## a一种低成本的个人图像生成方式

## 0.背景

生离死别是人生常态，伴随之的是人情感的跌宕起伏。通过新的AIGC（AI Generated Content）技术，我们可以一定程度地弥补情感遗憾。例如丢失的宠物可以再现于图片或视频中。即使背景更改（例如主人搬家），动作变化（产生新的姿态），通过对AIGC进行正确指示都能得到所求目标图片，从而达到慰藉宠物主人的目的。虽然ps技术也能生产新图片，但一是内容丰富度有限，二是ps的过程是人为主导、降低了带给人的惊喜程度。

2023年是AIGC元年，出现了文生图大模型DALL-E2和Stable Diffusion，出现了OpenAI的文本对话大模型ChatGPT。虽然AI技术的历史已经有30多年，当年IBM的“深蓝”下国际象棋也产生了巨大的轰动，但AI技术一直没有没有规模商用、且成本高昂。近年来，随着通用算力极大提升、算法的日渐成熟、数据的极大丰富，终于使得AI技术突破天花板、走入寻常百姓家。

## 1.介绍

在深度学习文本到图像（text-to-image）和图像到图像（image-to-image）生成领域，Stable Diffusion（简称SD）模型是当前主流的效果最好的模型之一，并且由于其完全开源（代码，数据，模型全部开源），它吸引了大量AI研究者和爱好者的参与，其模型参数达到了10亿级。一般而言，图像生成模型所生成的内容与其训练时所使用的数据集有非常大的相关性，比如使用原始的 SD进行人物或者动物生成，得到的结果往往会与网络上常见的一些图像比较相似，这是因为它的训练集中包含了大量类似的图像，模型直接从图库中提取获得了图片的元素。而如果想要求SD生成与个人相关的一些指定图像，比如与你的宠物猫相似的图像，就会十分困难，因为原始的SD模型并没有在训练集中见过你的宠物猫的样子。解决上述问题的一个方法就是模型二次训练。因为SD模型是开源的，我们可以使用包含个人宠物照片的数据集对它进行重新训练，使得它能产生类似的图像。但是重新训练SD并不是一件容易的事情，它对数据集的数量和算力都有不小的需求。据官方披露，从零开始训练的SD模型共花费了150000个A100 GPU 小时，这是一笔数百万的投资，不是一般企业和家庭能负担的起的。

SD是一个基于Latent的扩散模型，它在UNet中引入文本条件来实现基于文本生成图像。SD的核心来源于[Latent Diffusion](https://link.zhihu.com/?target=https://arxiv.org/abs/2112.10752" \t "/Users/romi_li/Documents\\x/_blank)这个工作，常规的扩散模型是基于Pixel的生成模型，而Latent Diffusion是基于Latent的生成模型，它先采用一个autoencoder将图像压缩到Latent空间，然后用扩散模型来生成图像的Latents，最后送入autoencoder的decoder模块就可以得到生成的图像。

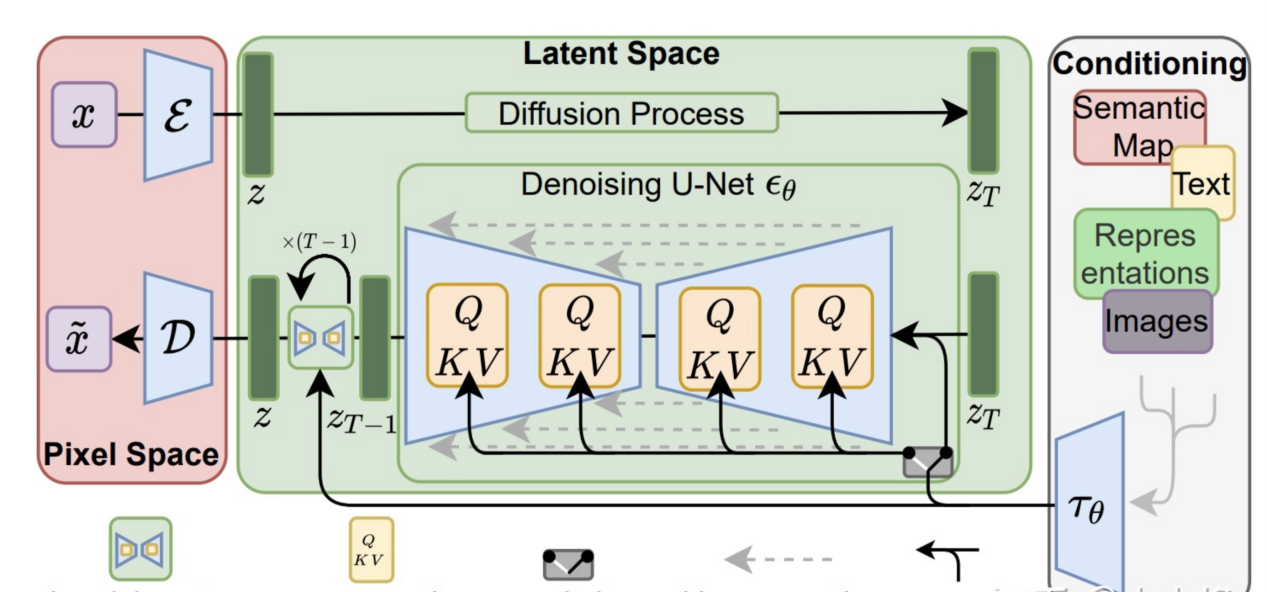


图1 SD原理

相比图像生成，AIGC的另一个领域：大语言模型（LLM），对二次训练和模型微调的需求更为强烈，从而促使了LoRA（Low-Rank Adaptation of Large Language Models）算法的出现。LoRA算法能够在原始的大语言模型的基础上，需求较少的训练数据和算力，对模型进行微调以使其适应一些定制化的应用场景。如图2所示，如果把一个预训练好的模型用矩阵表示，这个模型的输入用向量表示，它原本的输出则可以表示成。而LoRA算法的原理就是在原本模型的基础上添加一个旁路，这个旁路由两个向量和组成，添加了旁路后的模型的输出变为。而我们使用新的数据集进行二次训练的过程，就是对向量和训练的过程，由于向量和所包含的参数数量要远少于原本的模型，所以训练所需要的算力也就更小。

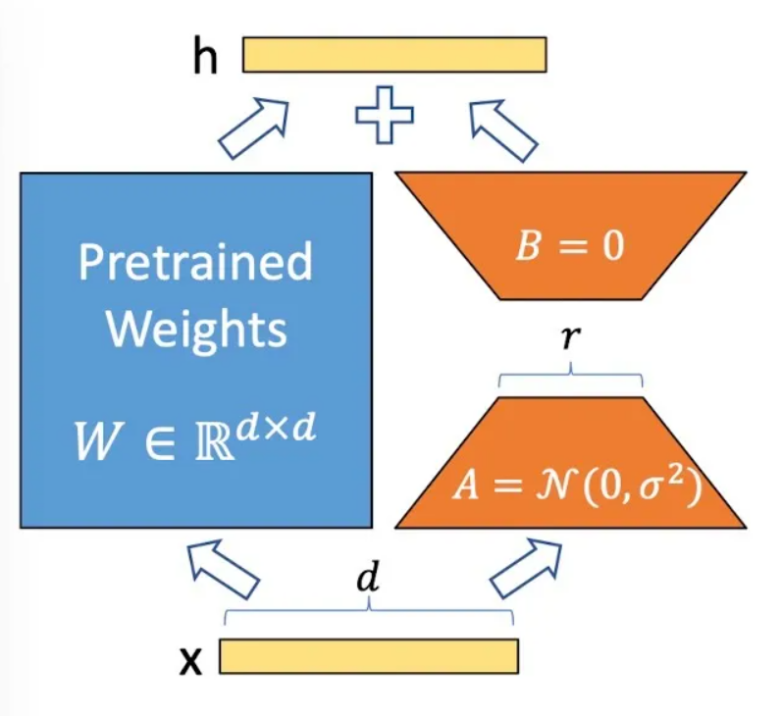


图 2 LoRA算法原理

通过LoRA算法的原理，我们不难发现其实它可以被应用于任何预训练好的模型，而不仅仅是大语言模型。所以本文中将尝试使用LoRA算法对图像生成领域的模型SD进行二次训练，以个人宠物猫图片数据作为输入，以期在个人电脑的有限算力上、获得一个能够生成符合个人需求图像的新模型。

## 2.原始模型测试

首先在在MD的Prompt文本框内输入提示词 ”masterpiece, best quality, cat, looking at viewer, full body, standing, outdoors, flower field, landscape, slit pupil“（翻译：”杰作，最好的质量，猫，看着观众，全身，站着，户外，花田，风景，狭缝瞳孔“），其余参数均使用默认设置，以此随机生成六张图像，结果如图 3所示。可以发现，在没有经过个人素材图片的训练前，原始模型使用不同的随机种子生成的不同图像之间猫的风格和种类都是迥异的、但是也感觉不切实际，这是因为其原始模型所使用的训练数据集本身具有较高的多样性。模型从大数据所搜索到的猫中选择了一部分作为基本数据生成这些并不具有指定性的图片，范围太过宽泛。这些图片可以作为实验用，但与实际场景不契合，无法与宠物主人心理共鸣。如果我们希望模型能够稳定生成特定的、与个人宠物猫和所在环境有强关联的图像，就需要使用个人数据集对其进行二次训练。

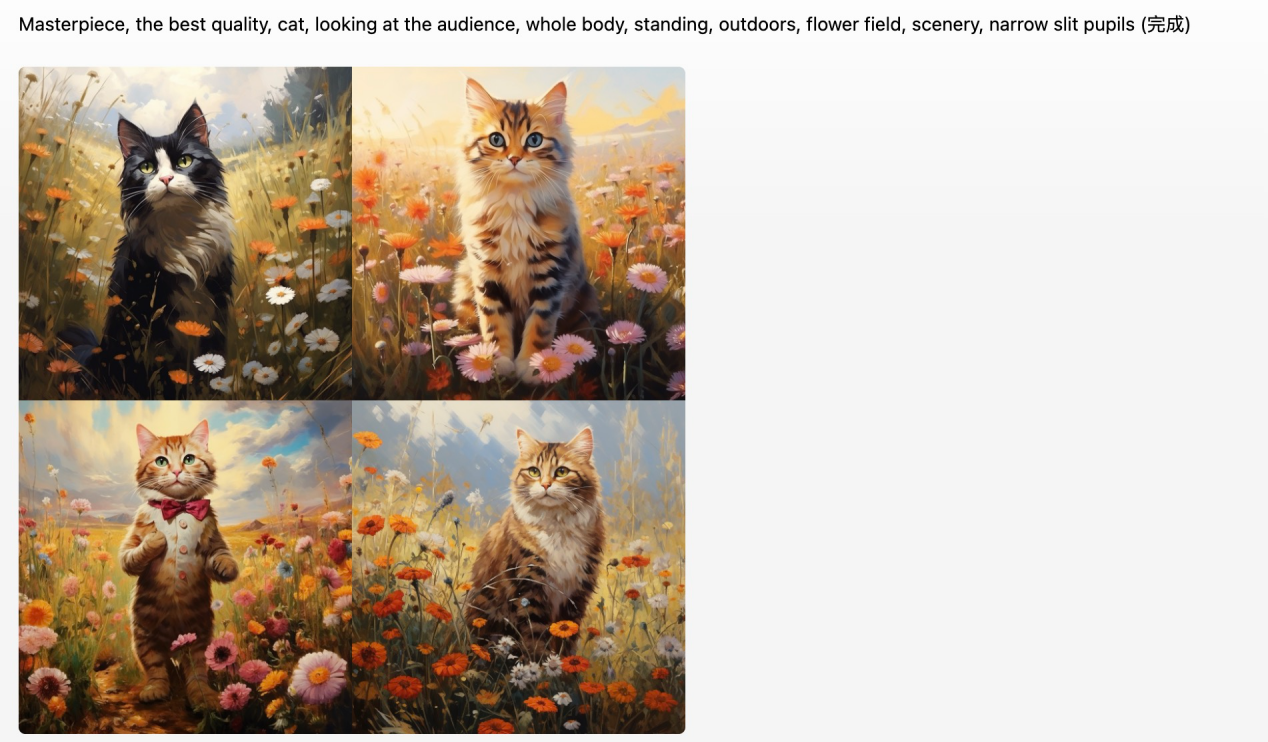


图 3 Stable Diffusion 原始模型生成图像

## 3.模型二次训练

## ·**数据集准备**

由于LoRA算法的特性，我们仅需要准备较少的训练数据就可以对模型进行二次训练。在本文中，我们共收集了32张同一只宠物猫的真实图像作为训练集，最终目的是生成以这只猫咪为蓝本的多元图片，不受动作和背景的限制。其中图 4展示了部分训练集样本。



图 4 部分训练集样本

但是用于训练SD模型的数据集不仅需要图像，还需要对应的描述该图像的文本标签。为了减少人工工作量，我们使用了另外一个AIGC模型CLIP。CLIP是一个图像到文本（image-to-text）生成模型，其功能就是通过输入一张图像来输出描述图像内容的文本。我们可以通过CLIP为训练集中的每一张图像生成文本描述。此处的目标是将此素材图片添加上花朵植物等要素以完成宠物主人的心愿。

#### ·正式训练

借助于Kohya's GUI这个开源工具，我们可以方便地使用LoRA算法对SD模型进行二次训练。

* 首先配置训练过程中的相关参数，比如学习率，batch大小等。如果我们使用LoRA常见的参数配置进行训练，还是难以在仅有6GB VRAM的图形卡上完成训练过程。
* 调节训练参数中的梯度检查点（gradient checkpointing）配置，使模型在训练过程中动态计算需要存储的最少梯度数据，以此减少训练所需要的VRAM空景。使用Configuration file加载功能加载该文件直接设置训练参数成目标。

在完成上述步骤的准备工作后，就可以正式开始模型的二次训练了，在作者的个人桌面电脑上、该训练过程约会持续15分钟至1小时。期间我们可以使用tensorboard观察模型loss的收敛情况。

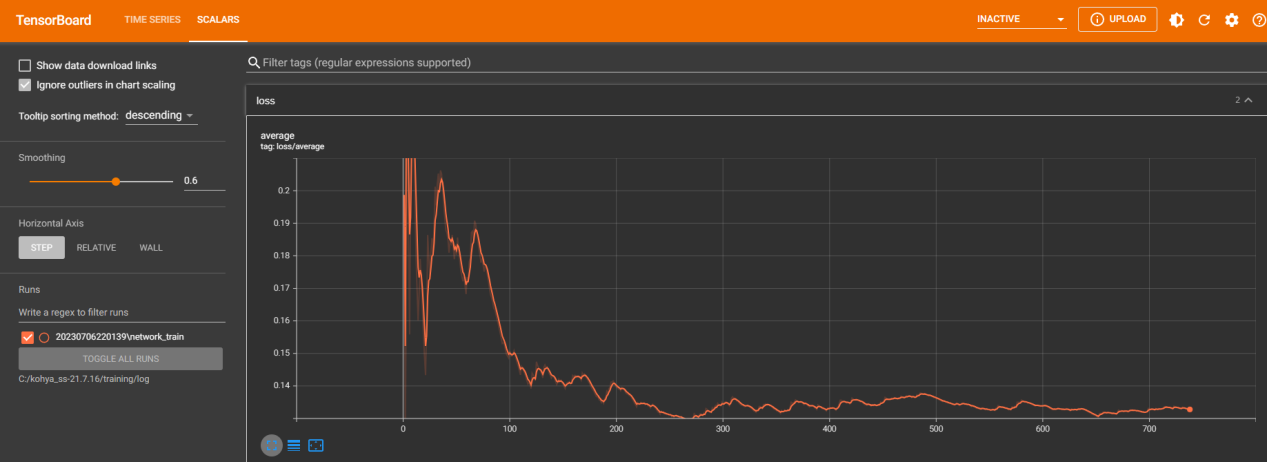


图 5 通过tensorboard观察模型loss变化

代码后台对作为素材的图片通过公式进行了解析，图6中所示为根据后台对素材图片的长宽和文字的长度等基础信息的统计所绘制的图片。

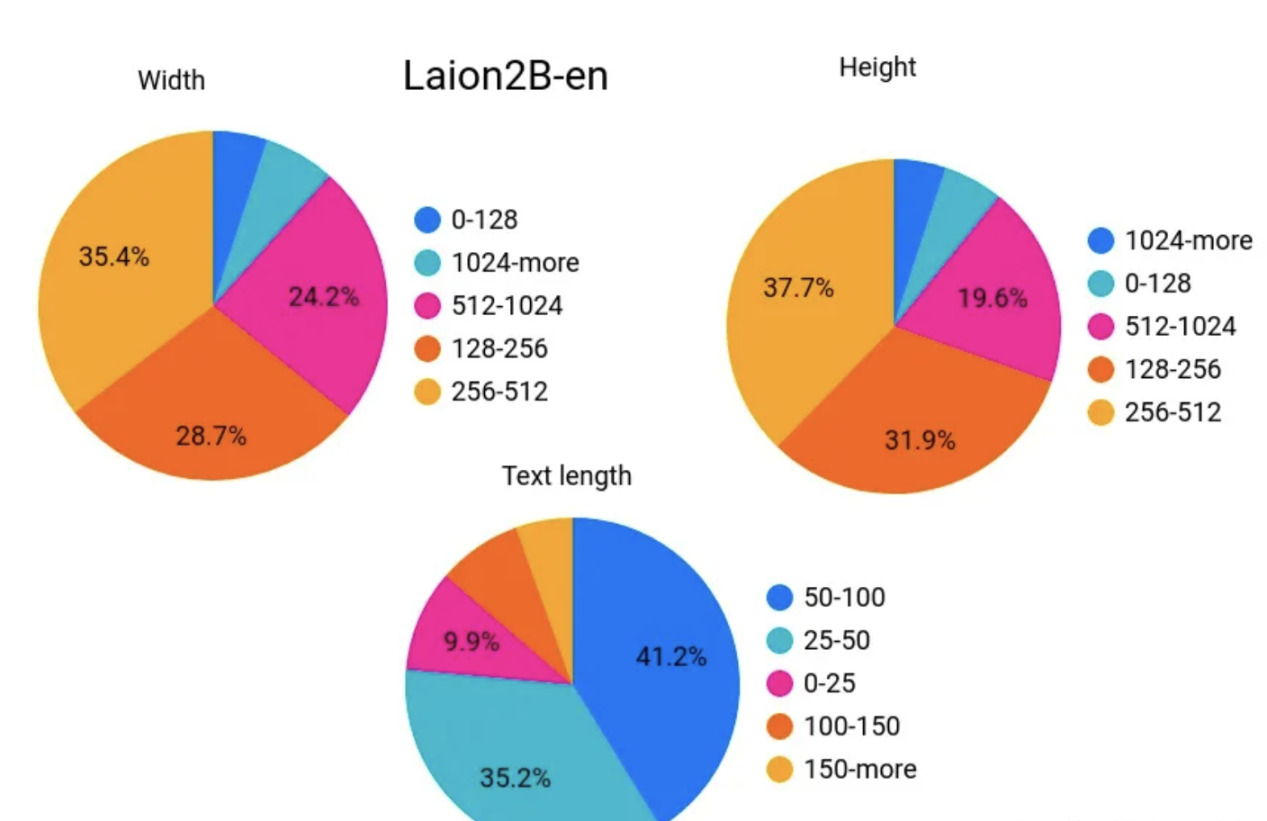


图6 样本数据

## 4.二次训练模型效果检验

在完成训练后，我们会得到一个约150MB大小的模型文件Addams.safetensors，这就是记录了LoRA算法中两类向量和参数的文件。将该文件放置于Stable Diffusion Web GUI下的目录后，我们就可以使用这个二次训练模型了。

还是使用第3节中用到的Prompt：”masterpiece, best quality, cat, looking at viewer, full body, standing, outdoors, flower field, landscape, slit pupil“对模型进行测试，得到的结果如图7所示。可以看到新的模型生成的图像稳定具有与训练集中的宠物猫相似的风格、同样的地板等。相比图3原始形态毛色各异的宠物猫，本次生成图片中的猫变成了确切指定的那只猫，从花纹到体型更加相似。并添加了花朵和阳光这两个变量，并且很好地融入了图片中。同时完成了选择指定猫和加入相关元素两个条件，以满足宠物主人内心对猫咪的期许，得到人文上的关怀。

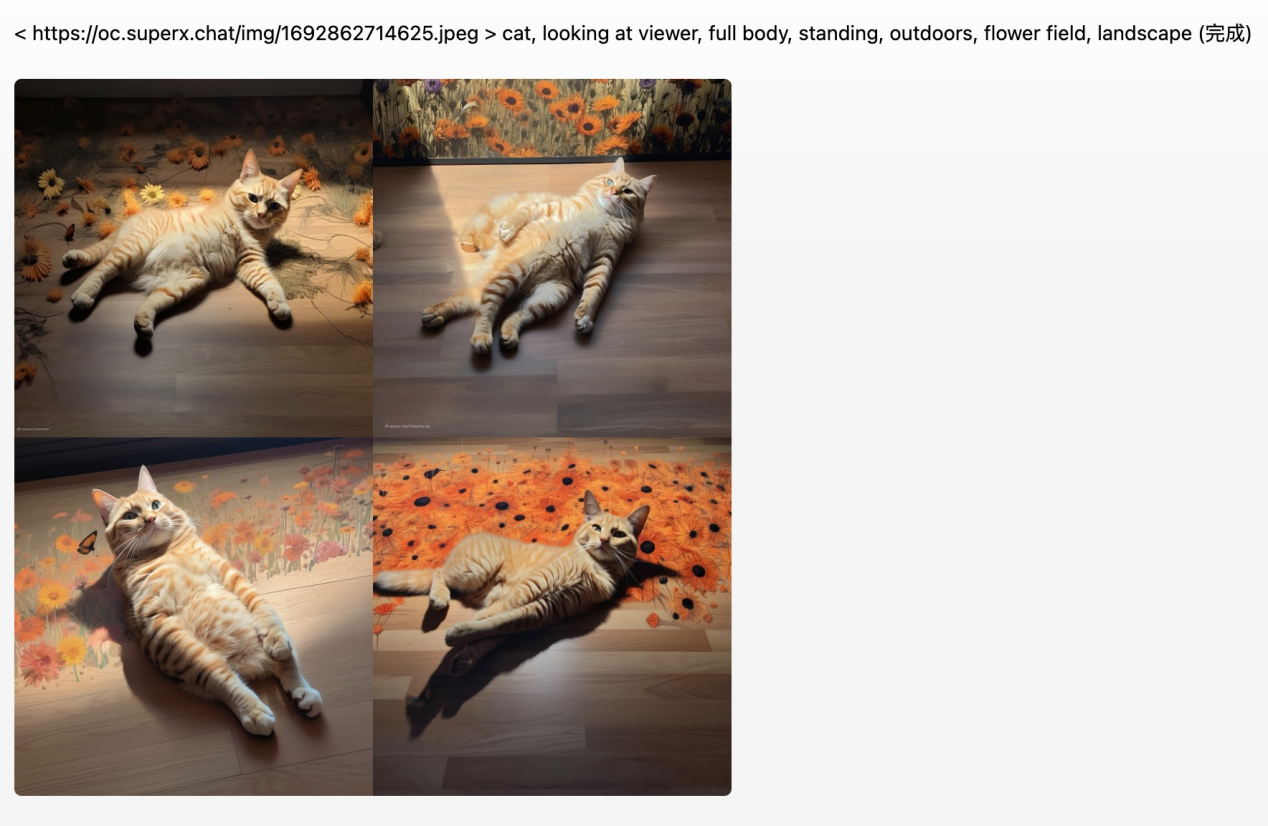


图 7 二次训练后的模型使用相同的 Prompt 生成的图像

## 5.总结与思考

虽然LoRA算法原是为大语言模型微调而设计的算法，但我们将其与图像生成领域的SD模型结合，仅使用32张个人图像构造一个训练集，在个人电脑上（仅有6G VRAM的图形卡）就成功产生了二次微调模型，使其能够稳定生成符合特定风格和种类的图像。在当前智能手机和个人电脑普及的年代，该方法让AI技术真正走入了百姓家，为人文需求服务。例如可以通过像文中我所做的一样通过训练获得可以通过文字标签直接修改动作和背景的自家宠物的图片，或许可以帮助到失去宠物的家庭抚平最后的遗憾。AI技术不仅仅是科学上的进步，也能弥补技术本身无法理解的人类情感。

下一步我将研究更多模型和技术，尝试生成个人特定的视频，生成可以与人互动的虚拟宠物（但和自己的宠物一个外形），甚至尝试让AI识别人声音中的情绪（喜怒哀乐）、从而让虚拟宠物自动产生与情绪对应的动作或行为。