

stable diffusion prompt 创作技巧，各参数效果
分析及一份样例

技巧/内容综合区域 (/t/skillcontent)
二次元图像模型 (/t/2d)



分享



Rion (/u/27740892032377223994) 5 天前

1楼

导语：

注意到一些群友对 prompt 中的各个参数的效果理解得比较模糊，所以稍微写一下各个参数的用途和效果，并串接一些作者使用到的技巧和理解，以及一份样例 prompt。

其实不了解也没有什么关系，毕竟即使是看上去很奇怪的 prompt，也不是不可能生成理想的图。

不过了解一下的话，也许可以写出更加稳定的 prompt 也说不定。

本文面向的对象是有一定经验的创作者来着，所以难度会比较大。（其实主要是写完之后感觉初学者不一定看得懂来着...）

注：其实我也不是很了解啦，各位随便看看就好了。

目录：

- (1) Clip skip
- (2) 模型
- (3) CFG scale
- (4) TAG
 - (4-1) 75
 - (4-2) 连续性
 - (4-3) 元素溢出
 - (4-4) 泛化 TAG
 - (4-5) 权重均衡
 - (4-6) 负面TAG
- (5) 小的总结
- (6) Sampler

- ◀ • (7) Step
- ◀ • (8) 分辨率相关
 - 样例

正文：

(1) Clip skip:

首先谈谈对整体影响最大的参数：Clip skip（在 setting 中设置）。

Clip Skip 是一个非常强的参数，它表示将 Tag 抽取到 Clip 模型后，在后面生成图像的过程中会跳过最后的多少层 Clip 提前停止。

具体的说，如果 TAG 描述为，“一个站在森林中的小女孩”，那么第一层 Clip 可能描述“一个人”，第二层可能描述“小女孩”，第三层可能是“一个站着的小女孩”，以此类推。后面的层会比前面的层描述的更具体，这些层会共同组成 Clip，参与后面的图像生成。而 Clip Skip 会控制系统忽略 Clip 的最后的若干层，这些层往往是描述的最具体，综合性最高的那几层。

简单的讲，它可以理解成一个增强版的 CFG，Clip Skip 越高，系统对 TAG 的服从度越低。这种服从度的降低主要表现在细节和各个 TAG 的协作上，对于一些复杂的修饰词，Clip skip 升高会导致这些词效果的骤降。

它作用于 CFG 之前，或者说，它作用于大多数参数之前。所以它的效力非常非常强，尤其是它的最后一层，Clip skip = 1 和 Clip skip = 2 的效果差距会非常明显。比如上面的 prompt，如果把 Clip skip 设为 1，那么就很有可能死图（指生成的图像不能被人类所接受）（原因是权重大的 TAG 太多了）。

（不过好像没有看到 Clip skip = 0 的设置，也许它的最后一层是必须跳过的？）

因为 Clip skip 的效力非常非常强，所以从某种意义上，它又不是那么重要了...因为它并不适合作为最后效果的精调参数，而作为一个粗调参数，绝大多数情况下它都是不需要改变的。如果要改变它，整个 prompt 就要重写以保证稳定性。这是不可接受也没有必要的。

（嗯，prompt 的稳定性指 prompt 稳定生成效果比较好的图像的能力）

比较推荐的设置是 Clip skip = 2，这样 TAG 上的权重均衡会简单很多。如果 Clip skip = 1 的话，各个 TAG 的效果就很强，对 TAG 进行加权时要非常谨慎（Clip skip = 2 时，TAG 的权重设到 1.5 都是可以的，但是 Clip skip = 1 时这么做往往会死图）（不过如果不管 TAG 间的权重均衡，不使用加权，那直接 Clip skip = 1 会更好也说不定？）

Clip skip 由 2->3（以及之后的增加）影响不会很大（相比 1->2），不过 Clip skip = 2 时，TAG 的效力就已经比较弱了，不建议再升高。

总之，这个参数既重要也不重要，重要在于它对最终效果影响很大，不重要在于反正我们一般也不会去动它...（它甚至没有被放到主界面，而是放到了 setting 里）

(2) 模型：

这个大家应该都知道是干什么用的（大概？），但是还有两点需要注意：

第一点是，不同模型对同一个 TAG 的敏感程度是不相同的。

比如说，在同一个 Clip skip 下，Anything 的词汇广度（指能够起作用的 TAG 词汇数量）明显优于 Counterfeit。

从一个方向讲，同一个 TAG 在不同模型下的效果也是不一样的，比如说 small breasts, Anything 和 Counterfeit 都知道这个词是在干什么，但是橘子会在做到这一点的时候，经常会顺便把这个部位裸露出来，就很让人讨厌...

第二点是，每个模型会有自己的特征权重。

这里的特征权重不是说模型的画风，而是说模型在细节上的偏向，比如 Counterfeit 偏向粉发，偏向大量的木头作为背景，偏向无意义的延申的衣服布料，橘子偏向紧身衣，偏向背光，甚至是放大器也会有这种偏向，比如说 Latent (nearest-exact) 偏向花，偏向条纹等等。

这种特征权重往往不是我们所需要的，所以可以通过增加负面进行删除。

不过要注意的是，特征画风一般不能作为负面，像 3D 这种过于宽泛并且会攻击橘子核心画风的词，是不能直接加到负面里去的（这一部分会在后面具体说明）。

(3) CFG scale：

这个参数的话，嗯，不好说。

简单的讲，它是对“系统对 TAG 服从程度”的精调，CFG 越大，服从程度越高。

理论上讲，CFG 只与 TAG 有关。对于不同的 TAG 权重组合，应该使用不同的 CFG 进行均衡。

但是不同模型对 CFG 的敏感程度也不一样，每种模型适应的 CFG 范围各不相同。

所以，比起同时调 TAG 和 CFG，也许根据模型将 CFG 固定，然后调整 TAG 会比较好？

模型对应的 CFG，一般在那个模型的说明书中会有提到。

（没有说明书的话...就慢慢试吧）

Anything 是个例外，正如其名，Anything 对任何 TAG，任何 CFG，都会有所表现。不会像某些模型一样，CFG 调出舒适区，就基本不能出图了。

(4) TAG:

这个是 prompt 的核心。

所谓的“普适的 TAG 表”是没有意义的，不同模型对 TAG 的广度和深度都不同。

所以哪些 TAG 有用，效果好，哪些 TAG 间有冲突，哪些 TAG 可以相容，都只能通过实践得出。

从这一点来说，对于 prompt 的创作，经验是非常重要的。

这里谈论一下 TAG 的一些特性：

(4-1) 75:

嗯，WebUI TAG 右上角会有一个词数显示。如果词数超过了 75 个，那么系统会将 TAG 每 75 个词数分一组，各组之间通过 AND 连接。（这里的“词数”并不是具体的描述，一个 TAG 可能对应多个“词数”）

这意味着，超过 75 个词数后，很难精确地控制权重分配。

所以我使用的 TAG 一般都是 75 以内的来着。

(4-2) 连续性：

同类 TAG 连在一起会比较好，吗？

如果保证了 75 以内的话，似乎很难说会不会更好。

不过距离近的 TAG 之间会更容易相连，这一点是确定的。

这意味着，耦合度高的 TAG 放到一起去，且不说会不会更好，至少不会出问题。而且 TAG 可读性也会增加...所以还是把相关联的写到一起吧。

(4-3) 元素溢出：

这个是元素冲突的一种。

举个例子，TAG 中既有“白发”，又有“黑发”，那么二者会产生冲突，导致元素溢出。

元素溢出不意味着元素融合，所以画面主体的头发并不会既黑又白（一般情况下）

一般而言，结果会是冲突的一个对象加载到主体上，另一个加载到客体上。

比如 TAG 中有一只女孩和一只猫，白发加载到女孩上，黑发就会加载到猫上。

如果没有客体，元素溢出就可能导致未曾预料的客体的产生。比如从单人变成多人。

利用这个特性，其实多人描写是很简单的。

另外，像“白发”，“黑发”这种明显的冲突是很容易被注意到的。但是导致元素溢出的原因往往不会这么显然。

比如说 (white hair), short hair，这个也可能导致元素溢出。

这一点的话，我之前的贴子里有说过的来着。

(4-4) 泛化 TAG:

有一类 TAG 的影响是整体性的：这些 TAG 的效果，并不是向图像中加入某些元素，而是向图像中的所有元素，加入这类 TAG 代表的特性，这里称呼它为泛化 TAG。

常见的泛化 TAG 包括大多数画风提示词，大多数全局抽象词等等。

这类词有很多有趣的用法，之前的贴子里有提到过，这里略过。

(4-5) 权重均衡:

TAG 的一个很理想的状态是：内部自均衡，外部与 CFG，模型，clip skip 相均衡。这种状态下，prompt 的稳定性会有显著的提高。

内部自均衡的话，要求各个 TAG 的权重之间相均衡，人物与背景的权重均衡，人物自身的权重均衡，背景自身的权重均衡。

外部均衡的话，要求 TAG 与模型特点相均衡，TAG 权重与 Clip + CFG 相均衡，CFG 与模型相均衡。

这个的话，不好说有什么诀窍吧...主要还是靠经验？

(4-6) 负面 TAG:

负面 TAG 是非常重要的。

和正面 TAG 一样，负面 TAG 也要受 CFG，Clip，模型的影响。

好的负面 TAG 会给 prompt 的表现带来较大的提升。

不过一句 EasyNegative 走天下...也不是不行...

嗯，负面的另一个作用在于均衡模型带来的特征权重。

比如样例中用 wood,stone 来均衡 Counterfeit 自带的木石背景，或者用特殊的负面限制橘子的 3D 效果（以前的帖子里有，不再赘述）。

最后，不要把泛化 TAG 写到负面里，而应该把它拆成多个实体 TAG 再写到负面里。

(5) 小的总结：

上面是 Clip skip，模型，CFG，TAG 的效果...

这四部分是 prompt 的重点所在，控制着“最终会产生关于什么的图片”。

(6) Sampler：

这个不是很好说...

它的影响其实不是那么大，而且一般来说，每个模型都会有比较适合的 Sampler。

说到底常用的 Sampler 只有那么几个...

而且 Sampler 不会影响权重均衡。

嘛，就...比较无所谓的。

不过不同的 Sampler 也会有不同的特色就是了。

upscaler 同理。

(7) Steps：

不同的模型对应不同的采样器，有不同的 Steps 要求。

超过 20 后，Steps 越大，一些地方会越精细，但是细节会越少。

少于 20 的话，嗯...

一般 20 - 30 就够了，具体还是要参照各个模型的说明书。

(8) 分辨率相关：

与分辨率相关的参数包括：Size, Hires upscale, denoising strength。

如果不使用高清修复的话，后面两个参数不存在。

这三个参数是要均衡的，其中 Size 还需要和之前的 TAG 均衡。

Size 过大的话，TAG 要考虑增加权重，人物与背景之间得 TAG 均衡需要重新调整。

嗯，假设 Size 和 TAG 已经达到均衡了，那么下一步是根据需要确定 Hires upscale，然后根据生成的图片的效果调整 denoising strength。

denoising strength 相对较小时，会有大量的噪点，upscaler 会按自己的喜好填充这些噪点。但是一般会有很明显的杂乱表现。

denoising strength 相对较大时，放大前后的图片区别会很大。这意味着...实际上真正影响图片结果的就是放大的过程了...一般不会认为这是很好的结果的...大概？

upscaler 的选择对二者的均衡会有影响。

比如 Latent (nearest-exact) 对应的 denoising strength 会显著大于 LDSR。

结语：

只是随便写一写来着，如果有问题可以评论区指出。

虽然大家练练 lora 很开心，但是 prompt 并不是随便就可以丢掉的东西来着。

初学者可能会看不懂，这个是正常的（大概？

样例：

如附图。

Prompt:

正面 TAG:

((masterpiece)),((best quality)),((illustration)),((Chinese_style)),

((beautiful detailed girl)),cat girl,tea,standing,

((delicate cute face)),blue_eyes,watery_eyes,

bob_cut,white_hair,flower,

◀ (white hanfu), silky, pink, luxuriant, layered dress,
(in_spring), lake, beautiful_landscape, light particles,

负面 TAG:

EasyNegative, ((wood, stone)),

其他参数:

Model: Counterfeit-V2.5,

CFG scale: 10,

Sampler: DPM++ SDE Karras,

Steps: 30,

Size: 432x704, Hires upscale: 2, Hires upscaler: Latent (nearest-exact), Denoising strength:
0.54

Clip skip: 2,



回复

👍 chy233666 (/u/3115539226929654852), 我太丑 (/u/21001520664106674402) 和 cocock9 (/u/1795413942823008383) 觉得很赞

5 天后



我太丑 (/u/21001520664106674402) 2 小时前

2楼

谢谢大佬，受教了