

一. 环境准备

在 autodl 上配置选择如下：

镜像 PyTorch 2.3.0 Python 3.12(ubuntu22.04) CUDA 12.1 [更换](#)

GPU vGPU-32GB(32GB) * 1 [升降配置](#)

CPU 16 vCPU AMD EPYC 9654 96-Core Processor

内存 60GB

硬盘 系统盘: 30 GB
数据盘: 免费:50GB SSD 付费:25GB [扩容](#) [缩容](#)

附加磁盘 无

端口映射 无

自定义服务

端口协议 6006端口: http 6008端口: http [修改](#)

网络 同一地区实例共享带宽

计费方式 按量计费

费用 ￥1.68/时 ￥1.77/时

二. 第一阶段物体去除模型的数据集

RORD: 真实世界物体移除数据集

完整数据可在以下网址获取：<https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=realm&dataSetSn=487>

低分辨率 (960X540): <https://drive.google.com/drive/folders/1-aCLHMWvb-8Hwv82v17wl4rVuDutwAv6?usp=sharing>

四十锁/RORD --- Forty-lock/RORD

我采用的是后者，但依然很大内存故每类选取前十张，总计约 3430 组数据

模型要求物体是白的，背景是黑的，这份数据集是反的需要反转处理一下；

-  **gt** 包含 gt 纯背景, img 带物体原图, mask 覆盖物体+阴影的掩码图
-  **img**
-  **label**
-  **mask**

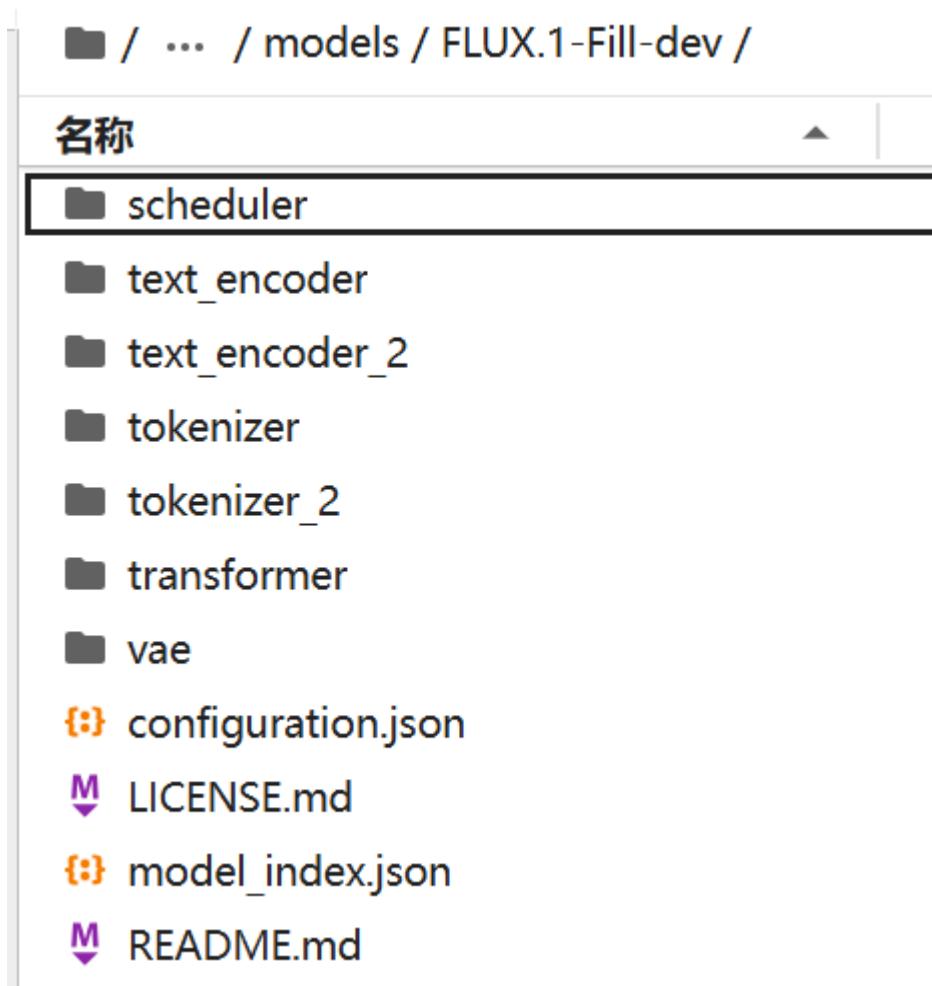
三. 物体去除模型训练

原文是全量微调，但 autodl 显存有限使用 **LoRA**

采用 FLUX 模型，git clone <https://github.com/lrz Jason/T2ITrainer.git>

本地下载模型并上传，载入训练代码

```
# 下载 FLUX.1-Fill-dev 到指定目录
model_dir = snapshot_download(
    'AI-ModelScope/FLUX.1-Fill-dev',
    cache_dir='/root/autodl-tmp/models',
    revision='master'
)
print(f"模型下载完成, 路径: {model_dir}")
```



将数据处理成索引 json(/root/autodl-tmp/prepare_data.py),
简单修改/root/autodl-tmp/T2ITrainer/train_flux_lora_ui_with_mask.py 的数据读取逻辑。
(以及其余几处修改)

之后配置 config_fill.json

- 输入: 图片 A (有物体) + Mask
- 目标: 图片 B (无物体, 背景图)

cd /root/autodl-tmp/T2ITrainer

accelerate launch train_flux_lora_ui_with_mask.py --config_path config_fill.json

云平台显存有限，训练过于漫长，采用预训练模型观察效果(autodl-tmp/T2ITrainer/A_MY_ADD/test_pretrained.py)



四 . 下一阶段物体插入模型数据集准备

路线为搜集海量网络图片扣除物体+阴影，生成纯背景图片，再把物体(不带阴影)图片贴到背景上，得到“悬浮感”这种无阴影的不太自然的图片，在这里利用现成的 gt 纯背景图片 /root/autodl-tmp/prepare_real_insertion.py 脚本自动化处理，利用 rembg 智能抠图



五 . 物体插入模型的训练

修改配置文件: /root/autodl-tmp/T2ITrainer/stage2_train.json

训练脚本: root/autodl-tmp/T2ITrainer/train_stage2.py

```
█ / ... / flux_real_insertion / flux_insertion_lora_stage2-3-3432 /
```

名称	已修改
flux_insertion_lora_stage2-3-3432.safe...	前天
optimizer.bin	前天
pytorch_lora_weights.safetensors	前天
random_states_0.pkl	前天
scheduler.bin	前天

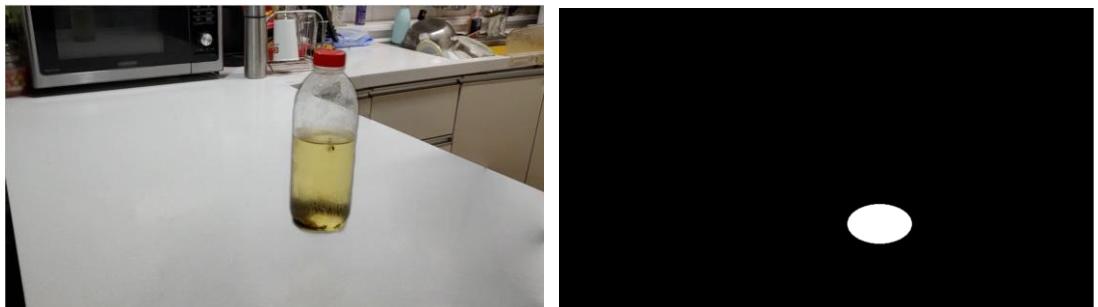
训练过于漫长(4epoch 要五个多小时)且效果不是很好，训练脚本需要进一步改进

暂时的通过预训练模型的观察其效果

/root/ autodl-tmp/T2ITrainer/A_MY_ADD/ add_shadow_local.py

python add_shadow_local.py \

```
--model_path "/root/autodl-tmp/models/FLUX.1-Fill-dev" \
--image /root/autodl-tmp/dataset/l-210715_l09026_W02_F0008.jpg \
--mask /root/autodl-tmp/dataset/l-210715_l09026_W02_F0008_M.png \
--output result.png \
--prompt "soft natural shadow"
```



总的来说，物体插入模型的训练脚本还需修改，autodl 环境可能需要更换，然后训练更多 epoch 看能不能达到更好效果。