四川大學

SICHUAN UNIVERSITY

基于生成式 AI 的个性化文创图像作品设计系统

项　目　总　结　报　告

任课教师 杨波

学　　院 计算机学院

专　　业 计算机科学与技术

组　　别 第一组

组　　长 郑仕博

成　　员 陈奕嘉，苏泳豪

2025 年 6 月 30 日

目录

1 绪论

1.1 项目背景

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。本项目旨在解决当前文创产品普遍采用预先设计制作模式，难以满足游客个性化需求的供需矛盾，探索相关技术在中文文创领域的应用潜力。此外，本项目还致力于为旅行者等用户提供便捷的工具，使其能够随时随地进行个性化文创设计和制作，满足市场需求。

1.2 相关技术现状

传统的图像编辑软件 (如 Adobe Photoshop, Illustrator)：这些软件功能强大，可以实现对图片的文字进行修改和添加。但它们通常需要专业技能，且操作相对复杂，难以满足普通用户快速、便捷地进行个性化文创设计的需求。此外，这些软件在生成与背景自然融合的文字方面也存在一定的局限性。

在线设计平台 (如 Canva, 稿定设计)：这些平台提供了丰富的模板和素材，用户可以进行简单的文字替换和排版。但其个性化定制程度相对较低，难以实现高度自由的创意表达。

已有的文字控制图像生成模型 (如 GlyphDraw, Textdiffuse, AnyText)：这些模型在解决字体与背景融合方面取得了一定的进展，但正如前文所述，它们仍然难以完全避免文字生成中的错误，并且缺乏专门针对中文的优化。当前模型通过集成大语言模型提升了文本生成的稳定性，然而，对文本生成位置的精细化控制以及基于图像内容的文本引导修改能力仍有待提升。

1.3 项目的主要开发工作（具体参见系统设计概述文档）

为解决上述问题，本项目的主要开发工作包括：

- **模型微调与优化**：采用 AnyText 模型作为基础，并针对中文应用场景进行专门的训练和优化。具体分为文字控制框架的训练和扩散模型的训练两部分。
- **数据集构建**：制作了两份数据集。一份是基于 AnyWord-3M 数据集筛选出的约 40 万条数据，用于微调 AnyText 模型。另一份是通过网络爬虫采集的约 1000 张与中华文化及文物相关的图片，用于微调 stable diffusion v1-5 模型。
- **系统实现与部署**：基于 Gradio 搭建了交互式的网页界面，实现了文字到图片生成和图片文字编辑两大核心功能。

1.4 项目成员、分工及完成情况

成员	主要分工	完成情况
郑仕博	• 项目统筹与规划 • 模型训练与调优 • 核心代码编写与上传 • 项目部署	• 负责修订项目计划书并明确团队分工 • 专注模型训练，研究并解决过拟合问题 • 与陈奕嘉共同进行扩散模型的训练 • 完成最终版模型的训练，并上传至 ModelScope • 编写模型测评文件代码并上传至 GitHub • 与组员共同完成前后端制作并上传至 GitHub • 使用 LaTeX 重新书写了系统概述和需求分析文档 • 参与编写 Dockerfile，将项目封装并上传至 Docker Hub • 将项目部署至 ModelScope • 完善文档，增加技术难点部分，并制作测试报告和 PPT

陈奕嘉	• 文档撰写与管理 (项目计划书、需求分析、软件著作权等) • 数据集搜集与整理 • 辅助模型训练与调节 • 软件测试	• 起草并完成项目计划书和需求分析书初稿 • 负责搜集与中华优秀传统文化主题相关的训练集 • 合并了两个模型的权重 • 与郑仕博一同完成模型部分的训练 • 撰写软件著作权说明书与项目注意事项的初稿 • 使用 LaTeX 修改系统设计文档 • 参与编写 Dockerfile 并封装项目 • 参与将项目部署至 ModelScope • 筹划并执行软件测试 • 参与制作最终汇报 PPT
苏泳豪	• 前后端框架开发与实现 • 数据集搜集与标注 • 辅助模型调试 • 项目部署支持	• 寻找并修改了前后端模板，确定基本框架 • 负责收集并整理所有成员找到的数据集 • 编写脚本对标签进行修改，并标注数据集 • 完成前后端代码的书写，实现了文本生成和图像编辑两大核心功能 • 使用 LaTeX 重新排版需求分析文档 • 参与编写 Dockerfile，解决网络问题，成功将项目封装并上传至 Docker Hub • 编写了 requirements.txt 文件，并参与将项目部署至 ModelScope • 对项目产品进行测试，并参与制作汇报 PPT 和完善文档

2 项目策划、分析与设计工作情况

2.1 项目立项计划及完成情况

项目旨在解决文创产品个性化不足以及现有技术在中文处理上的缺陷。项目计划通过微调先进的生成式 AI 模型，开发一个能够让普通用户轻松设计个性化文创图像的工具。目前，项目已经完成了模型的训练、功能实现、系统部署，并申请了软件著作权。

2.2 系统需求分析工作和结果

2.2.1 功能需求

系统主要提供两种创作方式：

- **基于用户输入的文字**：用户输入文字内容，在预览区指定文字位置和大小，生成创意图片。
- **基于用户输入的成形图片**：用户上传图片，在图上标注文字位置并输入内容，系统生成最终的文创图片。

两种方式都支持结果的预览、调整、保存和分享，具体可见使用手册。

2.2.2 性能需求

模型需要具备较高的文本准确性、卓越的图像生成能力，并且能够避免过拟合，保证良好的泛化性能。此外，该系统对硬件有一定的要求，如下：

类别	基本要求
服务器端	Intel Core i5（或更新）； 内存 32G 以上； GPU NVIDIA RTX 4060，内存 8G 以上； 硬盘剩余空间不低于 50G；
客户端	能运行现代网页浏览器即可

2.3 系统设计工作及结果

本系统基于阿里云开源的 AnyText 模型框架进行开发与扩展，整体架构由三个核心模块组成：辅助潜在模块（Auxiliary Latent Module）、文本嵌入模块（Text Embedding Module）以及文本控制扩散生成模块（Text-control Diffusion Pipeline）。在此基础上，我们对文字控制能力和图文融合性能进行定向增强，并完成了多项训练与优化工作。

2.3.1 技术框架

本系统使用 AnyText 模型，结合了 SD1.5 的图像扩散模型与文本控制机制，能够实现图像中文字的精确生成与编辑。模型架构如下 (图1)：

- **Auxiliary Latent Module**：处理输入的文字渲染特征，与图像潜变量融合，指导扩散过程。
- **Text Embedding Module**：将文字内容编码为向量特征，结合 OCR 与 Tokenizer 保持语义一致性。
- **Text-control Diffusion Pipeline**：基于 UNet 和控制网络实现文字引导的图像生成，并引入 Text Perceptual Loss 实现图文融合感知优化。

该架构支持文字生成图与图像中指定区域的文字编辑，包括添加、修改与删除，具有良好的图文融合能力。

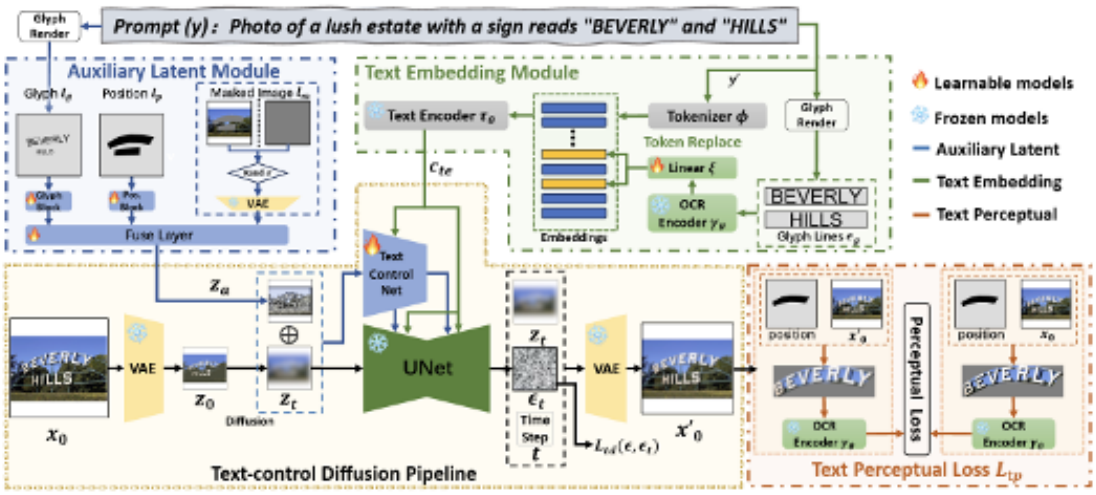

图 1: 模型示意图

2.3.2 界面设计

提供 Web 端界面，两种主要操作模式：

1. 文到图像的生成：输入提示词，并将需要渲染的文本用“”标注，可以通过画布绘制文本位置、拖框选择文本位置或随机选择文本位置；
2. 图片文字编辑，手动掩盖需要修改区域，输入文本并进行编辑。

界面上有说明、参数设置、文本输入框、模式选择、文字位置标注、样例（点击即可）、运行按钮、图片结果展示和加强训练的物品，用户可调整 CFG-Scale、Steps 等参数，查看结果并保存。详情见图 2，图 3.

详情见系统设计概述文档，用户使用手册。

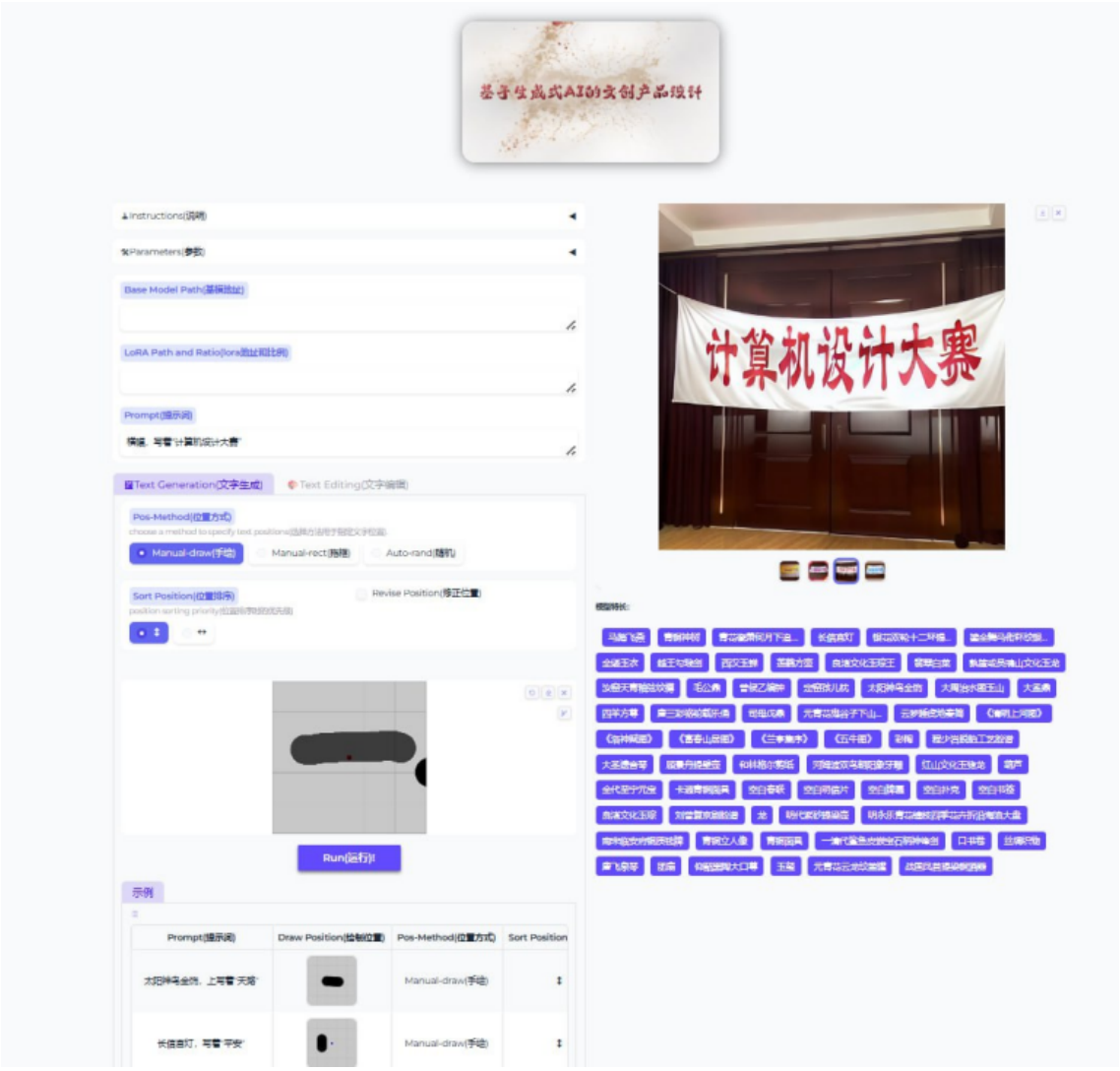
基于生成式AI的文创产品设计
Instructions(说明)
Parameters(参数)
Base Model Path(基模地址)
LoRA Path and Ratio(lora地址和比例)
Prompt(提示词)
Text Generation(文字生成)
Text Editing(文字编辑)
Pos-Method(位置方式)
Manual-draw(手绘)
Manual-rect(拖拽)
Auto-rand(随机)
Sort Position(位置排序)
Revise Position(修正位置)
Run(运行)!
示例
Prompt(提示词)
Draw Position(绘制位置)
Pos-Method(位置方式)
Sort Position
Manual-draw(手绘)
Manual-draw(手绘)
计算机设计大赛

图 2: 人机界面


基于生成式AI的文创产品设计
Instructions(说明)
Parameters(参数)
Base Model Path(基模路径)
LoRA Path and Ratio(lora路径和比例)
Prompt(提示词)
Text Generation(文字生成)
Text Editing(文字编辑)
Ref(参考图)
Ori(原图)
海纳百川
有容乃大
壁立千仞
无欲则刚
Run(运行)!
示例
Prompt(提示词)
Ori(原图)
Ref(参考图)
Strength(控制力度)
海纳百远
有容高大
壁存千仞
志欲则刚

图 3: 人机界面

2.4 系统具体实现工作及结果

2.4.1 数据集制作

本项目制作了两份数据集，第一份数据集用于微调 AnyText 模型，另一份数据集用于微调 stable diffusion v1-5。

- **数据集来源**：本项目利用爬虫技术从 Google、Bing、百度等搜索引擎抓取数据，自动化访问网站并提取网页内容（如文本、图片、链接等），筛选并保存有价值的信息。项目重点聚焦于文物图像数据的采集，旨在抓取并整合图文结合的信息，以构建高质量的数据集。大规模多语言数据集 AnyWord-3M，于论文 AnyText 中提出。数据来源涵盖 Noah-Wukong、LAION-400M 及多个 OCR 任务数据集（如 ArT、COCO-Text、RCTW 等），覆盖街景、书籍封面、广告等多种文本场景。OCR 数据直接利用已有标注，其余图片经 PP-OCR 处理并由 BLIP-2 生成文本描述。经过严格筛选与后处理，最终获得 303 万余张图片，包含 900 万行文本及 2000 万余字符。

- **数据规模**第一份基于 AnyWord-3M，按照水印、文字块、语言（保留大部分中文数据集，少部分英文数据集）等标准精筛后，保留约 40 万条数据。第二份数据集由爬虫采集约 1000 张与中华文化及文物相关的图片，并通过水印筛选和修整优化质量。随后，使用 wd14-convnextv2-v2 进行自动标注，并对标注结果进行了适当调整，以提升准确性和适用性。

2.4.2 模型微调

针对中文场景，本项目将对 AnyText 模型先与 Realistic_Vision_V4.0 （基于 sd1.5) 进行权重的合并，在中文训练集上的专门训练，使模型更好地适应中文应用场景，确保其在实际应用中展现卓越的表现能力。此外为了更贴合主题，本项目在 Realistic_Vision_V4.0 （基于 sd1.5) 权重基础上使用 dreambooth 的方法进行微调，并与微调后的 AnyText 模型的权重进行整合。（第一步的 sd1.5 只是用于训练的，其权重并不会改变，而第二步才微调 sd1.5 然后进行权重的整合）第一步使用第一份数据集，第二部采取第二份数据集。

2.4.3 部署与框架实现

本项目将使用 Gradio 搭建模型的交互式部署框架。结合任务需求，设计并开发直观易用的网页和交互界面，便于用户体验模型的功能和效果，同时支持进一步的迭代优化。

3 系统部署与测试工作情况

3.1 系统部署与调试

系统通过 Gradio 成功部署，提供了一个用户友好的 web 界面，支持文字生成图片和图片编辑功能 (可以使用 Docker)。

3.2 系统测试

系统测试使用的是 AnyText Benchmark，包含文字生成和修改两部分。测试结果表明，模型在文字生成和编辑方面表现良好，但在文字控制框架的训练过程中，由于算力限制，训练效果未达到预期。

- **文字正确率测试**：采用 OCR 识别与标注对比的方式进行评估。训练后的模型在文字生成和修改的正确率及编辑距离上没有显著提升，分析认为与算力、参数及训练批次有关。因此，项目最终采用了官方的 AnyText 权重。

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

表 2: 文字控制框架评估结果

- **扩散模型评估**：采用 FID (Fréchet Inception Distance) 进行评估。训练前后的 FID 值接近，证明模型没有发生严重的过拟合。

状态	FID ↓
训练前	31.558
训练后	34.242

表 3: 扩散模型评估结果

- **结论**：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

3.3 生成对比测试

见图4和图5。测试结果表明，模型在生成与背景自然融合的文字方面表现良好，能够满足用户个性化文创设计的需求。

图 4: 生成对比测试

图 5: 生成对比测试

4 项目管理相关工具使用情况

本项目在管理和协作过程中，充分利用了多种项目管理与协作工具，提升了团队的沟通效率和项目推进效率。具体如下：

- **代码与文档管理**：项目核心代码、模型测评文件、前后端代码、系统文档等均托管于 **GitHub**，实现了版本控制和团队协作开发。
- **模型与容器管理**：训练完成的模型上传至 **ModelScope**，便于模型的共享与部署；项目容器化后上传至 **Docker Hub**，实现跨平台部署和环境一致性。
- **文档与报告编写**：项目计划书、需求分析、系统设计等文档采用 **LaTeX** 进行协作撰写，保证了文档的规范性和可维护性。
- **任务分工与进度管理**：团队通过定期会议和在线协作工具明确分工，及时同步进展，确保各项任务有序推进。
- **测试与评估**：项目测试报告、PPT 等成果文档也通过 GitHub 进行版本管理和共享。

经费使用情况表格显示，项目主要开销用于租用云服务器和 GPU 资源，共计 1513.22 元。

5 项目讨论与体会

5.1 对项目过程的体会

- **作品特色与创新点**：项目实现了高度的个性化定制，操作便捷，生成的文字能与背景自然融合。其创新点在于通过创新的文字渲染模型，革新了图片修改的应用场景，并支持对图片中文字的便捷编辑。
- **挑战与不足**：在文字控制框架的训练过程中，由于算力限制（至少需要 8 卡 V100），训练效果未达到预期，这是一个主要的挑战。此外，该模型缺少对文字更多方面的控制，如字体、颜色等，且对中文的支持仍有待加强。

5.2 项目和项目管理过程的优点与不足

优点：

1. **高度个性化与创新性**：项目的核心创新点在于实现了高度的个性化定制，能够让用户轻松设计独特的文创图像作品。通过创新的文字渲染模型，它革新了图片修改的应用场景，并支持对图片中文字的便捷编辑，满足了当前市场对个性化文创产品的迫切需求。

2. **技术路线清晰且先进**：项目基于阿里云开源的 AnyText 模型框架进行开发与扩展，结合了 SD1.5 的图像扩散模型与文本控制机制，技术选型先进且合理。通过模型微调与优化，使其更好地适应中文应用场景，展现了我们在 AI 前沿技术应用方面的能力。

3. **系统功能完善**：系统实现了文字到图片生成和图片文字编辑两大核心功能，并且通过 Gradio 搭建了直观易用的 Web 交互界面，方便用户体验。同时，项目还支持结果的预览、调整、保存和分享，功能设计全面。

4. **成果可转化性强**：项目已产出实际可用的系统，并申请了软件著作权。这表明项目不仅停留在理论研究层面，更具备实际应用和市场推广的潜力。

5. **规范化管理与协作**：项目在管理和协作过程中充分利用了多种项目管理与协作工具，如 GitHub 进行代码与文档管理，ModelScope 和 Docker Hub 进行模型与容器管理，LaTeX 进行文档编写，提升了团队的沟通效率和项目推进效率。

不足：

1. **计算资源受限**：项目在关键的模型训练环节受到了计算资源的严重制约，特别是文字控制框架的训练，由于算力限制（至少需要 8 卡 V100），训练效果未达到预期。这导致了文字生成正确率没有显著提升。

2. **模型功能仍有提升空间**：现有模型在文字更多方面的控制上仍有待加强，例如缺乏对字体、颜色等更精细的控制。同时，对中文的支持仍需进一步优化。

3. **部署优化空间**： ModelScope 上无法持久化存储的问题导致每次启动速度较慢，影响了用户体验，未来需要进一步解决部署的效率问题。

展望：

尽管存在一些不足，但本项目为未来的发展奠定了坚实的基础。我们已计划未来采用性能更优的 AnyText2 模型，该模型增加了字体选择功能，有望解决当前模型对字体控制不足的问题。同时，我们计划在编辑板块补充更多精美的图片素材，以激发用户创作灵感，进一步提升用户体验。这些展望表明我们对项目的未来发展充满信心，并已明确了后续的优化方向。

6 参考资料

- 软件项目计划书 (submit)
- 需求分析
- 系统设计概述

- 基于生成式 AI 的个性化文创图像作品设计系统说明书
- 个人周报（内含问题报告）
- 报告 PPT