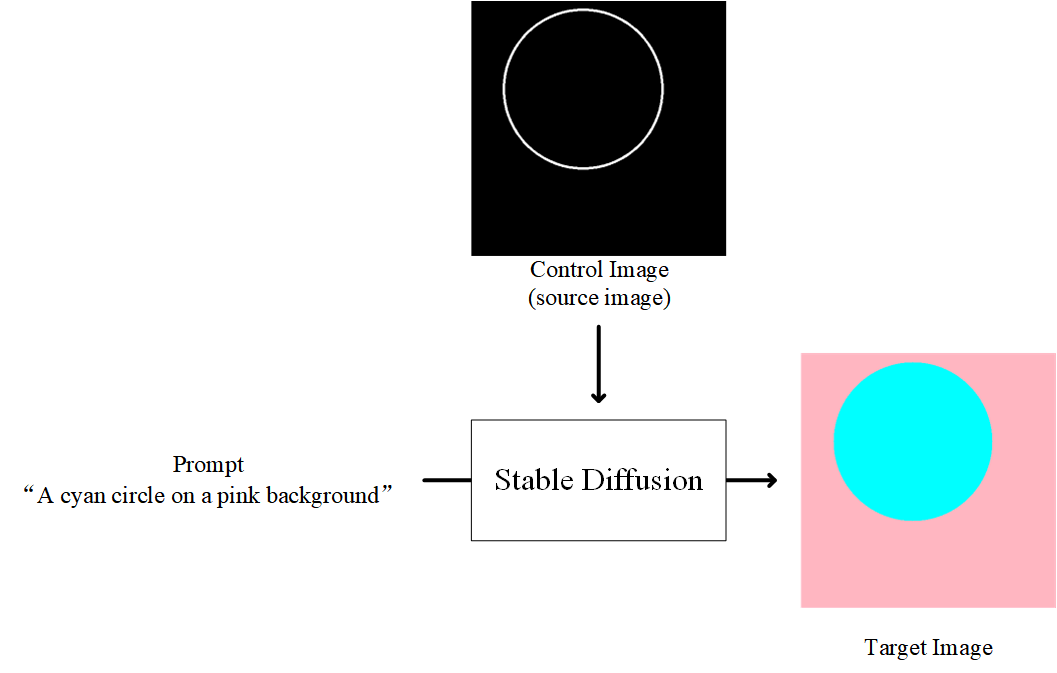
**3. 功能测试**

3.2 实验测试

我们小组一共进行了3次训练实验测试。每次训练我们采用相同的训练参数，训练的batch\_size = 4，learning\_rate = 10^-5，Stable Diffusion 模型部分锁定，训练ControlNet部分。训练流程如下：

我们想控制Stable Diffusion模型用颜色填充circle，prompt包含我们target图片的文字描述。Stable Diffusion模型是已经预训练好的，它已经知道提示文字的含义，但不知道“控制图像”的含义。我们的目标就是让他知道控制条件的含义，然后生成期望的图片。

1. 第一次训练实验：

我们使用作者提供的数据集fill50k进行训练。数据集的结构包括：控制图 source/ 文件夹、原图 target/ 文件夹、提示词 prompt.json 文件。这个数据集不是专门针对线稿上色模型的，因此对线稿上色效果较差。此次实验的目的一是摸清实验整体流程，二是了解从哪些方面进行评估分析。我们训练了3个epoch，根据训练日志得到的曲线走势图如下：

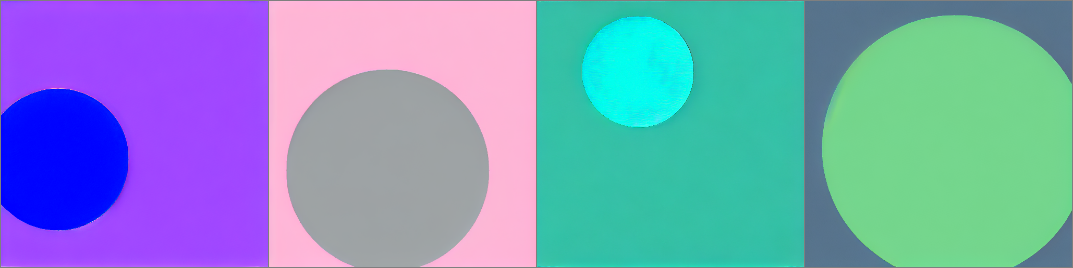
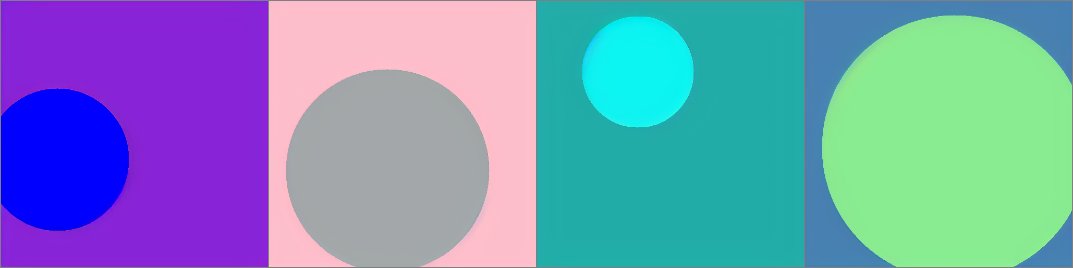
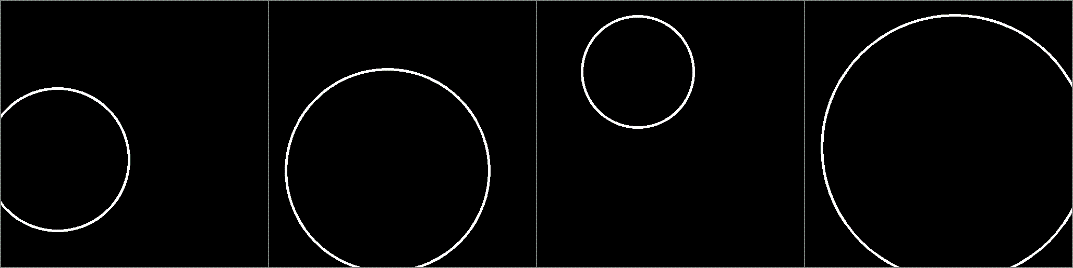
图表, 折线图

描述已自动生成

图表

描述已自动生成

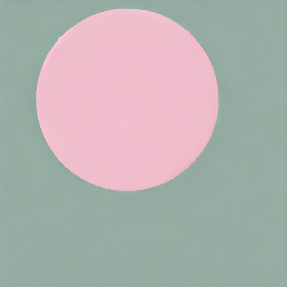
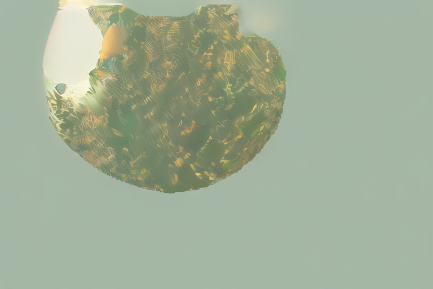
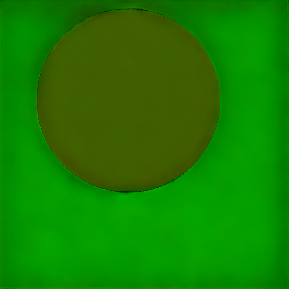
整体上loss成下降趋势，说明真实图片和预测图片的误差在减小，模型的性能逐渐向好。我们随机选取了训练过程中模型对训练集图像的预测图片，结果如下：



第一排是控制图像，第二排是真实图像，第三排是训练模型的预测图像。可以发现预测图像和真实图像的误差不算大，模型对circle的边缘信息理解较好，对图像填充的颜色信息理解稍差。可能由于训练集都是circle的原因，测试时选用非circle效果并不理想，于是我们选择训练集中的图进行测试。测试结果如下：

形状, 圆圈

描述已自动生成



第一张图是输入的控制图，第二张图在没有提示词条件下的生成图像，后两张图是在有提示词条件下的生成图像。可以看到模型对控制图的信息理解较好，但是在没有提示词条件和有提示词条件时相比，对图像边缘信息的理解较差，我们认为在迭代次数较少的情况下，模型还是比较依赖prompt描述才能有较好的生成效果。

1. 第二次训练实验：

我们在danbooru数据集基础上建立自己的数据集。我们从中随机选取50k对原图和线稿图片对，通过提示词反推工具对原图生成提示词，编写python脚本将提示词进行合成和结构化，并仿照作者给出的fill50k数据集的结构，建立自己的数据集。此次实验我们训练了11个epoch，得到的曲线走势图如下：

图表

描述已自动生成

图表, 折线图

描述已自动生成

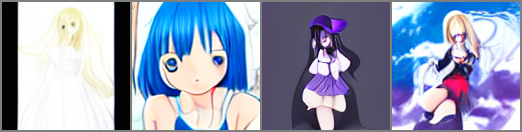
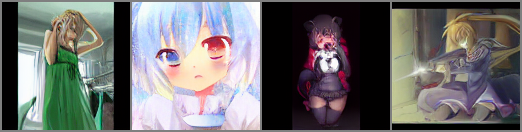
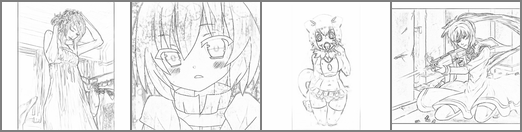
图表

描述已自动生成

图表

描述已自动生成

其中一个指标是loss\_vlb\_epoch，它反映的是扩散模型的变分推断中近似分布和真实分布的差异。vlb是变分下界，用于计算近似分布。训练过程中模型对训练样本的预测如下：



从上到下分别是线稿图、真实图和预测图。可以看到模型对人物边缘信息理解还是较好的，但是细节部分还是不容乐观，譬如人物的眼睛、脸、发型等，并且人物的颜色上色差距很大。之后测试的结果也不理想。我们将这个数据集和fill50k数据集进行对比后，发现我们的数据集控制图是白底黑线，而fill50k数据集是黑底白线，我们了解到计算机是将黑像素值为0，白像素值为255，把黑色当作mask掩膜也就是不需要处理的部分，因此我们决定将线稿图进行反色处理后，作为控制图进行训练。

1. 第三次训练实验：

此次训练实验中，我们将输入图像进行反色，并随机选取更少的10k对图片对，用更多的迭代次数进行训练。最终共迭代了30个epoch，得到的曲线走势图如下：

图表, 折线图

描述已自动生成

图表

描述已自动生成

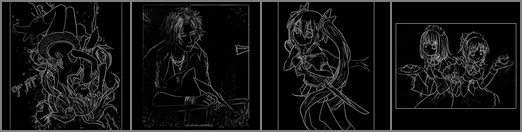
图表, 直方图

描述已自动生成

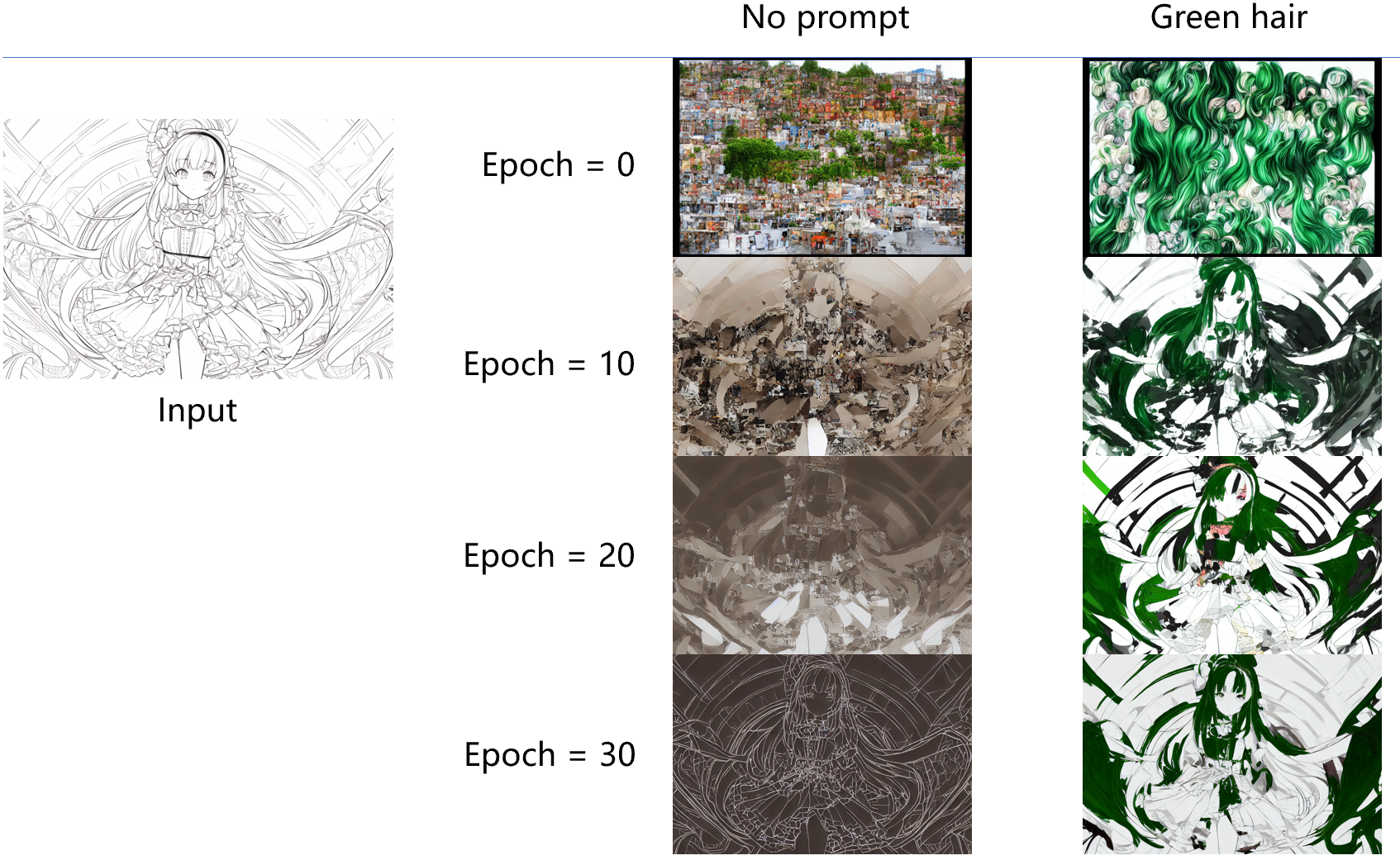
图表, 直方图

描述已自动生成

四张曲线图的整体趋势都是下降的。而在第二张图中，可以发现loss\_vlb\_epoch的值发生了抖动现象。我们小组讨论研究后认为，出现抖动情况可能是因为每个迭代步骤只利用了一个小批量数据，在计算变分下界损失函数时，需要对整个训练集进行求和，计算出训练集的损失函数。如果样本数据较少，那么每次迭代计算得到的损失函数受到该少量数据中的噪声等因素的影响就会相对更大，从而导致损失函数值抖动，并且可能会导致过拟合现象的发生。解决抖动现象就需要更大批量的数据，或者使用更优的算法。以下是随机选取的训练过程中模型的预测图像：



可以看到，此次训练实验与第二次训练实验相比，对图像的边缘信息理解更好，误差更小。但仍然存在一个问题，也就是预测图像的颜色问题。我们发现在自己建立的数据集中，prompt对图像的内容信息有较好的描述，但是颜色信息却存在较大误差甚至是缺失。我们尝试解决该问题，但解决方案主要还是人工筛选优化，对于样本较多的数据集来说，这项工作非常繁杂。该模型的测试结果如下：



我们用非训练集中的图像，在epoch = 0，10，20，30及分别有提示词和无提示词条件下进行了测试。可以看到模型生成结果整体趋势是向好的，ControlNet的作用部分相对来说性能在逐渐提升。无提示词条件下，从最初的毫无关系的图像，逐渐生成边缘较清晰的图片，但是在对图像上色的颜色单一，这可能是由于模型训练不足，导致对prompt的依赖，如果没有prompt提示，就只会对整张图片填充一种颜色；有提示词条件下，ControlNet的作用部分对人物头发的理解逐渐提升，从最初的大范围上色，到最后的上色范围缩小，但缩小后，部分本应该上色的地方没有进行上色。同时，由于对prompt的依赖，当只输入一个prompt时，其他部位几乎没有进行颜色的填充。

**4. 问题分析**

在本文中的几个训练实验测试中，我们发现几个模型的训练与测试都有一定的效果，但是效果并不是很好，在对图像的特征信息的理解上还存在着明显的偏差，譬如将人物背景识别成人物头发、将人物颈部的装饰物识别为眼睛等，并且在对线稿进行上色时可能出现颜色单调或者没有成功上色的情况。我们小组对几个训练实验测试进行了思考和总结，认为在以下方面存在不足和可以改进的地方：

1. 训练过程中迭代的次数以及训练的步数。当训练的迭代次数越多，模型越能够充分学习训练集中的模式和特征，提高模型的准确性。但是当迭代次数过多时，会出现过拟合的现象，导致在训练集上的效果好，在测试集上效果变差。因此我们尝试根据训练过程中loss\_epoch的变化，以及在每进行n个左右的epoch后，保存checkpoint并检查当前的模型性能是否有所提升即上色效果变好，如果没有提升，则终止训练。
2. 训练集中的图像质量。图像的分辨率低会导致模型无法正确识别图像特征，图像的噪声过多会导致模型的准确性受到影响，而样本图像之间的差异过大，也可能导致模型过度学习这些样本图像，从而导致在新的测试集的表现不佳的过拟合现象。fill50k数据集图像与我们的数据集图像相比，图像分辨率更高，并且样本图像之间的差异较小，因此在相同的训练参数下，理论上前者的训练效果较好。较好的解决方案就是搜集分辨率更好样本差异较小的数据集进行训练。
3. 训练集中的提示词质量。与fill50k数据集中的提示词相比，我们自己建立的数据集的提示词存在问题主要是在对图像的描述的准确度方面。我们的提示词是通过当前已存在的提示词反推模型推导而来，因此在提示词的准确度方面欠佳，譬如没有对颜色进行正确识别、对图像一些细节存在识别错误等。解决该问题的方法主要还是需要通过人工筛选优化。之后我们可以尝试使用更小的训练集，人工对提示词进行筛选优化后进行更多的epoch训练，对训练后的模型进行对比和分析，判断提示词的优劣对模型的影响的强弱。

收获与体会

在过去的两个多月中，我们小组有幸参与到一项论文复现课题《基于ControlNet引导的线稿上色》的研究中。我们小组的成员都对动漫有一定了解和兴趣，便希望借此机会学习当前领域的前沿知识和技术，积累相关学习经验。

我们小组成员都是初次接触与科研有关的论文复现课题，因此在进行课题研究的过程中，我们一直学习如何进行科研相关工作，如查找相关文献等；同时组内经常进行探讨和交流，分享自己的学习成果，最终完成了论文复现课题。这为我们今后的研究生阶段的学习奠定了经验基础。

我们从兴趣出发，探索目前AI绘画领域的前沿知识和技术，比如Attention机制、Text-to-Image、扩散模型等。我们从ControlNet论文出发，了解了构建ControlNet的动机：为了在有限资源下也能产生较好的生成效果，对预训练好的大模型进行微调；理解了ControlNet的实现原理，通过复制Stable Diffusion的网络结构，锁定原来的网络以保留原网络能力，训练复制网络来进行性能优化等。在此基础上，我们拓展学习了相关工作的论文，如Attention机制，CLIP文图匹配，Fine-Tuning的其他模型等，并与ControlNet进行联系或对比，最后总结出小组关于上色任务的见解。

在进行论文复现课题的过程中，我们遇到了一些困难和挑战，比如本地设备性能不足、运行环境的配置和搭建、python包的版本的不一致导致各种报错、模型的训练效果不理想、论文中相关概念不理解或缺少相关资料等。我们对出现的问题进行网络检索或是请教学姐，不断探索和尝试，最终一一解决了这些问题。从中我们获得了一些与工程相关的技术能力以及对论文的阅读理解能力，希望在将来我们能够将这些能力运用到实践中去。在解决这些困难和挑战的过程中，我们还通过定期汇报和组内交流探讨等形式，提高了团队合作和沟通交流的能力，明白了科研工作不仅仅是依靠个人的努力，它更需要团队的协作、集多人的智慧，每个人的专业知识和独特视角都是团队研究成功的关键。

最后我们感谢老师和学姐为我们提供指引和帮助，使我们在这次课题中收获良多。