?

The images people post on the internet may have also been edited to fix imperfections. This can be done with either the Stable Diffusion img2img tool or with conventional image editing software like GIMP or Photoshop. In either case some manual effort and skill is required to get good results.



Prompt 3 (shortstack), cherrypicked result (best of 9). Hands are deformed, this is a very common problem with images generated by convolutional neural networks.

What To Write

Write text that would be likely to accompany the image you want. Typically this means that the text should simply describe the image. But this is only half of the process because a description is determined not just by the image but also the person writing the description.

Imagine for a moment that you were Chinese and had to describe the image of a person. Your word of choice would likely no longer be "person" because your native language would be Chinese and that is not how you would describe a person in Chinese. You wouldn't even use Latin characters to describe the image because the Chinese writing system is completely different. At the same time, the images of people that you would be likely to see would be categorically different; if you were Chinese you would primarily see images of other Chinese people. In this way the language, the way something is said, is connected to the content of images. Two terms that theoretically describe the same thing can be associated with very different images and any model trained on these images will implicitly learn these associations. This is very typical of natural language where there are many synonymous terms with very different nuances; just consider that "feces" and "shit" are very different terms even though they technically describe the same thing.

Prompt 4 (anime girl), cherrypicked result (best of 9).

TLDR: when choosing your prompt, think not just about what's in the image but also who would say something like this.

Prompt Length

Be descriptive. The model does better if you give it longer, more detailed descriptions of what you want. Use redundant descriptions for parts of the prompt that you care about.

Note however, that there is a hard limit regarding the length of prompts. Everything after a certain point - 75 or 76 CLIP tokens depending on how you count - is simply cut off. As a consequence it is preferable to use keywords that describe what you want concisely and to avoid keywords that are unrelated to the image you want. Words that use unicode characters (for example Japanese characters) require more tokens than words that use ASCII characters.

Punctuation

Use it. Separating keywords by commas, periods, or even null characters ("\0") improves image quality. It's not yet clear which type of punctuation or which combination works best - when in doubt just do it in a way that makes the prompt more readable to you.

Attention / Emphasis

There is a lot of confusion around attention, meaning ways to increase or decrease the weight of specific parts of a prompt. Some people assert that putting a keyword in round brackets increases its effect while putting a keyword in square brackets decreases its effect; Using more brackets supposedly results in a stronger change. However, others can frequently **not** reproduce this effect in their own prompts.

As it turns out the reason for the discrepancy is that*different scripts process brackets differently.* This fork of webui (https://github.com/AUTOMATIC1111/stable-diffusion-webui) for example explicitly processes brackets while this fork (https://github.com/hlky/stable-diffusion-webui) would only get the effect of brackets that the model implicitly learned from the training data (a different syntax is used instead). For this reason: **make sure to check whether the syntax of a prompt that you copy from someone else matches the syntax of the script that you use!**In a quantitative analysis (https://github.com/JohannesGaessler/stable-diffusion-tools/tree/master/emphasis) square brackets did not have a consistent effect unless explicitly processed.

The repetition of a certain keyword seems to increase its effect regardless of the specific script that is being used.

Specificity

The model has essentially learned the distribution of images conditional on a certain prompt. For the training of neural networks the quality of features is important: the stronger the connection between the inputs and the outputs is, the easier it is for a neural network to learn the connection. In other words, if a keyword has a very specific meaning it is much easier to learn how it connects to images than if a keyword has a very broad meaning. In this way, even keywords that are used very rarely like "Zettai Ryouiki" can produce very good results because it's only ever used in very specific circumstances. On the other hand, "anime" does not produce very good results even though it's a relatively common word, presumably because it is used in many different circumstances even if no literal anime is present.

Choosing specific keywords is especially important if you want to control the content of your images. Also: the less abstract your wording is the better. If at all possible, avoid wording that leaves room for interpretation or that requires an "understanding" of something that is not part of the image. Even concepts like "big" or "small" are problematic because they are indistinguishable from objects being close or far from the camera. Ideally use wording that has a high likelihood to appear verbatim on a caption of the image you want.

Movement and Poses

If possible, choose prompts that are associated with only a small number of poses. A pose in this context means a physical configuration of something: the position and rotation of the image subject relative to the camera, the angles of the joints of humans/robots, the way a block of jello is being compressed, etc. The less variance there is in the thing that you're trying to specify the easier it is for the model to learn. Because movement by its very definition involves a dramatic change in the pose of the subject, prompts that are associated with movement frequently result in body horror like duplicate limbs. Also because human limbs and especially human hands and feet have a lot of joints they can assume many different, complex poses. This makes their visualization particularly difficult to learn, for humans and neural networks alike.

TLDR: good image of human standing/sitting is easy, good image of human jumping/running is hard.

Height and Width / Cut-Off Heads

Images of humans generated with Stable Diffusion frequently suffer from the subject's head being out-of-frame. The reason for this is that the training data was cropped to square images; if the image height was larger than the image width this oftentimes cut off a person's head and feet. The extremely simple prompt "runway model" is a good example of this. The images associated with this prompt are almost all in portrait mode and shot by professional photographers. As professionals the photographers know how to properly frame their subjects in such a way that the subjects are in frame but without wasting too much space at the top and the bottom - and this is precisely why cutting off the top and the bottom of such photographs consistently removes the head and feet.

Non-cherrypicked samples for the prompt "runway model" at the default resolution of 512x512. Heads are mostly out of frame.

A solution to cut-off heads is to change the aspect ratio of the generated images: if the image height is increased the images are extended beyond their typical borders and heads and feet are more likely to be generated. Note the usage of the term "likely": because it's essentially random whether the image gets extended at the bottom or the top you may end up with images where the head is still cut off but you get a lot of space below the subject's feet.

Some specifics for the prompt "runway model" (keep the limited sample size in mind):

- With a width of 448 pixels the subjects' heads were mostly in frame at a height of 704 pixels. Image coherence started to degrade at a height of 896 pixels.
- With a width of 512 pixels the subjects' heads were mostly in frame at a height of 768 pixels. Image coherence started to degrade at a height of 832 pixels.
- Increasing both the image width and height at the same time greatly reduce image coherence compared to increasing just one of the two.

Miscellaneous

- Unicode characters (e.g. Japanese characters) work.
- Capitalization does not matter.
- At least some Unicode characters that are alternative versions of Latin characters get mapped to regular Latin characters. Full-width Latin characters as they're used in Japanese (e.g. ＡＢＣ) are confirmed to be converted. French accents (e.g. é and è) and German umlauts (e.g. ä and ö) are **not** mapped to their regular counterparts.
- Extra spaces at the beginning and end of your prompt are simply discarded. Additional spaces between words are also discarded.
- Underscores ("_") are **not** converted to spaces.

Non-cherrypicked samples for the prompt "runway model" at a modified resolution of 448x832 (7:13 aspect ratio). Heads are mostly in frame.

Finetunes

Specific usage instructions for Stable Diffusion finetunes (SD trained on special training data, e.g. anime).

Waifu Diffusion

The Waifu Diffusion finetune was trained on Danbooru (https://danbooru.donmai.us/) images. The optimal way to build prompts is not yet clear but a good starting point is to use the same method (https://github.com/harubaru/waifu-diffusion/blob/main/danbooru_data/scrape.py) that was used for the finetune training data:

1. Select one or more *copyright tags*, meaning tags that specify which anime/manga/game the image subjects are from. Use "original" for original content.
2. Select *character tags* if applicable, meaning tags that specify the characters that SD should generate.
3. Select *general tags*, meaning regular tags that describe the content and style of the image. **Do not use meta tags such as "highres" as these were not used in training.**
4. Select *artist tags*, meaning tags that specify which artists' style to emulate.
5. Sort the tags within each of the aforementioned categories alphabetically (numbers before letters, same order as on Danbooru).
6. Replace spaces within tags with underscores.
7. Build the prompt by appending the tags from each category, separated by spaces (no commas or periods). Put the copyright tags first, then the character tags, then the general tags, and then the artist tags.

Example prompt based on this (https://danbooru.donmai.us/posts/2294702) picture:

> kono_subarashii_sekai_ni_shukufuku_wo! megumin 1girl ^_^ backlighting blush brown_hair cape closed_eyes collar facing_viewer fingerless_gloves flat_chest gloves hat lens_flare short_hair short_hair_with_long_locks smile solo staff sunset witch_hat dacchi

Example results are shown to the right. Overall the samples are not very accurate both in terms of the character and details like the eyes or staff; getting specific results with diffusion models is always difficult though. Emphasis might help but as of right now some manual editing or img2img is still required for optimal results.

Non-cherrypicked samples for the example prompt to demonstrate the Danbooru recipe.

Keywords

The most reliable way to find good keywords is to look at the keywords that are used to generate images that are similar to what you want. Alternatively there are multiple websites that let you explore various art styles and other modifiers. Adding the names of artists that you like to your prompt seems to be the most reliable way of getting good-looking results (be aware that the model may not know the artist you are talking about). Below are some (unconventional) known good keywords (as determined by using keywords as prompts without other keywords or in very short and simple prompts). The underlying assumption is that the keywords will also be good as part of large prompts; if they are not, please provide feedback. When the list tells you to avoid keywords the reason is that they simply produce bad outputs. Keywords that produce unexpected unsafe outputs have an explicit warning. An archive with the samples used to judge these keywords can be found here (https://mega.nz/folder/oRM1xAAJ#MeZYuu-lkKNMC3fgrvhsmw).

Generic Keywords

Keywords are considered generic if they are applicable to a wide range of contexts. This is in contrast to specific keywords that have clear themes and thus are only useful for certain kinds of images, e.g. mythology, fantasy, or anime.

Clothing

Clothing options for your waifu. Note that by specifying clothing you are implicitly specifying physical characteristics. For example, "bikini" is generally produces breasts that appear large on camera while "camisole" produces breasts that appear comparatively small on camera.

- "bikini armor": not understood, explicitly describe the armor instead.
- "blazer": women in jackets that have a somewhat professional look.
- "bodycon": women in tight-fitting dresses. Biased towards hourglass figures.
- "bodystocking": women in stockings that cover the whole body. **Produces unsafe outputs.**
- "bodysuit": essentially a one-piece swimsuit made of regular fabric. "lace bodysuit" produces the lingerie version.
- "bralette": lightweight bra without an underwire.
- "camisole": women in loose-fitting, sleeveless tops. Makes the breasts appear smaller.
- "cheongsam", "qipao", "mandarin gown": women in traditional Chinese dresses.
- "corset": piece of clothing that squeezes the torso into more of an hourglass figure. **Can produce unsafe outputs.**
- "diaphanous": clothing made from translucent fabric. **Can produce unsafe outputs.**
- Dress length can be controlled by specifying a certain dress type. From longest to shortest: "gown" (hemline touching the ground), "long dress"/"maxi dress" (hemline at the ankles), "midi dress" (hemline between ankles and knees), "knee-length dress", and "mini dress" (hemline above the knees). Avoid "microdress".
- "floral print": fabric with flower pattern.
- "Gothic Lolita": women in frilly black dresses.
- "jumpsuit": women in loose-fitting one-piece garments that cover the torso and legs. Makes the breasts appear smaller.
- "loincloth": muscular men wearing loincloths.
- "maid": does not give you the stereotypical maid aesthetic. Use "French maid" instead. Avoid "メイド".
- "negligee": women in loose-fitting, see-through garments. **Produces unsafe outputs.** Avoid "babydoll", "baby doll dress", and "négligée".
- "plugsuit": pilot suits from Neon Genesis Evangelion. Questionable accuracy without further context.
- "reenactor": interpreted similarly to cosplayer, also works in ridiculous combinations like "XCOM reenactor". Compared to "cosplayer" there are more natural environments, less sexualization, and more practical outfits.
- "sarong": skirt/dress obtained by wrapping fabric around the waist, frequently worn in e.g. Asia and Africa. Images are biased towards women and colorful fabric.
- Skirt length can be controlled by specifying a certain skirt type. From longest to shortest: "long skirt"/"maxi skirt" (hemline at the ankles), "midi skirt" (hemline between ankles and knees), and "miniskirt" (hemline above knees). Avoid "ankle-length skirt", "knee-length skirt", and "microskirt".
- "slip dress": women in loose-fitting, sleeveless dresses. Makes the breasts appear smaller.
- "slit dress": women in dresses that have slits on the side which expose the legs. Variants like "front slit dress" where the slit is in the front are not understood.
- "sports bra": women wearing sports bras. Makes the breasts appear smaller.
- "sweater": warm upper-body wear with long sleeves. For the design with lines try "ribbed sweater".
- "Sweet Lolita": women in frilly pink dresses.
- "thong": women in underwear that does not cover the cheeks. Avoid "C-string", "T-back", and "T-front".
- "zentai", "morphsuit": people whose entire body is covered in tight-fitting, monochromatic suits.
- "Zettai Ryouiki": short skirt in combination with stockings or socks, visible thighs. Avoid "絶対領域" (kanji spelling).

Perspective

Ways to get certain camera angles/viewpoints. Note that these keywords also tend to bias the content of images. The specific effect of these keywords often depends on the subject. The results listed here are based on the output from the keyword without context as well as combinations with "French maid", "bronze statue", and "World Heritage Site":

- "aerial shot", "bird's eye shot": pictures of the ground shot from high in the sky. "French maid" and "bronze statue" were lying on the ground and they were huge compared to their environment.
- "close-up": zooms in on the subject, but not necessarily in the way as the term is used in filmmaking.
- "freeze frame shot": mostly pictures of people as they are jumping. Avoid "freeze frame".
- "from behind": shows the backside of the subject. Works very well with humanoid subjects, with "World Heritage Site" it was unclear whether the actual backside was shown. **Can produce unsafe outputs without context.**
- "full body", "full body portrait", "full-body portrait": shows the entire body of a humanoid subject. Remember that this results in cut-off heads. "full body portrait" forced humanoid subjects in combination with "World Heritage Site".
- "high angle": pictures taken from above at relatively close range, worked ~50% of the time with "French maid". On its own or with "World Heritage Site" similar to aerial shot. Did not work in combination with "bronze statue" (images were low-angle shots instead).
- "isometric view": depicts 3D on a 2D screen by giving all axes equal length.
- "looking at camera": humanoid subject looking at the camera. Worked well with "French maid", somewhat with "bronze statue" and "World Heritage Site".
- "low angle": shows the subject from below. Produces pictures of buildings and trees without context, works sometimes with "French maid" and "bronze statue", worked well with "World Heritage Site".
- "portrait": paintings of peoples' faces or upper bodies. With "bronze statue" it zooms in on the face. With "World Heritage Site" it can add faces to the buildings.

Style

Ways to get certain image styles:

- "bokeh", "depth of field": blurs the foreground and background (photography).
- "bronze statue": shiny statues of people.
- "daguerreotype", "ambrotype", "tintype": early photographic processes, not sure if the model actually learned the differences between these.
- "Gorillaz": somewhat reproduces the associated art style.
- "hourglass figure": female body type. "rectangle body shape" and "inverted triangle body shape" work somewhat. Avoid combinations with "spoon" and "rectangular" as well as combinations not listed here.
- "ink": high-contrast black-and-white illustrations.
- "infrared": images taken with an infrared lens. To get images somewhat resembling those taken with an infrared camera use "thermal vision". Avoid "thermal sight" and "LWIR".
- "iridescent", "prismatic": trippy colors, works well with illustrations but can be difficult to apply to photorealistic images. Specify what part of the image the colors should be applied to, e.g. "iridescent skin".
- "Lovecraftian", "Necronomicon": occult imagery. "Lovecraftian" produces more tentacles than "Necronomicon".
- "science": garbage. Use "cyclotron" or "synchrotron" to make things look like science instead.
- "Snapchat selfie": selfie with just one person in frame. "selfie", "selfie, Snapchat" and "Snapchat, selfie" produce images with multiple people. Avoid "Snapchat".
- "studio photography": makes photographs look better, simple backgrounds.
- "thicc": sexualized pictures of women with comparatively high body fat compared to what is considered conventionally attractive. Looks slightly better than "thick".
- "patronus": animals with a blue glow. Could combine with "dog" and "pikachu", not with "Richard Stallman", "geoduck", "cockroach", or "[car brand]".
- "professional photo shoot": makes photographs look better, blurs the background.
- "wide angle lens", "wide-angle lens": pictures taken with a high angle of view, nearby objects appear larger, straight lines become rounded. Avoid "fisheye lens".
- "wikiHow": can replicate the style of the illustrations.
- "x-ray": images of x-ray medical imaging.

Specific Keywords

Keywords that only apply to certain contexts/images. This is in contrast to generic keywords that are applicable to a wide range of contexts/images.

Mythology / Fantasy

SD has learned terms from mythology. Unfortunately not all of them yield good results. Terms that are too obscure like "Argonaut" are simply not recognized while terms that are too popular like "Thor" or "Odyssey" produce unrelated garbage. The following terms yielded good results:

- "666": satanic imagery.
- "Dungeons & Dragons", "Dungeons and Dragons": illustrations related to the RPG. "D&D" and "DnD" also worked but looked slightly worse.
- "Fenrir": wolf from Norse mythology, produces illustrations of fierce-looking wolves.
- "Hercules": produces the Disney character. Use the Greek equivalent "Heracles" instead.
- "Iliad": one of the two ancient Greek epic poems written by Homer, results in ancient Greek visuals. Avoid "Odyssey".
- "Medusa": one of the gorgons (women with snake hair) from Greek mythology, produces generic women with snake hair. Avoid "gorgon".
- "mind flayer", "illithid": doesn't produce the iconic Dungeons & Dragons monster but can be used for generic tentacle humanoids.

Scientific Names

SD has learned several scientific terms from biology. The names of species can produce images simply showing that species. Terms that contain more than one type of animal (for taxonomical reasons or e.g. because dogs and wolves belong to the same species) can produce hybrid creatures.

- "Canis lupus": dog-wolf hybrid creatures.
- "Felis catus": cats.
- "Homo sapiens": genus that modern humans belong to, disturbing humanoid hybrid creatures. "Homo sapiens sapiens" (anatomically modern humans) produces slightly less disturbing humanoid creatures.
- "invertebrate": hybrids of insects, spiders, etc.
- "Loxodonta africana": African bush elephant, looks better than "elephant" and .

Subreddits

Stable Diffusion has learned which kind of image gets posted to which subreddit. Unfortunately for most subreddits it has learned incomprehensible garbage, typically because the images contain a lot of text. Subreddits that are essentially just image dumps work pretty well though:

- "r/AbandonedPorn", "AbandonedPorn": pictures of abandoned buildings.
- "r/aww", "r/awwducational": cute images of cats and dogs. Avoid "aww".
- "r/battlestations", "battlestations": pictures of desktop PCs.
- "r/creepy": creepy images, mostly drawings of faces. Avoid "creepy".
- "r/EarthPorn", "EarthPorn": landscape photography.
- "r/evilbuildings", "evilbuildings": buildings that look like they're owned by a super villain or evil corporation. "evil buildings" is random skyscrapers.
- "r/eyes": bright blue eyes + conventionally attractive faces.
- "r/fitgirls", "r/Fitness": muscular women. "Fitness" is just pictures of women working out. "Reddit fitness" seems to be interpreted similarly to "r/Fitness".
- "r/gardening": pictures of home gardens. "gardening" is pictures of garden work.
- "r/GirlsWithGlasses": selfies of women wearing glasses.
- "r/interestingasfuck": can give you cool textures but can also fuck up your images.
- "r/InternetIsBeautiful": abstract colorful images.
- "r/MachinePorn", "MachinePorn": industrial machinery.
- "r/OldSchoolCool": vintage photographs, has more varied and interesting subjects compared to "vintage photograph".
- "r/SkyPorn", "SkyPorn": pictures of the sky.
- "r/spaceporn": pictures of space.

All of the 100 largest subreddits were tested. The ones not listed here produced either garbage or unremarkable results.

Note that for some subreddits it has been confirmed that "/<subreddit>" and "<random letter>/<subreddit>" produce nearly identical results. These may be adversarial examples: in the training data there are presumably many images associated with the string "r/<subreddit>" and basically none with other letters. Instead of learning the meaning of "r/<subreddit>" SD may therefore have simply learned a meaning for "/<subreddit>" because with the training data the two terms were virtually interchangeable.

Weebshit

Anime and other Japanese things:

- "ahegao" somewhat produces the meme face but honestly it doesn't look very good. **Can produce unsafe images.**
- "anime": generic, mediocre anime-style images, looks somewhat like the 2000s. Since "anime" is associated with many low-quality/unrelated images a common strategy is to just specify a drawing and use Japanese words in your prompt to associate your prompt with what a Japanese person would be likely to draw (i.e. anime). For style variations try "アニメ" (Japanese way to write anime, looks more modern), "chibi", "ドルフィー" (Japanese doll brand), "Kyoto Animation", "light novel illustration", "shonen", "Studio Ghibli", "visual novel CG", or "Yusuke Murata" (artist of the One-Punch Man manga). Avoid "manga", "tankobon", and "waifu". Order of keywords is simply alphabetical.
- "cosplay": pictures of western people cosplaying. "コスプレ" is pictures of Japanese people cosplaying.

- "gyaru": Japanese women with tanned skin and dyed hair.
- "hentai": bad. "エロアニメ", "エロゲ", "エロ同人", and "エロ漫画" less bad. "エロゲ" and "エロ同人" also produce 3D.
- "ikemen": handsome Japanese men. Avoid "イケ面" (Japanese spelling).
- "manga", "tankobon": generic anime-style images, artifacts from text and paneling, also associated with pictures of physical copies. "漫画" and "マンガ" look better but also have artifacts. "漫画" seems to be more associated with manga for adults while "マンガ" is more associated with manga for children.
- "Nekopara": cat girls from the franchise. Avoid "ネコぱら".
- "nendoroid": brand of plastic figures for characters from anime, manga, and video games. Avoid "ねんどろいど".
- "oneshota": cute anime boys.
- "pantsu": vaguely Japanese-looking women in underwear. "panties" looks better. Avoid "パンツ". "shimapan" means striped panties**but the result is Japanese porn.**
- "to-love-ru": characters from the franchise. Avoid "toraburu" and "To LOVEる".
- "Touhou", "Touhou Project": characters from the franchise. Avoid "東方".
- "waifu": modern Japanese women.
- "アイドル", "aidoru": Japanese idols. "アイドル" is mostly 3D, "aidoru" is mostly 2D.
- "フィギュア": general plastic figurines of anime/mange/video game characters.
- "ガンプラ", "gunpla": Gundam plastic models. Avoid "ganpura".
- "イラストレーション", "イラスト": illustrations in Japanese style (I think, definitely not "anime" style), "イラスト" looks more abstract than "イラストレーション". In limited testing "イラストレーション" looked slightly better than "イラスト".
- "悪魔": frequently translated as "demon", demonic imagery in anime/Japanese style. "akuma" produces the street fighter character.
- "美女", "美人": Japanese women, classical beauty standard.
- "男性": literal meaning is just man/male gender**but the result is Japanese gay porn**.
- "彼女", "kanojo": Japanese women, kanojo also contains 2D and pictures of couples.
- "可愛い", "かわいい": pronounced "kawaii", cute things. On its own "可愛い" produces pictures of birds, "かわいい" general cute things. However, good results were achieved with "可愛い" in long prompts so testing the single keywords may be inaccurate. Avoid "kawaii" and "カワイイ".
- "女": Chinese/Japanese women.
- "巨乳", "爆乳", "おっぱい": Japanese women with large breasts, either topless or wearing a bra.

Miscellaneous

- A random UUID essentially gives you a random type of digital photograph. These are very weak though so you can append a random UUID to your prompt for a slight variation.
- "asperitas clouds": skies with wild, stormy-looking clouds. "asperitas" seems to be associated with the pattern in a more abstract way.
- "bobs and vagene", "Mr. Dr. Durga sir", "please do the needful": do not redeem the prompt
- "cock and ball torture": with Craiyon this produces a meme result**but Stable Diffusion produces the actual thing**
- "cheeki breeki": *bandit_radio.mp3 starts playing in the background*
- Emoji work well:

  produces pictures of dogs. produces images of fast food.
  unsafe outputs. Avoid
- "geoduck": edible clam with a weird shape.
- Geographic locations work well. For example, "Karlsruhe" is a city in Germany with ~300000 inhabitants and SD can reliably reproduce the style of the buildings.
- "hodl": memecoins. "diamond hands" and "paper hands" are taken literally.
- "E=mc2": Albert Einstein.
- "masked": people wearing COVID masks. Can be used to suppress humanoid faces when e.g. generating humanoid robots.
- "miniature wargaming": many small plastic figures.
- "miniature figure": a single small figure.
- "painting by Adolf Hitler": couldn't even get into art school, what a loser.
- "Pepe the Frog": difficult to generate good ones. Related keywords that also work to some degree are "Matt Furie" (guy that created Pepe), "rare Pepe", "Angry Pepe", and "Sad Pepe". Avoid any generic keyword related to real-life frogs as well as "Pepe", "feels good man", "feels bad man", "kek", and "reeee" (did not work with 10 or 20 e characters either).
- "sandals": feet in sandals. Compared to e.g. "feet" the body horror is greatly reduced, presumably because the sandals restrict movement of the toes. can produce similarly good consistently, sometimes on bare feet).
- "sunbeam", "crepuscular rays", "god rays": large, visible rays of light, extra damage against vampires. Avoid "light shafts".
- "thumbs up": gesture in which hands seem to look less bad.
- "World Heritage Site": ancient buildings.

Tips for Finding Keywords

- Querying the LAION dataset here (https://rom1504.github.io/clip-retrieval/) can be used as a quick way of checking which keywords are in the dataset but keep in mind that this is not 100% reliable.
- Stable Diffusion may not know a certain art style but artists of that art style and vice versa.

Useful Links

- GFPGAN (https://github.com/TencentARC/GFPGAN): Tool for fixing faces
- krea.ai (https://www.krea.ai/): Website that lets you explore keywords
- promptoMANIA prompt builder (https://promptomania.com/stable-diffusion-prompt-builder/)
- clip-retrieval (https://github.com/rom1504/clip-retrieval): Project that lets you determine the relationship between images and keywords, works in either direction. Online versionhere (https://rom1504.github.io/clip-retrieval/)
- Archive of samples produced by individual keywords (https://mega.nz/folder/oRM1xAAJ#MeZYuu-lkKNMC3fgrvhsmw)
- Google Arts & Culture (https://artsandculture.google.com): can be used to discover artists (https://artsandculture.google.com/category/artist), art movements (https://artsandculture.google.com/category/art-movement), mediums (https://artsandculture.google.com/category/medium), etc.
- [1] (https://rapidgator.net/file/733331357588e7aacc10485fe3d37a49/Add_text.ps1.html): Powershell script to add text to pre-processsing .txt files (place in process dir, right click, edit)
- [2] (https://rapidgator.net/file/e5b544779538257ac9cd0163a8d96af6/Remove_text.ps1.html): Powershell script to remove text from pre-processing .txt files (place in process dir, right click, edit)

Retrieved from "https://wiki.installgentoo.com/index.php?title=Stable_Diffusion&oldid=52589"

- This page was last edited on 21 October 2022, at 23:14.
- Content is available under Public Domain unless otherwise noted.