Stable Diffusion模型配置

2023年4月20日 19:03

stable diffusion和Midjourney的异同：

1、两者都是AI绘图软件

2、Stable diffusion是一个开源的、偏向于娱乐的软件。它一般部署在本地，也可以部署在云服务器上。但是stable diffusion的出图质量不稳定，需要有controlnet等进行控制，否则可能会有错误，具体看运气；配合lora有很多种玩法，可以炼丹，想画什么就画什么，包括生成涩图。难以生成同一人物的微调图片

3、midjourney专门为商业设计，需要付钱，跑在云端，啥电脑都能用。出图质量佳，基本没有错误；不能生产敏感图。可以使用近乎口语的关键字进行描述，并配合少量的关键字即可出图。可以通过命令生成同一人物不同表情或动作

注意点：

1、之前的模型后缀都是使用的.ckpt，现在为了安全已换成.safetensors，即安全张量，但两者基本上是一致的，都可放置在Stable Diffusion文件夹下使用。.ckpt是序列化的，有可能包含恶意代码，而.safetensors是用numpy保存的，只包含张量数据，更安全

2、webUI实质上是一个便于操作的本地页面，实际上是在cmd里运行的，所以生成图的过程中cmd也会动

3、Stable Diffusion简称SD，是一类模型的总称，而checkpoint和Lora是它的具体实现模型，其中checkpoint是存放着模型权重参数的文件，本质是一个变量名-变量值的词典，是主模型，在大方向上决定了最终生成的效果，例如novelAI的二次元风格模型、chilloutmix-Ni真人模型等；而Lora是一个较小的参数模型，用来调整checkpoint模型的产出，以达到想要的效果，如各种具体角色偏向的（例如神里）以及一类人群偏向的（如japanese likeness）。

主模型往往只适用于某一类作画，比如chilloutmix模型主要画女生正面图，因为它的训练集里基本上就是单人女生图，就算prompt里写了有male或者其它，而且把权重提得很高也没法画出来或者画的很垃圾。

4、模型后面跟着的字符串为哈希值，用来标注模型的唯一性

5、fp16和fp32对生成图形影响不大，可以选择fp16。fp32模型的大小是fp16的接近两倍

6、有些选项需要依赖于未进行下载的包，例如面部修复。勾选后将自动进行下载（CMD里可以看到）。所以不要随便就使用某个功能

7、EMA即指数移动平均法，可提高模型的鲁棒性

8、图像模糊不要随便更改图片分辨率，而是选择Hires.fix选项（高清修复）

9、一定要输入正负面画面质量的提示词，例如best quality和low quality；另外有些模型需要使用配套的EasyNegative嵌入式模型，不然效果可能不佳；在性能足够的情况下尽量跑更高的分辨率，能得到更好的效果。当前AI出图不再仅仅是关键词了，各种插件（例如动态CFG等）、各种embedding（如easynegative）、clip跳过层等都是影响画面质量的关键

10、VAE：变分自编码器，是一个美化模型，作用为滤镜+微调，一般该文件的名字中带有vae，后缀名一般为ckpt\pt等。有的大模型如Chilloutmix是自带vae的，再加入vae可能会适得其反

11、Embeddings和Hypernetworks个性化模型

12、影响图像成品最主要的几个因素：CFG的大小、Lora的使用和权重的分配。其他的因素比如分辨率、steps、采样方法等对结果的影响远没有前面三者的大

13、图片信息（PNG info）：上传AI生成的PNG图片，即可看到图片生成时所使用的各种参数，比较详尽，可以用来重现类似的图片

14、可用打开两个webUI窗口，但同时只能有一个在运行，毕竟显卡只有一张。可用在一个窗口运行的时候，

用另一个窗口填写下一组要画的东西的prompt，节省时间

15、有的模型比如chilloumix只能画带有真实感的图，而有的比如novelai只能画二次元风格图，注意模型的使用。Lora也要配合用，虽然二次元角色的Lora在真实风格的图中有体现，但是效果并不好

16、注意prompt在不同类型图片中的使用，某些词在二次元风格图片中就可以取得较好的效果，而在真实风格图片中，可能因为没有这样的样本而导致效果不好，真实中可能很难找到这种图片来训练，而二次元中可能就比较容易画出这样的效果

17、对于CFG而言，不同模型、不同风格的图片可能会有不同的适应度，即最佳的CFG可能不一致，需要调整

18、潜变量和噪声：潜变量是AI所能理解的提示词，是AI结果学习后能“理解”并抽象提取出来的东西，我们直接看就是噪声，但它能被AI理解

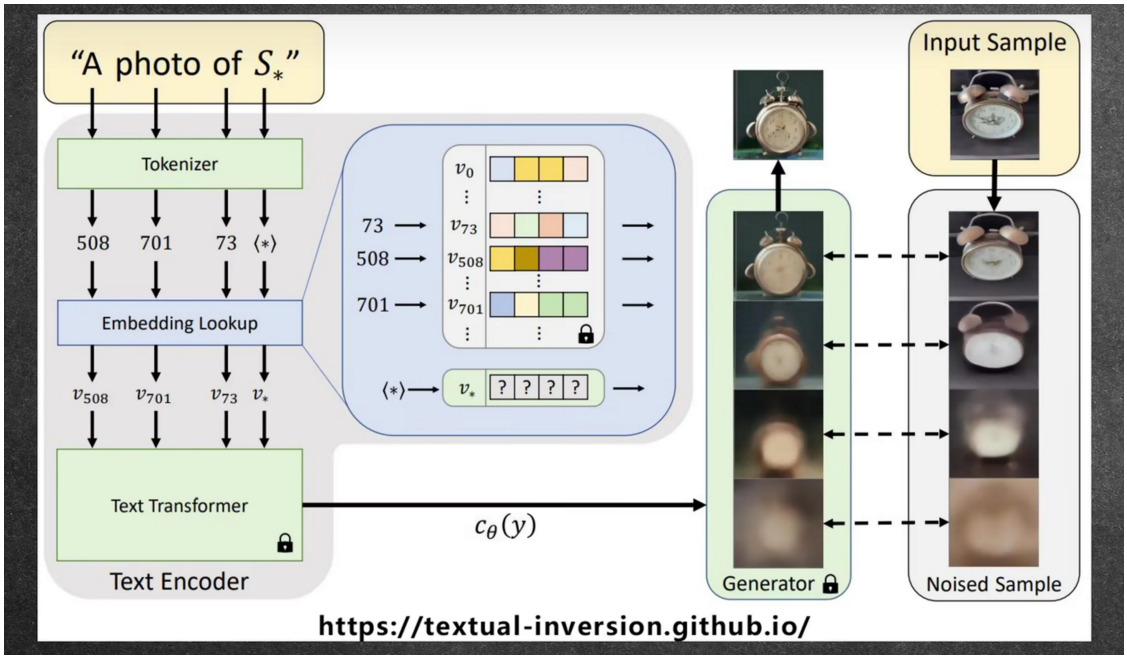
19、要画多人，需要在prompt里多加几个多人的关键词，不然很难画出多人的画作

Dreambooth：

直接用新的图片和关键词去调整U-Net里的参数，保存整个修改过的模型，所以文件很大

Lora：在U-net里增加一些层，训练这些层去调整U-net的输出，只需保存添加的层，所以文件比较小，但需要基础模型

Textual Inversion：调整clip，让他输出符合新图片的文本特征，只需要保存到学习到的特征，文件非常小



Controlnet：训练另外一个神经网络来控制U-net的输出

安装Stable Diffusion web UI

NovelAI使用的是Stable Diffusion模型，目前也有基于Stable Diffusion模型的一些开源项目，其中Stable Diffusion web UI是使用体验最好的一个

1、首先下载安装git，然后在命令行中输入

git clone https://github.com/AUTOMATIC1111/stable-diffusion-webui

直接将github上的项目拉到本地

2、创建并激活anaconda虚拟环境

3、在虚拟环境中下载cuda和pytorch

4、进入项目根目录，下载依赖库

python -m pip install -r requirements.txt

5、下载模型文件，根据种类放在不同文件夹，satble diffusion解压的文件夹中，embeddings存放提示词或其他，models文件夹中lora存放lora文件，Stable-diffusion存放主文件

6、运行web UI：可以双击webUI-user.bat，但使用python launch.py（在launch.py所在的文件夹下运行）感觉速度更快，但必须先激活虚拟环境

以当前电脑所安装的为例：

打开cmd，cd C:\Users\SupernovaGo\Desktop\stable-diffusion-webui

然后输入：venv\Scripts\activate　(这里activate是一个文件，不是命令，所以不需要加虚拟环境名称)

此时已进入虚拟环境，再输入：python launch.py，即可接第七步

2023.5.27更新：修改webui-user.bat，也可以更改启动参数，而且打开也不慢，所以以后优先使用这个

参数：在该文件的第六行，修改第六行的值即可，参数之间以空格隔开

--xformers：使用该模块

--precision full：全精度（32位），可以提高质量（但实测基本没有，不如花心思在其他方面），但是会显著提高显存占用以及内存占用

--no-half：不要使用半精度，一般和上一个参数一起使用，注意中间有一短线。如果显存太小又想画高分辨率的图（不考虑时间），则不要加上这两个参数

--lowvram和--medvram：有关显存的，前者可以以时间为代价画高分辨率的图（降低显存占用），后者以显存占用为代价提升速度。不过前者如果生成高分辨率的图，可能会生成错误的图片（也可能是采样迭代步数太小了？）

实践证明，在这台4G显存的机器上，可以跑150w像素的图，需要开启lowvram和xformers，不雅全精度那两个参数。不过速度贼慢还朴实无华，并且也比较吃内存

如果lowvram和medvram都不加，则生成的最大分辨率会自动根据你的显存来，4G显存默认应该是medvram，最高大概是90万像素，启用lowvram和xformers可以达到150万像素以上

7、当出现Running on local URL: http://127.0.0.1:7860 时，表示已打开（cmd会持续运行中），将该网站复制到网址栏并打开，即可看到页面（cmd别关了）

常见问题：

1、需要在虚拟环境中运行：

打开webui-user.bat后，第一行有venv即在虚拟环境中

2、确保依赖包安装到了虚拟环境中：

在cmd中使用stable-diffusion-webui\venv\Scripts\pip.exe list查看gfpgan、clip（open-clip-torch）是否已安装完成

3、没有检测到git

需要将git所在文件夹的bin文件夹路径加入到环境变量，然后重启cmd

4、运行launch.py中可能要下载其他的包，需耐心等待

5、第六步卡Installing gfpgan、clip或者其他（总是报错无法下载）

原因：应该是墙的问题导致无法连接上从而无法下载

解决办法：从github下载这些包的压缩包到本地，进行本地安装

1、将包解压后，放到stable-diffusion-webui\venv\Scripts目录下

2、打开cmd，cd到stable-diffusion-webui\venv\Scripts\GFPGAN-master下；

使用命令stable-diffusion-webui\venv\Scripts\python.exe -m pip install basicsr facexlib安装GFPGAN的依赖；再使用d:\\stable-diffusion-webui\venv\Scripts\python.exe -m pip install -r requirements.txt安装GFPGAN的依赖；使用d:\\stable-diffusion-webui\venv\Scripts\python.exe setup.py develop安装GFPGAN

3、再次打开stable diffusion目录下的webui.bat会发现不再要求安装gfpgan

4、clip同理：将open_clip的源文件下载到本地，解压到stable-diffusion-webui\venv\Scripts目录下

打开cmd，cd到d:\\stable-diffusion-webui\venv\Scripts\open_clip-main下

使用d:\\stable-diffusion-webui\venv\Scripts\python.exe setup.py build install安装open_clip

安装完毕后，再打开stable diffusion根目录的webui-user.bat会发现不再报错。如果还是卡在installing clip，则尝试先激活虚拟环境，并使用launch.py打开：stable-diffusion-webui\venv\Scripts\activate命令激活虚拟环境，如果命令行每行开头有(venv)的字样说明激活成功。激活成功后，直接使用python d:\\stable-diffusion-webui\launch.py即可运行。中间若报错则多次重试（如果没报错就耐心一点等，尤其是安装webUI这一行）

6、启动时显示没有名为“xformers”的模块，运行时不会使用它。

xformers模块可以减少显存需求，但似乎会降低速度

不管它也可以，不影响使用

原文链接：https://blog.csdn.net/weixin_40735291/article/details/129333599

配置Stable Diffusion web UI：安装简体中文包

1、点击Extension选项卡，Availbale子项，取消勾选localization，再把其他勾上，然后点击橙色按钮laoad from

2、安装：在下面的扩展中找到zh_CN Localization并install（在中偏下部分）

3、配置：settings选项卡，选择reload UI按钮刷新扩展列表，在extension选项卡确认已勾选中文包

4、启用：settings选项卡左侧，找到User interface选项，去到页面最底部，找到localization选项，选择中文包。回到顶部，apply settings保存设置，在reload即可生效

使用Lora：

Lora是用少量的图像，训练一个非完整的生成模型，可以根据自己的需要在一定程度上控制生成的结果，达到自己想要的特殊的、独特的效果

注意！有些Lora是有trigger words的，若prompt未包含触发词，则就算加入Lora也不会有效果

1、安放：将下载好的Lora模型放置到stable-diffusion-webui\models\Lora这个文件夹

2、加载：在“生成”的按钮下有五个小按钮，最中间的按钮即是存放各种网络模型的，例如超网络和Lora。点击Lora模型，就会自动添加到prompt

3、也可以手动调用Lora

4、微调Lora：在prompt中，lora的第三个字符串表示所占权重，即lora在图像生成中占据多少权重

xformers：可以降低显存占用

安装：在指定位置（虚拟环境）中，cmd输入pip install xformers==0.0.16rc425即可安装

使用：在平常运行webUI的指令后加上 --xformers 即可让程序在运行时使用 xformers 来进行优化，例如：python launch.py --xformers

当打开web UI后底下出现xformer的版本号时即启用成功

经过测试后发现：xformers确实可以降低显存占用，从而生成更高的分辨率而不爆显存

txt2img

2023年4月25日 22:13

提示词（Prompt）：分为正负两方面，可在https://tags.novelai.dev/ 寻找合适的英文关键词

1、AI 会按照 prompt 提示词输入的先后顺序和所分配权重来执行去噪工作；

2、AI 也会依照概率来选择性执行，如提示词之间有冲突，AI 会根据权重确定的概率来随机选择执行哪个提示词；

3、越靠前的 Tag 权重越大，比如景色Tag在前，人物就会小，相反的人物会变大或半身；

4、生成图片的大小会影响 Prompt 的效果，图片越大需要的 Prompt 越多，不然 Prompt 会相互污染；

5、Prompt 支持使用 emoji，且表现力较好

6、正负面prompt都不是越多越好，这里的太多指的不是数量多，而是重复prompt太多。太少可能达不到想要的效果，太多会使画面中出现奇怪的东西

常用正向prompt：

正向提示词	描述
HDR, UHD, 8K (HDR、 UHD、 4K、 8K和64K)	这样的质量词可以带来巨大的差异提升照片的质量
best quality	最佳质量
masterpiece	杰作
Highly detailed	画出更多详细的细节
Studio lighting	添加演播室的灯光，可以为图像添加一些漂亮的纹理
ultra-fine painting	超精细绘画
sharp focus	聚焦清晰
physically-based rendering	基于物理渲染
extreme detail description	极其详细的刻画
Professional	加入该词可以大大改善图像的色彩对比和细节
Vivid Colors	给图片添加鲜艳的色彩，可以为你的图像增添活力
Bokeh	虚化模糊了背景，突出了主体，像 iPhone 的人像模式
(EOS R8, 50mm, F1.2, 8K, RAW photo:1.2)	摄影师对相机设置的描述
High resolution scan	让你的照片具有老照片的样子赋予年代感
Sketch	素描
Painting	绘画

可参考：喂饭级stable_diffusion_webUI调参权威指南 - 知乎 (zhihu.com)

负面prompt在某一类绘图中基本都是通用的，比如画人、画风景等，如果为了简便都可以用一类prompt通用

常用negative prompt:

mutated hands and fingers, deformed, bad anatomy, poorly drawn face, mutated, extra limb, ugly, poorly drawn hands, missing limb, floating limbs, disconnected limbs, malformed hands, out of focus, long neck, long body

提示词的进阶使用：提示词的语法、分隔符和组合符（都必须是英文半角的）

一般流程：先把要描述的画面写下生成一次，根据生成结果边试边改不满意或遗漏的描述，要强调的概念用 (xxx: 1.x) 语法形式来提升权重，其中 xxx 是你要强调的词 1.x 代表要提升的比例，如 1.5 就是提升 150% 的

权重。权重取值范围 0.4-1.6，权重太小容易被忽视，太大容易拟合图像出错

1、分隔符：逗号--,

分割词缀，有一定的权重排序功能，逗号前权重更高，因此建议排序：

1、综述（图像质量+画风+镜头效果+光照效果+主题+构图）

2、主体（人物&对象+姿势+服装+道具）

3、细节（场景+环境+饰品+特征）

且尽量不要直接使用动词，而是使用名词或形容词

2、组合符：

冒号：自定义权重数值格式，例如(1girl:0.75)表示单人女孩权重为0.75

圆括号：权重乘1.1，如(1girl)，表示权重乘1.1

花括号：权重乘1.05，与圆括号用法类似

方括号：权重除1.1

复合括号：多次叠加权重，如(((1girl)))，表示权重乘1.1的三次方

3、连接符：+, and, |, _ 都可连接描述词，但各自细节效果有所不同

Sampling method和Sampling Steps：采样方法和采样步长

采样方法本身没有绝对意义上的优劣之分，只有是否合适一说，只要能达到预期的效果，就是好方法

采样步长也必须合适，太小会导致随机性很高，太大会导致效率降低（和低的steps效果差不多但是生成图像所需时间更长）

Restore face、Tiling、Highres.fix：三种优化技术

Restore face是优化面部的，首次使用会自动在cmd界面中下载扩展包，不够效果貌似一般

Tiling：CUDA的矩阵乘法优化，一般不用

Highres是在内部生成低分辨率的图，放大并添加细节后再输出

因为直接使用高分辨率画图，可能会出现奇怪的情况，比如画出了多个人。这种情况在风景图等也存在。。。超过800*800的图都使用高清修复会更好出结果

高清修复本质就是生成了一张低分辨率的小图，然后放到图生图里提升分辨率和细节

当需要画出高分辨率图像时，最好使用高清修复。不过这个方法可能还是会出现问题

放大算法若是人物类图片，优先选择R-ESRGAN 4x+，潜变量latent有时候也还行

Denosing strength重绘幅度，越大则和原来的图片越无关，若太大可能让原本好的细节变垃圾；越小则越保留原图细节，但若原图崩得厉害，重绘幅度太小就救不回来

batch count和batch size：

count是出图的轮次，而size是每一轮出图的数量。区别在于，每一轮中的出图都是同时运行的，即需要更高的配置才能运行

CFG scale：AI对描述参数的倾向程度，越高则越倾向于prompt，而随机性下降。该参数会显著影响最终结果

基本上都在4——8之间最好

低CFG：图片糊、看起来雾蒙蒙的，色彩对比弱，构图也比较差，图片结构未定型

高 CFG：图片对比度非常强，色彩非常饱和，甚至会过饱和，颜色和结构失调

Denoising strength：重绘程度，即对原图片的保留程度，和CFG有些相似。越接近于0，则在生成高分辨率图时会保留更多的低分辨率的图。而越接近1则重绘程度越高。可把数值看成百分比

一般在0.4——0.7之间比较好

注意：1、对于真人而言，如果重绘程度和CFG都很小，图片会有一种朦胧感；2、只有适合的CFG和适合的重绘程度才能生成好图，不然图片会变得非常诡异；3、CFG一般不超过10和不低于2，最大不能超过20，重绘程度一般不大于0.8，否则可能会出问题。4、重绘程度越大，则会有越多的细节，但是细节太多也不是一件好事，比如出现奇怪的东西；5、重绘幅度在0.7以前一般都是一个画风，而到了某个值以后就会发生画风变化

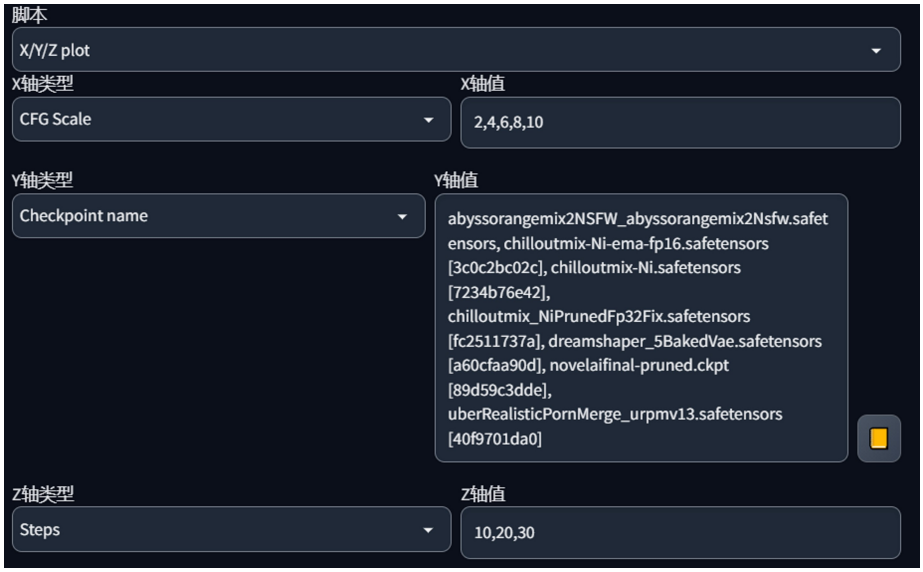
可用脚本X/Y/Z plot进行测试，发现最好的参数

脚本：方便比较使用各种参数所得到的结果

1、“|”连接符+开启“提示词矩阵”：此时会生成两张图片，第一张包含所有提示词，第二张图片不包括|之后的提示词，方便对比有无某些提示词的差异

2、X/Y/Z plot：方便对比各种不同参数所得到的图片的差异

例如以下这个例子：共生成x、y、z值的乘积张图片，其中不同值之间用逗号分割



其规律为：Z、Y、X重要性依次降低，即每次生成总是先取定一个Z或Y，再生成X的各个值

prompt S/R：是xyz图的一个子项，即提示词搜索和替换，比提示词矩阵更加好用

用法：若在prompt中输入了提示词A，则在其值中可输入A，B，C，此时会分别以提示词A、B、C生成3张图，便于比较不同提示词之间的差异

ps：也可以用该功能来方便跑不同类型词的图，比如跑5张修女和5张jk图。这样的话，记得勾选随机种子为1

画图要点：

先总体再部位，分层设置提示词

比如要画一张女生图，一般的提示词板块设置为：

1、使用的小模型：lora等

2、画面质量：highresolution、raw、best quality、CG、8K等

3、画面环境：bedroom、kitchen等

4、主体：1girl、beautiful girl、full body等（如果是单人特写可与画面环境交换次序以提升重要性）

5、动作：lying、beg、arms behind back等

6、其它细节：如bunny girl、各种装饰、长发还是短发等

img2img

2023年4月25日 22:13

图生图即使用现有的图形来生成新图形

1、可以用来批量生成类似的图

2、其主要参数和文生图的基本一致

重点关注：

1、图生图中不能使用高清修复，即图生图所产生的图片，其分辨率将低于原图的分辨率

2、Denoising strength重绘程度：最重要的一个参数，根据prompt和negative prompt在原图基础上进行重绘。当其值为1时，若有prompt，则会完全变成prompt，若没有，则会变成和原图相似元素很少的一幅新图。重绘幅度不能太大，否则会生成很奇怪的效果

3、resize mode缩放模式：当生成的图长宽比和原图不一致的时候采用什么方式进行缩放

4、反推提示词：第一次需要下载相关依赖

CLIP反推：一般是一句话描述，描述也比较准确

DeepBooru反推：结果也比较准确，但不是一句话描述而是关键词堆叠，很像prompt的风格

图片信息板块只适用于携带了图片原始生成信息的PNG图片，若是生成时没勾选保存生成信息，或者生成的不是PNG图片，则不能使用图片信息，这个时候反推就可以派上用场

Sketch绘图：会将画蒙版的画笔的颜色信息带入，重绘幅度越低，则颜色信息表现得越明显

1、绘图仍然会将输入图片的所有部分进行重绘（区别于局部重绘），比如原始衣服是白色的，用蓝色画笔在衣服上画了蒙版，则生成的时候若重绘幅度不高，则衣服倾向于蓝色

2、绘图功能甚至可以手动画图，比如用红笔画个房子的草图，然后把该图片进行上传，并在prompt里加入房子这个关键词，就可以生成一张带有房子的图片（可同时画多个元素）

inpaint局部重绘：img2img的一个功能

对于已经生成的图，若总体良好而局部有瑕疵，则可使用该功能

上传图片后，右上角的三个按钮分别为：关闭图片、撤销一步、调整画笔（蒙版）大小

处理图片过程：

1、确定蒙版区域

2、蒙版区域预处理：让蒙版区域的颜色更加贴近想要生成结果的颜色，提高成功率

3、对蒙版区域进行模糊：对蒙版区域进行模糊，产生颜色过度的效果，便于后期去噪，提高成功率

4、迭代生成图片：去噪

prompt：最简单的就是使用相关的物件对画面进行填充，遮挡住瑕疵部分，比如在prompt里填入flower、camera等

参数解析：

1、缩放模式：如果重绘区域是全图，且下面设置的长宽和原始图像的长宽不一致，则采取缩放

若重绘区域仅蒙版，则该参数不生效

2、蒙版模糊：调节图片的模糊程度，对应处理第三步，数值越小模糊的程度越高，基本不用修改

3、蒙版模式：是重绘被选中的区域还是非选中的区域，对应处理第一步

4、被蒙版蒙住的内容：选择预处理方式，对应第二步

填充：将蒙版部分模糊并进行填充

原图：不进行预处理，直接局部重绘，此时蒙版模糊参数无效

潜变量噪声：将蒙版部分使用噪声进行填充。使用潜变量会和原图差异较大，如果需要进行较大的改动则可使用潜变量处理方式

潜变量数值为0：将潜变量的数值设置为0，再填充噪声

5、重绘区域：

全图：将图片修复完成后将整张图片的宽高进行调整（如果设置的长宽和原图不匹配）

仅蒙版：仅对蒙版区域调整而不涉及全图，由于调整范围小，可以处理更大分辨率的图片。此时宽高可不用设置（仅影响像素密度，看不出差别就行）

6、仅蒙版模式的边缘预留像素：数值越小填充像素密度越高，越大则越低

像素填充密度越大，所生成的内容越多，比如生成完全不一样的东西，甚至是图中图

修改区域越大，像素填充密度则高一点更好，使得区域内内容更加完善

如果填充过后发现填充和原图之间有比较明显的边界，则试试加大采样次数（适用于低分辨率图片）。如果分辨率较高，则需要配合边缘预留像素来解决这一问题

手涂蒙版：可自由选择颜色，并进行蒙版的涂画，且一次修复可有多种颜色。最终修复时，会按照所使用的颜色的类似的颜色进行填充

Controlnet

2023年5月29日 19:16

controlnet插件：通过额外的输入来控制预训练的大模型，可以提高产出的精度，极大避免了各种差错

注意：使用该插件有着更高的显卡要求，因为要使用额外的模型，显存没个6G就不要想了

可以使用它来控制人物的骨骼、动作等，因此可以用来做动画

使用谷歌云可以满足配置要求，且colab版的webui已集成了controlnet，不用额外去启用。不过模型有的还是会在使用到的时候自动进行下载

在插件里可打开这插件，并勾选“enable”以启用

有两个模型选择区域，第一个是预处理，可以根据上传的图或者手绘草图，并对其进行预处理，以便于对生成图像的控制。一般情况下预处理器和模型基本上都是选择一样名字的。

预处理器分辨率控制强度，分辨率越高越吃显存，生成的也越精细

预处理主要就是从草图或图片中提取出边缘或深度或其它信息，从而控制新生成的图的形式

预处理主要有：

canny边缘检测：绘制出边缘

此外还有HED边缘检测、PiDiNet边缘检测，都大差不差

MiReS深度信息估算：绘制深度图，一般用于有纵深的图，尤其是照片类型

LeRes深度信息估算，只是算法不同，也是深度图

M-LSD线条检测：绘制的预览图是很正直的线条，适合建筑和室内设计

normal法线图：预览图是通过RGB颜色通道来表示凹凸

openpose姿势检测：预处理图为姿势图，对应动作

openpose姿态及手部检测：带手部的姿势检测

scribble涂鸦：通过画布画一张草图，通过tag限定场景，即可创造出图像

语义分割：将画面的不同物体以不同部分进行分割

controlnet多通道混合模式：

在设置--controlnet--最大网络数调高，比如3，即可启用多个控制网络同时工作

weight：控制controlnet强度

引导介入时机（start）和引导退出时机（end）：生成图像的百分之多少开始controlnet的介入和退出

Deforum动画生成

2023年6月1日 0:18

Deforum是Stable diffusion WebUI的一个插件，可以根据提示词生成动画

该插件（又或许是所有动画生成软件都具有的）一个明显特点就是使用“key frame”，即关键帧，几乎每种变动，比如视角转变、controlnet介入等，都可以设置入点出点，以此达到关键帧的目的

注意：左上角的主模型在Deforum里仍然适用，如果没有在keyframes的Checkpoint选项卡中设置，则默认使用左上角的主模型

主要功能板块：

1、Run：AI模型基本参数调节，如采样方法和步数、视频宽高等。这些基本参数是不可以打关键帧的、最基本的参数

2、Keyframes：设置视频其它参数比如最大帧数，并且可以打关键帧

下边还有一些其它的小功能板块，比如镜头的移动，噪声的选取等，甚至可以给主模型打关键帧，在不同时刻使用不同的主模型

3、Prompts：提示词板块，设置贯穿全视频的正向和反向提示词，还可设置在不同帧时不同物体的开始出现，也就是提示词关键帧

注意Prompts的格式是JSON格式，不能更改。如果要加新的关键帧也要符合原来的格式

4、Init：设置初始帧或视频，即以某张图或某段视频为开始，继续生成动画

5、Controlnet：使用控制网络达到更好的效果

可同时启用多个网络，且可设置控制网络介入的关键帧

6、Hybrid Video：hybrid表示“混合”，即混合视频或融合视频

7、Output：输出选项，可设置FPS、是否超分辨率以及超分辨率的算法、是否保存生成的每一帧的图片等

设置好后，点击右上角Generate即可生成视频，生成完成后可点击查看或到文件夹里去找

视频一般在outputs的img2img文件夹里边。如果未勾选delete imgs，则在文件夹里也会保存生成的每一帧

Latent Diffusion模型原理

2023年4月25日　23:00

Diffusion的本质就是加噪的过程，也就是给训练的图片添加随机噪声，重复T步；而unet根据加噪后的图片一步步denoise还原成原来的图像的过程，也就是去噪的过程，即反向扩散

stable diffusion基于latent diffusion模型。Diffusion是一种相比GAN更容易训练的模型，然而它是自回归的（自回归体现在diffusion部分，也就是加噪部分），需要反复迭代计算，这就导致训练和推理的代价都很高。Latent diffusion model（LDM）在diffusion模型基础上引入了潜在空间（latent space），将图片从像素空间压缩到潜在空间并进行处理，能大大减少计算复杂度，同时效果也很好

Latent Diffusion模型可划分为三层

1、Text Encoder模块：它本质是一个clip模型

a：Clip（Contrastive Language-Image Pre-training）是将文本和图片联系起来的预训练、多模态模型，由openai开发，结合了自然语言处理和计算机视觉，旨在理解图像和文本之间的语义关系，在文生图中是非常重要的组件。

b：在stable diffusion模型中，实际上只用到了文本编码部分（因为已经是训练好了的，图像编码部分就不需要了）

c：CLIP skip，即Clip模型的跳层连接（skip connections），是在神经网络中添加直接连接，将输入直接传递到某些层或层之间，允许低级特征和高级特征进行融合，增加模型的表达能力和学习能力

d：文本编码层将输入文本进行编码，作为图像生成器的输入。在stable diffusion中，向量的维度为768

2、Image Generator图像生成器：

a：由UNet（一个噪音预测网络）和Scheduler（调度器）组成，在潜在空间中逐步处理信息

b：输入：向量表示的文本（嵌入后）以及一个初始化的多维数组（张量）组成的噪声（noise）

c：输出：经过处理的信息数组

d：Schedule决定每个加噪的步骤添加多少噪声，可以是相同的量，更多的是先少后多

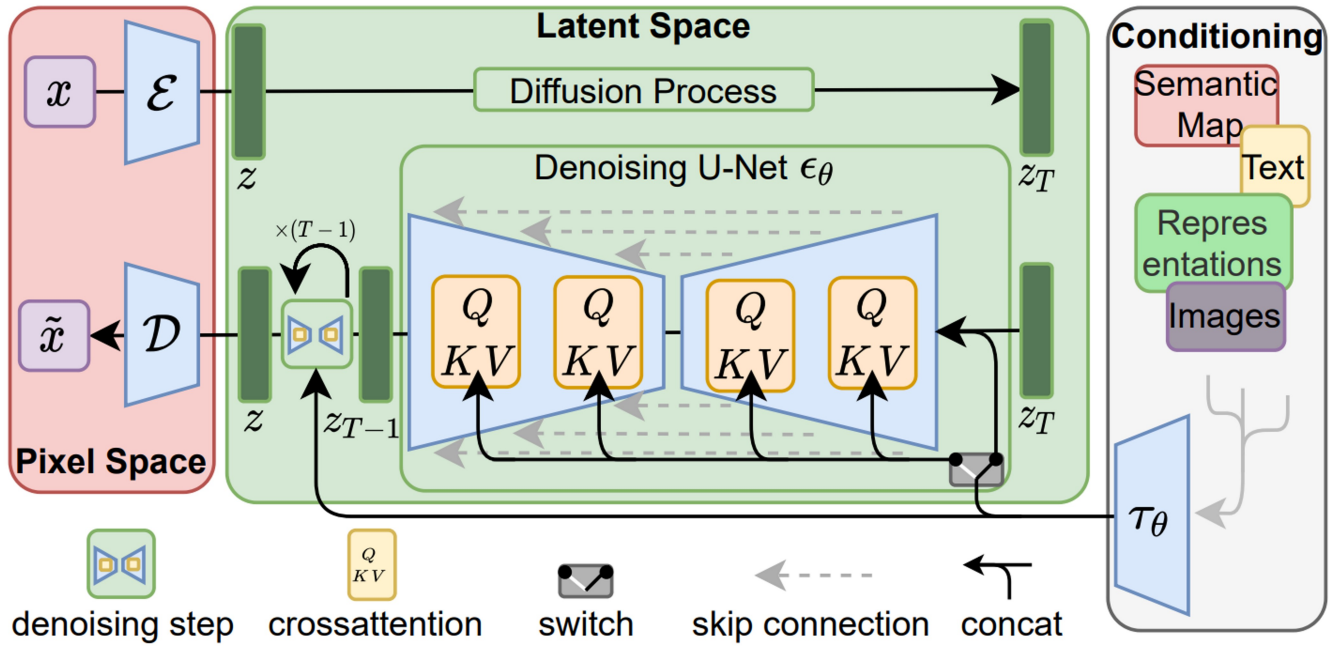
3、基于VAE的图像编解码器

a：由自动编码解码器（Autoencoder Decoder）组成

b：输入：经过处理后的信息数组

c：输出：生成的图像，维度为：红绿蓝+宽高

diffusion模型结构：



解释：红框、绿框、白框可以分别代表VAE编解码器、图像生成器、clip文本编码器

1、x和$\tilde{x}$即分别代表训练用的图片和输出得到的图片的特征，z代表被映射到隐空

间的图片特征，z_T代表加了T步噪声后的隐空间图片特征

2、$\mathcal{E}$和$\mathcal{D}$分别代表将像素空间的特征映射到隐空间特征的编码器和将隐空间特征映射到像素空间的解码器

3、cross attention，即交叉自注意力，和自注意力类似，也是键值对注意力的一种。不同的是，交叉注意力是计算多个不同的输入特征不同位置之间的关联性，将其中一个输入用作查询Q的计算，另外一个用作键和值的计算，有利于信息交流与融合

4、concat表示张量拼接，这里将来自不同特征的两个张量进行拼接，有利于特征融合、信息传递等

5、每个denoising步骤都是多层的交叉注意力机制+跳层连接，得到该步的输出

训练Diffusion的过程：

1、随机挑选一张训练集中的图像，并为其添加随机的高斯噪音，重复T步，此即扩散过程

添加噪声的原因：使模型更加容易学习到图片中的特征，并且随机噪声还增加了模型生成时的多样性

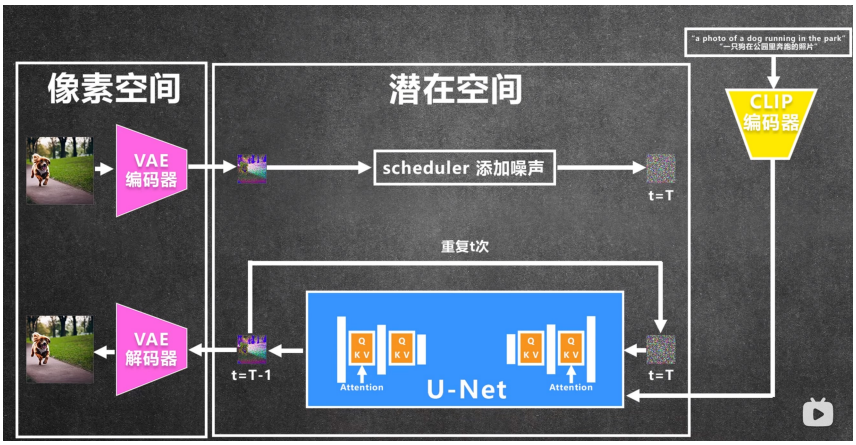
2、将加噪后的图像作为训练图像进行训练，训练的是U-net网络

a：u-net是一个噪声预测器，其功能是预测出图像中有哪些噪音，如果它能完好地从加噪后的噪音图像中准确地预测出噪音来（也就是还原原来的图像），则训练成功

b：u-net的输入是加噪图片以及加噪的步数，也就是加了多少噪声。u-net预测出噪声后，将加噪图片减去噪声即可得到原图

c：但如果噪声很多，u-net无法完全预测出原图的细节，而只能得到一个大概的轮廓。这时把预测的图当作原图，再加上比之前少一点的噪声进去，比如加噪过程总共是50，这一次就加49，再次预测。重复该过程，得到原图

d：为了把文字的内容加入去引导图片生成的内容，首先需要把文本使用clip模型转换成文本特征，然后把文本特征也加入到u-net的输入中。在u-net中添加交叉注意力机制，通过文本特征来引导噪声的预测。但只是这样，只能得到和文本相关的，不能精确得到精确描述文本内容的图片。这时需要使用classifier free guidance（无分类引导）：首先预测两个噪声，一个有文本特征引导，一个没有，然后两个相减，即得到了再文本引导下改变的不同的地方。为了让生成的图像更加精确，就把这种改变的信号放大，这个放大倍数就是GFC的值，一般在7倍左右，放得越大就越精确。将放大的信号与没有文本特征引导的噪声相加，就得到了加强文本引导的噪声。这之后，减去噪声得到模糊的原图，重新加上噪声继续预测，重复多次。若有负向提示词，则两个预测的噪声变成分别带有正向和负向提示词的噪声，然后相减得到远离负向、接近正向提示词的噪声



3、解码生成图像：经过潜在空间的反向扩散过程后，得到一幅图像

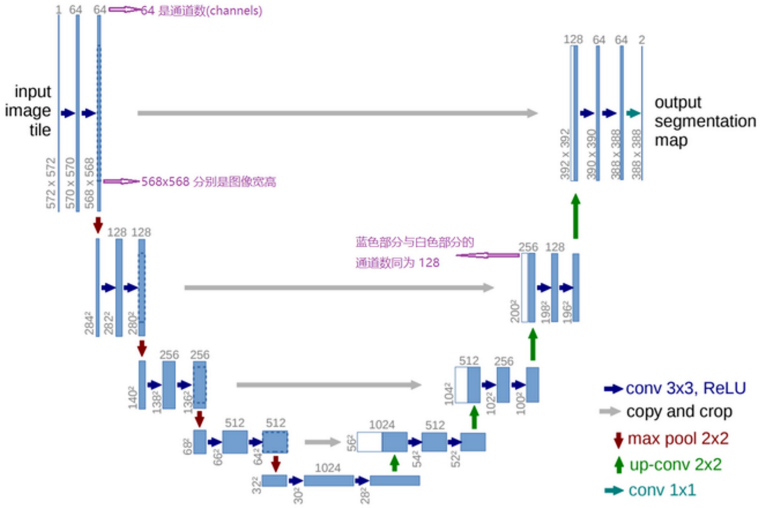
4、优化迭代模型：每次生成图像后，求loss，用梯度下降优化参数

生成图片的过程（推理）：不再需要上方的图像编码和扩散的过程

随机生成一个噪声，加上文本引导（正向prompt和负向prompt），生成结果

U-Net网络：一个基于卷积的语义分割网络

该网络的结构是对称的，因为形似英文字母U而被称为U-Net，采用Encoder-Decoder框架



解释：

conv 3x3，relu表示使用3x3的卷积进行信息提取，并在最后使用relu激活函数

copy and crop：表示复制和裁切，是一种针对图像数据的数据增强技术，尤其是在语义分割任务中常常被使用。语义分割是对图像中每个像素分配一个类别标签，而copy and crop则是先复制图片，再进行裁切，裁切出感兴趣的区域，再对其进行标记。这里复制和裁切后，再进行跳层连接（拼接张量）

其它和计算机视觉有关的概念：

1、高频细节与低频细节：高频细节通常指的是图像中快速变化、细小的细节信息。这些细节通常包括边缘、纹理、细线等高频变化的部分；低频细节指的是图像中缓慢变化、较大范围的细节信息。这些细节通常包括图像的整体亮度、颜色分布和大范围的平滑区域

2、patch-based（基于块）方法：将输入图像分割为小块（patches），然后对每个小块独立地进行处理或分析。每个小块可以被视为一个独立的样本，并且可以使用传统的或深度学习的方法进行单独处理

3、图像流形（image manifold）：高维图像数据空间中的一个低维嵌入子空间，其中图像具有良好的连续性和结构性

4、VQregularized：VQ-VAE（Vector Quantized Variational Autoencoder）的正则化方法，VQ-VAE是VAE的一种扩展

5、FID、IS、PSNR、SSIM、LPIPS：均为衡量图像的指标

6、degradation（降质）：指的是对原始图像进行有意或无意的损坏、改变或降低质量的过程。这种降质操作可以是人为手动进行的，也可以是由于噪声、压缩、模糊等外部因素引起的

7、Classifier-free guidance（无分类引导）：在生成图像的过程中，不依赖分类器来指导或约束结果，可以提供更灵活和多样化的图像生成能力

各种疑难杂症

2023年5月27日 22:12

1、生成的图像崩了，优先调：lora的权重、增大到合适的采样迭代步数、增大分辨率，CFG不要太高。尝试使用高清修复而不要直接生成高分辨率的图

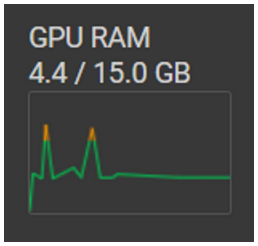
2、三种提升分辨率的方法

1、高清修复hi-res fix：文生图的一个功能，本质上是先生成了原始分辨率的图，再使用图生图的功能，以原图为蓝本生成一张全新的更高的分辨率的图

2、SD upscale：图生图的一个功能。本质是把原图切分成各个小块，然后每个块分别超分辨率。这个功能在“脚本”里边，需手动开启。缩放系数就是放大倍数，而图块重叠的像素，即每个块与块之间的“缓冲带”，避免最后拼接成图时出现接缝

3、附加功能放大算法：只是对原图进行AI分辨率放大，没有生成新细节或物体的功能，效果一般。不过负担小，方便，一般用于图片生成后进行分辨率放大

3、每次绘图，当一张图片快画完时，显存的占用会显著上升。若一开始显存占用就很高了，那么极有可能在即将结束的时候爆显存



4、正向Prompts问题：

1、正向prompts不是越多越好，尤其是含有互相冲突的prompts的情况下，比如同时含有“arms behind head”和“press face”，这样冲突的词，一般模型会以权重为依据随机挑选一个，但若冲突部分太多，则可能导致画面崩坏。简单即是多。

2、根据模型来选择合适的prompt，比如风景checkpoint就可以用wide lens等，而人物checkpoint则可用动作、外貌的提示词。其它的，比如一个正经画妹子的模型，强行塞不正经的提示词，因为训练的时候就基本没有这种样本，所以很难画出好东西