

大纲

1. 比赛背景简介
2. 当前思路
3. 后续优化点

背景简介

目标: text \rightarrow SVG 图像

评估:

vqa 分: 图像与给定文本的一致性

ocr 分: 图像中不能出现文字
(超过 4 个字符会被惩罚)

美学分: 生成的图像是否好看

当前思路

- ① text $\xrightarrow{\text{LLM}}$ SVG \times (#)
- ② text $\xrightarrow{\text{sd/flux}}$ Image $\xrightarrow{\text{转换}}$ SVG \checkmark (#)

技术点:

1° 更强大的“文生图”模型: $\text{sd2.1} \rightarrow \text{sdXL} \rightarrow \text{sd3.5 (mlc)}$
flux.1-dev

2° 迭代提升指标

* 指标计算 $\left\{ \begin{array}{l} \text{vqa} : \text{需要代理 (TODO)} \\ \text{ocr} : \checkmark \\ \text{aes} : \checkmark \end{array} \right.$

* 提升单次推理时间, 以便更多地尝试

$\left\{ \begin{array}{l} \text{图片尺寸} \\ \text{denoise steps} \\ \text{量化} \\ \text{减配} \\ \text{并行 (两张 GPU 并行推理)} \end{array} \right.$

* Image \rightarrow SVG 转换算法

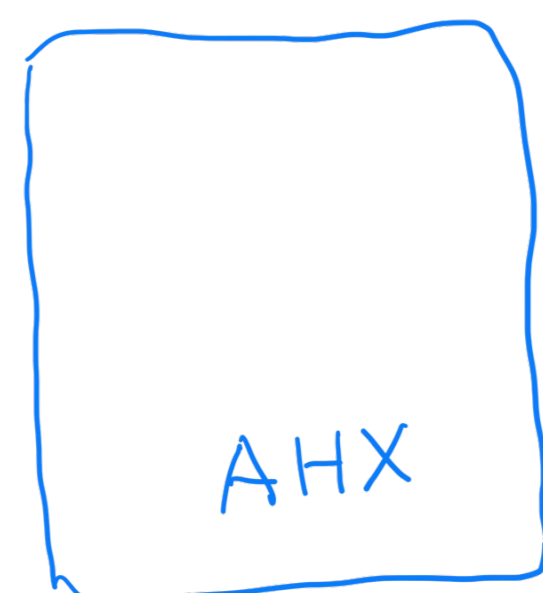
$\left\{ \begin{array}{l} \text{vtracer} \\ \text{基于 opencv 方案} \end{array} \right.$

* Lora $\left\{ \begin{array}{l} \text{sd} \checkmark \\ \text{Flux.1 (TODO)} \end{array} \right.$

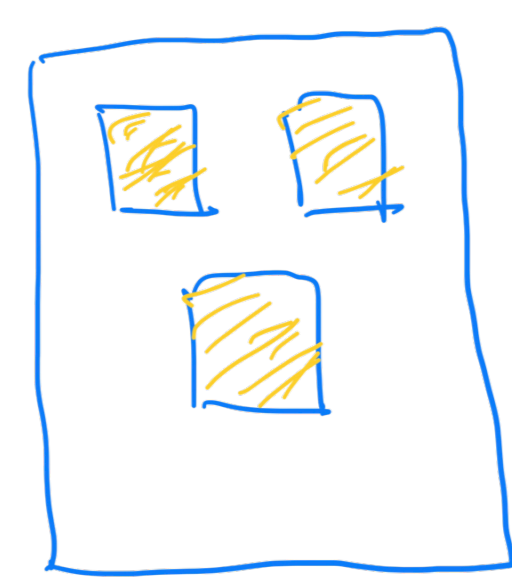
* prompt 调优: 调了 5 版, 效果不明显 (TODO)

* tricks

— 在生成的 SVG 中加入 4 个以内的字母 \checkmark



— 在生成的 SVG 中嵌入高美学分的小方块 (探索中)



aes: 0.5 \rightarrow 1.2

后续优化点

— 美学分优化

— prompt

— Lora

— trick 方式 嵌入

— vqa 分优化:

— 生成问题

— proxy

— 迭代优化

$\left\{ \begin{array}{l} \text{提速与提升单次图片质量 trade-off} \\ \text{双 GPU 并行 (2 程)} \end{array} \right.$

— 指标分析

$\left\{ \begin{array}{l} \text{日常类} \checkmark \\ \text{风景类} \checkmark \\ \text{抽象类} \times \end{array} \right.$

如何加 X 进来

1. 熟悉 文生图 简单原理
2. 熟悉 diffusers 库 API
3. 熟悉 Kaggle notebook 操作
4. 大致了解 图 \rightarrow SVG 原理
5. 看懂目前 top 方案 (Submit)
6. 从以上“后续优化点”入手