

# 实验五(Text2Image)

## 实验要求

使用 pytorch 或者 tensorflow 的相关神经网络库编写 text2image 代码，并在相应的数据集上实验。本次 text2image 实验主要使用两种方法，一种是基于 Conditional GAN 的方法，另一种是基于 Conditional Diffusion Model 的方法。基于 Conditional GAN 的方法需要自己训练，基于 Conditional Diffusion Model 的方法则可以使用训练好的模型。基于 Conditional GAN 的方法可以参考论文 [RAT](#) 和其开源代码库 [RAT-GAN]([GitHub - senmaoy/RAT-GAN: A conditional GAN for text-to-image](#))。基于 Conditional Diffusion Model 的方法可以参考 [StableDiffusion]([Stability-AI/stablediffusion: High-Resolution Image Synthesis with Latent Diffusion Models \(github.com\)](#))，可以直接调用其训练好的模型。本次实验以动手实践为主，报告主要记录实践过程和关键代码以及实验结果。

## 实验步骤

- 网络框架：**要求选择 pytorch 或 tensorflow 其中之一，依据官方网站的指引安装包。并且依据参考的内容安装相应的库。
- 数据准备：**本次实验使用的数据包含一个常用的简单的 text2image 数据集
  - Oxford-102：**该数据集包含8189张102种花的图片。每张图片包括10种文字描述。可以从 [text\\_flower](#) 中下载文字描述，从 [flower](#) 中下载图片。使用其中82种类别当作训练集，剩下 20 种类别当作测试集。
- 数据预处理：**你需要通过pytorch 或 tensorflow所提供的标准数据接口，将原始数据处理为方便模型训练脚本所使用的数据结构，如 torch.utils.data.Dataset 等。
- 模型训练：**基于 Conditional GAN 的模型需要自己训练，因此需要绘制 loss 随着训练进行的图像。
- 模型测试：**对于测试集中的数据。使用自己训练好的模型(CGAN)生成图片，并展示部分实验结果（包括原始图片，原始文字描述以及生成后的图片）。同时还需使用开源的 StableDiffusion 模型生成图片，并展示实验结果。两个模型使用相同的文本描述。

## 实验提交

本次实验截止日期为 **7月1日 23:59:59**，需提交代码源文件及实验报告到邮箱：[proton00@mail.ustc.edu.cn](mailto:proton00@mail.ustc.edu.cn)，具体要求如下：

- 全部文件打包在一个压缩包内，压缩包命名为 学号- 姓名 - exp5.zip
- 代码仅包含 .py 文件，如果有多个文件，放在 src/ 文件夹内。对于自己训练好的模型权重，可以上传到睿课网的云盘，并在实验报告中提供共享的地址。

3. 代码中提供一个可以直接运行的 `main.py`，可以根据输入的文本，生成文本所描述的图片。
4. 代码中提供一个描述所有需依赖包的 `requirements.txt`，手动列入代码中用到的所有非标准库及版本或者使用 `pip freeze > requirements.txt` 命令生成。
5. 实验报告要求 pdf 格式，要求包含姓名、学号。内容包括**实验过程**和**关键代码**展示，训练集损失图以及生成的实验结果展示（挑出较好的结果）。

## 参考资料

---

实验数据下载：

[text\\_flower](#)

[flower](#)

参考论文：

[RAT](#)

[2112.10752.pdf \(arxiv.org\)](#)

参考代码：

[RAT-GAN]([GitHub - senmaoy/RAT-GAN: A conditional GAN for text-to-image](#))

[Stability-AI/stablediffusion: High-Resolution Image Synthesis with Latent Diffusion Models \(github.com\)](#)

[Stable unCLIP \(huggingface.co\)](#)