



四川大學
SICHUAN UNIVERSITY

基于生成式 AI 的个性化文创图像作品设计

成员：郑仕博 黄家曦 许圳达 叶果 陈奕嘉

答辩人：郑仕博

指导老师：赵启军

时间：2025.04.18

A background image showing a park scene with trees and a monument. The monument is a large, dark, rectangular structure with a lighter-colored panel in the center featuring a stylized, abstract design. The trees have yellow leaves, suggesting autumn. A red semi-transparent banner is overlaid on the image.

目录 CONTENT

01

项目创意来源

02

项目进展

03

项目展示与测试

04

相关成果以及下一步计划



PART
ONE



项目创意来源

项目创意来源



四川大學
SICHUAN UNIVERSITY



本作品的核心创意来源于当前市场上文创产品同质化严重，难以满足游客日益增长的个性化需求的痛点。习近平总书记关于推动文化和旅游融合发展，将文化旅游业培育成为支柱产业的指示，以及《如果国宝会说话》等成功案例，激发了通过创新方式“激活”文化遗产，赋能个体创造独特文创作品的想法。



青铜立人像



青铜神鸟鼠标垫



PART
TWO



项目进展

系统算法实现流程



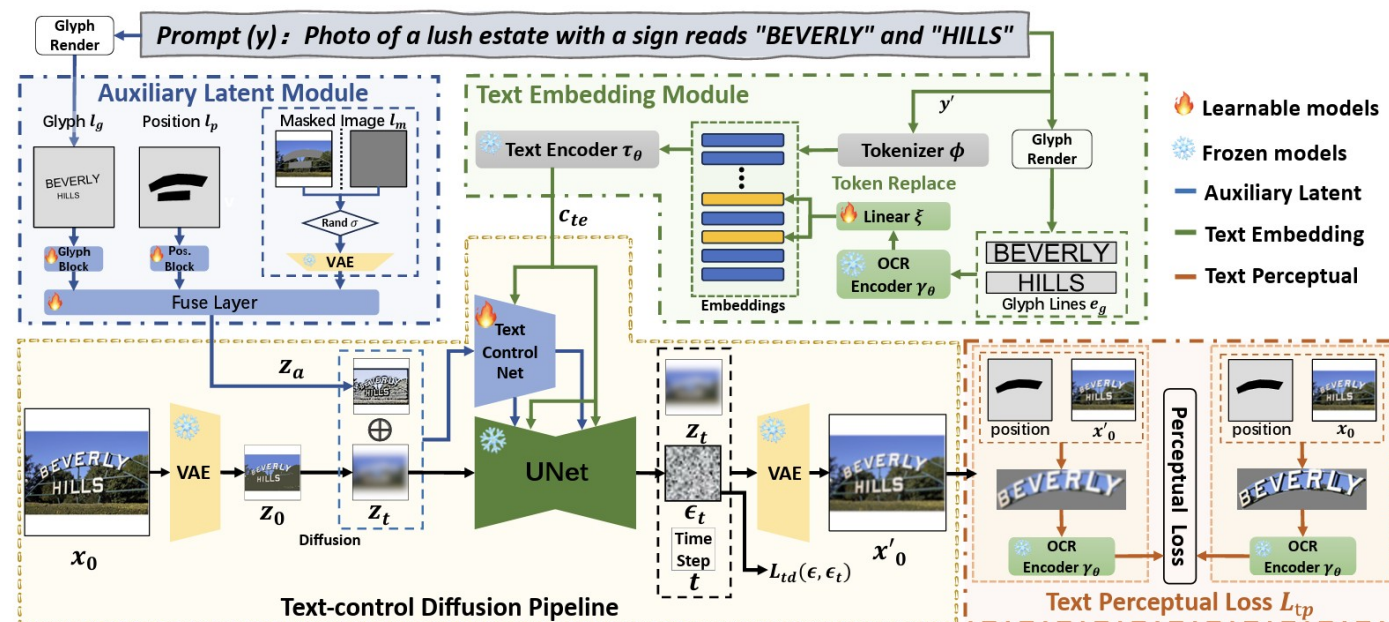
四川大学
SICHUAN UNIVERSITY

模型

使用阿里云的 *AnyText* 模型，主要由三部分组成——*Auxiliary Latent Module*, *Text Embedding Module* 和 *Text-control Diffusion Pipeline*

(*sd1.5* 扩散模型加上文字控制框架)

本项目将会对框架和扩散模型部分分开进行训练。



截自论文“ANYTEXT: MULTILINGUAL VISUAL TEXT GENERATION AND EDITING”

功能

文字生图：通过文字说明进行图片生成，可以指定文字生成的位置

文字编辑功能：对图片进行文字添加、修改和去除

特点：能够图文融合



图文融合示例

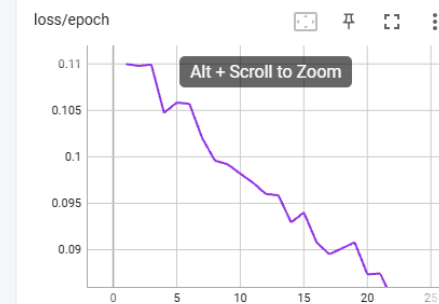
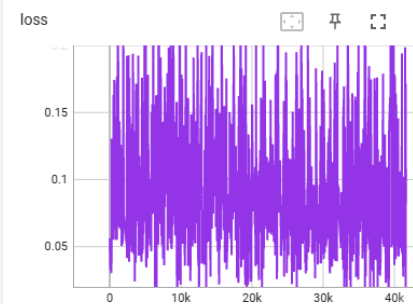
文字控制框架训练

对开源数据集 AnyWord-3M 进行筛选
(根据水印, 有效性, 同时保证中文训练
数据集远大于英文数据集), 保留了约
400k 的数据量, 并用 8 卡 V100 进行训
练。

扩散模型训练

在 google、edge、百度等搜索引擎上用 python
爬取图片, 经过筛选后, 对保留的约 1 千张的图片进
行去水印, 调整分辨率为 512*512, 用 *wd14-convnextv2-v2* 进行标注, 进行简要修改后, 从
hugging-face 上下载 Realistic_Vision_V4.0 作为起
始权重, 采取 DreamBooth 的方式进行权重的微调。

```
{
  "data_root": "/root/autodl-tmp/AnyText/data/data/imgs",
  "data_list": [
    {
      "img_name": "043517ab1c1560c0178a34e9108c33f467868286.jpg",
      "annotations": [
        {
          "polygon": [
            [
              72,
              99
            ],
            [
              430,
              97
            ],
            [
              431,
              210
            ],
            [
              73,
              212
            ]
          ],
          "text": "KDG",
          "language": "Latin",
          "rec_score": 0.9990164041519165,
          "valid": true
        },
        {
          "polygon": [
            [
              198,
              261
            ],
            [
              322,
              261
            ],
            [
              322,
              278
            ],
            [
              198,
              278
            ]
          ],
          "text": "KEDAGROUP",
          "language": "Latin",
          "rec_score": 0.9964058995246887,
          "valid": true
        }
      ]
    }
  ]
}
```





经费使用情况



四川大学
SICHUAN UNIVERSITY

目前经费仅用于租卡和服务器共计 1513.22 元

A	B	C	D	E	F	G	H
日期	交易类型	交易金额	优惠金额	余额支付	代金券抵扣		
2025-04-09	消费	-¥0.39	¥0.02	¥0.39	¥0.00		¥1,513.22
2025-04-08	消费	-¥7.97	¥0.41	¥7.97	¥0.00		
2025-04-07	消费	-¥0.78	¥0.00	¥0.78	¥0.00		
2025-04-06	消费	-¥22.89	¥1.21	¥22.89	¥0.00		
2025-04-05	消费	-¥16.86	¥0.89	¥16.86	¥0.00		
2025-04-04	消费	-¥25.09	¥1.31	¥25.09	¥0.00		
2025-04-01	消费	-¥2.11	¥0.11	¥2.11	¥0.00		
2025-03-31	消费	-¥46.14	¥2.43	¥46.14	¥0.00		
2025-03-30	消费	-¥49.92	¥2.64	¥49.92	¥0.00		
2025-03-29	消费	-¥41.77	¥2.15	¥41.77	¥0.00		
2025-03-28	消费	-¥43.01	¥2.23	¥43.01	¥0.00		
2025-03-27	消费	-¥10.08	¥0.53	¥10.08	¥0.00		
2025-03-25	消费	-¥0.11	¥0.03	¥0.11	¥0.00		
2025-03-24	消费	-¥0.11	¥0.03	¥0.11	¥0.00		
2025-03-23	消费	-¥0.11	¥0.03	¥0.11	¥0.00		

截取了一部分表格

后续经费安排:

1. 计算机设计大赛: 280 元
2. 软件著作权: 约 250 元
3. 租卡和服务器: 约 500 元



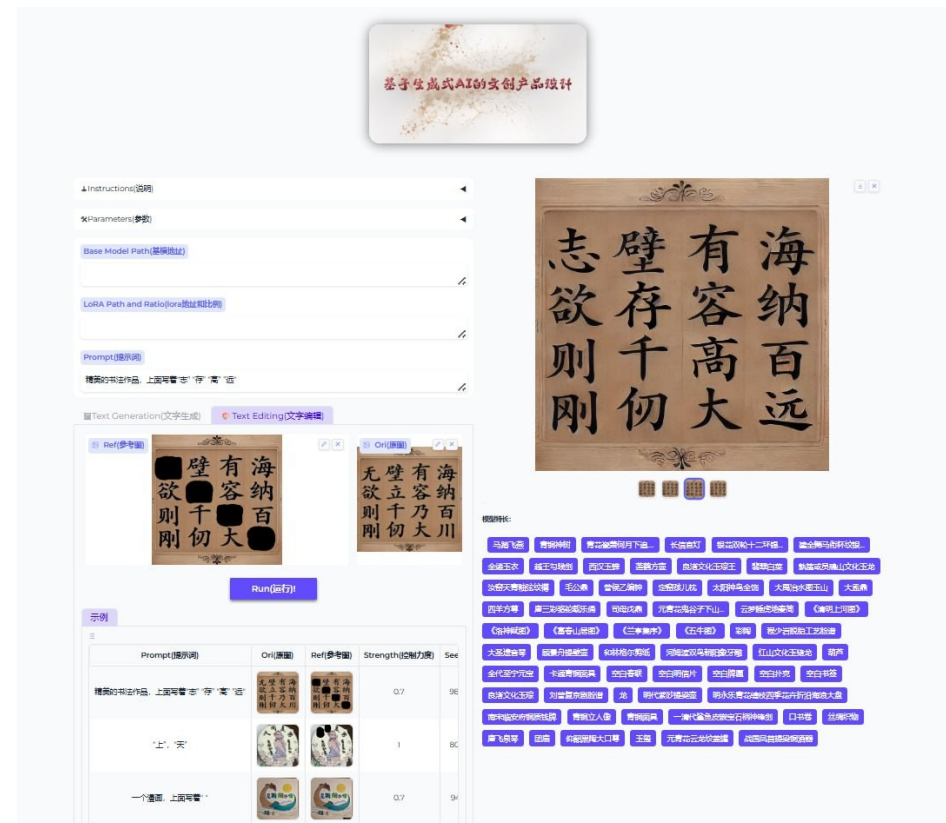


PART
THREE



项目展示与
测试

文字到图片到生成



图片文字编辑





生成对比

Ours

- 提示词：太阳神鸟金饰上写着“天路”

文字生成功能

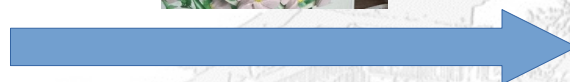


Ours

- 提示词：卡通青铜树，上方写着“神树”



文字编辑功能



- 提示词：“天”，“上”



更多例子





项目测试



四川大学
SICHUAN UNIVERSITY

	状态	正确率 ↑	编辑距离 ↑
文字生成	训练前	0.6957	0.8402
	训练后	0.6644	0.8282
文字修改	训练前	0.6671	0.8298
	训练后	0.6644	0.8282

文字控制框架文字生成和文字编辑评估正确率通过使用 OCR 对文字进行识别，与正确的进行比对，计算正确率（完全正确才行）和编辑距离。

发现训练效果并不理想，与算力、参数以及训练批次有关（经费问题）。

	FID ↓
训练前	31.558
训练后	34.242

sd1.5 扩散模型评估采用 *FID*（评估时与框架合并了）。证明了扩散模型经过训练，并没有遗忘掉太多先前学到的知识。



均采用 AnyText-benchmark!

因此目前项目使用官方 AnyText 的权重并与训练好的扩散模型进行合并，得到最终的模型权重



PART
Four



相关成果以
及下一步计
划



• 相关成果



四川大学
SICHUAN UNIVERSITY



四川大学
SICHUAN UNIVERSITY

Chengdu, 610065,
Sichuan, P.R.China
<http://www.scu.edu.cn>

相关成果

我们计划用本项目参加计算机设计大赛，目前作品已经完成，并录制了演示视频，准备报名并提交作品。

此外，我们也在积极推进本项目的软件著作权申报工作。

<input type="checkbox"/>	作品与答辩材料	-	2025-04-11 15:51
<input type="checkbox"/>	作品演示视频	-	2025-04-11 15:51
<input type="checkbox"/>	素材与源码	-	2025-04-11 15:51
<input type="checkbox"/>	设计与开发文档	-	2025-04-11 15:51
<input type="checkbox"/>	05-1人工智能应用作品提交要求（2024版）V2（人工智能实践赛、挑战赛...	15KB	2025-04-11 15:51

百度网盘链接：[计算机设计大赛作品](#)



计算机软件著作权登记申请

为原始取得或从他方继受取得的电脑软件、嵌入式软件、APP等软件申请著作权登记



• 下一步计划



四川大學
SICHUAN UNIVERSITY

为了进一步提升用户体验和创作自由度，本项目计划使用更好的模型。原有的 AnyText 模型准备替换为性能更优越的 AnyText2。这款新模型不仅在文字识别的准确率上实现了细微而显著的提升，更令人欣喜的是，它创新性地加入了选择字体的功能，让用户能够根据自己的创意和喜好，为文字赋予更丰富的视觉表达。

与此同时，为了激发用户的创作灵感，本项目计划在文字编辑板块补充一系列精美的图片素材，包括富有文化气息的书签、趣味十足的扑克牌以及充满个性化的明信片。这些丰富的视觉元素将为用户的文字创作提供更多元的选择和更广阔的想象空间，无论是制作独特的文创产品，还是设计个性化的节日贺卡，都能轻松实现。



谢谢各位老师
敬请批评指正

Thanks for your listening



四川大學
SICHUAN UNIVERSITY

答辩人：郑仕博

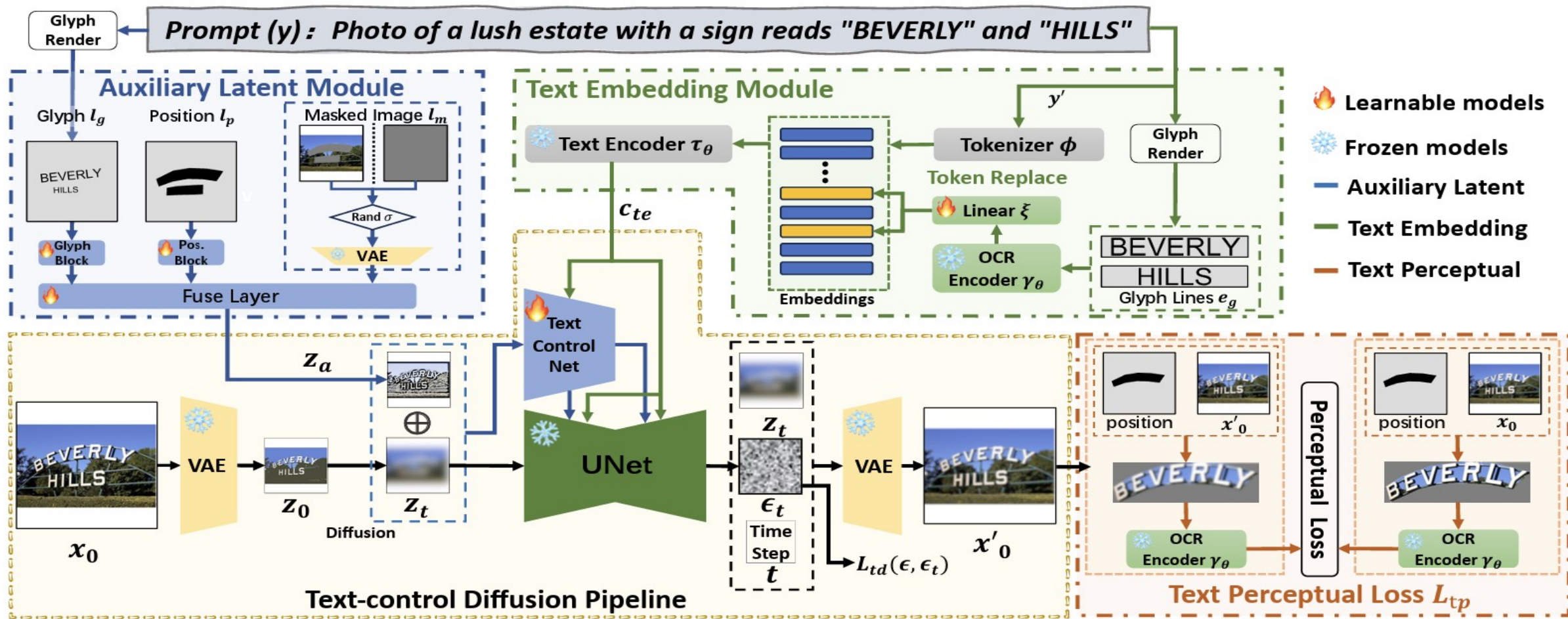
指导老师：赵启军

时间：2025.04.18

• 备用



四川大学
SICHUAN UNIVERSITY





• 备用



四川大學
SICHUAN UNIVERSITY

FID 的核心思想是利用预训练的 *Inception* 网络来提取真实图像和生成图像的特征，然后比较这两组特征在特征空间中的统计距离。

公式为：
$$FID = \|\mu_r - \mu_g\|^2 + \text{Tr}(\Sigma_r + \Sigma_g - 2 * \text{sqrtm}(\Sigma_r @ \Sigma_g))$$

编辑距离指一个字符串修改成另一个字符串所需要的最少操作，操作包括删除、添加和修改一个字符串。

DreamBooth 是由 *Google Research* 和波士顿大学于 2022 年提出的一种个性化微调技术，旨在增强文本到图像生成模型（如 *Stable Diffusion*）的能力，使其能够生成包含特定主体的图像。通过对预训练模型进行微调，*DreamBooth* 能够在保持模型原有能力的同时，学习并再现特定主体的视觉特征。

wd14-convnextv2-v2 是一种基于 *ConvNeXtV2* 的预训练模型，通常是用于计算机视觉任务中的深度学习模型。*ConvNeXt* 是一个基于卷积神经网络（CNN）的模型，专为图像分类和其他计算机视觉任务（如物体检测、分割等）设计。*ConvNeXtV2* 是 *ConvNeXt* 的改进版本，它采用了现代卷积网络架构中的一些新技术，例如与 *Vision Transformer (ViT)* 相结合的理念，使得其在性能上有所提升。



• 备用



四川大學
SICHUAN UNIVERSITY



首先通过gradio搭建并进入演示网站

百度网盘链接：[计算机设计大赛作品](#)

