PSP25-1  
基于生成式AI的个性化文创图像作品设计系统

# 软件设计文档



姓名：郑仕博,陈奕嘉，苏泳豪  
实验部分：视觉与多模态小组  
工作站：Windows 11, Intel i5-12400F, 32GB RAM, NVIDIA RTX 4060  
日期：2025年4月22日

目录

[PSP25-1 基于生成式AI的个性化文创图像作品设计系统 1](#_Toc745808026)

[软件设计文档 1](#_Toc2080929090)

[1. 引言 3](#_Toc256540588)

[1.1 目的 3](#_Toc1677982987)

[1.2 项目范围 3](#_Toc1104810105)

[1.3 文档概览 3](#_Toc1399822773)

[1.4 参考资料 3](#_Toc1137992926)

[1.5 术语与缩略语 4](#_Toc757747100)

[2. 系统概览 4](#_Toc877482990)

[3. 系统架构 4](#_Toc1086408981)

[3.1 架构设计 4](#_Toc1369776873)

[3.2 分解描述 5](#_Toc815208671)

[3.3 设计原理 5](#_Toc266465637)

[4. 数据设计 6](#_Toc984557064)

[4.1 数据说明 6](#_Toc1089074513)

[4.2 数据字典 6](#_Toc1072216610)

[5. 组件设计 7](#_Toc622263243)

[6. 人机界面设计 8](#_Toc132964211)

[6.1 界面概览 8](#_Toc1346501397)

[6.2 界面截图 8](#_Toc466307293)

[6.3 界面控件与操作 9](#_Toc1058845548)

[7. 需求矩阵 10](#_Toc1967626194)

[8. APPENDICES 11](#_Toc792762405)

# 

# 1. 引言

## 1.1 目的

本软件设计文档描述了“基于生成式AI的个性化文创图像作品设计系统”的架构与系统设计。面向开发、测试、维护本项目的工程人员及项目管理者，作为技术实现和系统集成的参考依据。

## 1.2 项目范围

该软件旨在利用生成式AI技术解决个性化文创产品供给不足的问题，核心功能包括：根据用户输入的文本和指定的位置生成创意图像，或编辑现有图像中的文本。重点目标是实现中文字符的高精度渲染，便于游客与文创从业者快速创作独特图像作品，助力文旅融合与传播。

## 1.3 文档概览

第1章介绍目的、范围、参考资料和术语；第2章提供系统概览；第3章详细阐述系统架构；第4章描述数据设计；第5章介绍各组件设计；第6章讲解人机界面设计；第7章为需求矩阵；第8章为附录。

## 1.4 参考资料

信息来源于网页https://www.sohu.com/a/823541100\_234564。技术细节参考了 AnyText、TextDiffuser、DDPM 等文献。文档结构基于IEEE Std 1016-2009和上传的SDD模板。

## 1.5 术语与缩略语

AI（人工智能）、SDD（软件设计文档）、VAE（变分自编码器）、UNet（网络结构）、扩散模型、Stable Diffusion（SD）、AnyText（图文生成模型）、Text-control Diffusion Pipeline、Auxiliary Latent Module、Text Embedding Module、Gradio（UI库）、Prompt（文本提示）、Glyph（字形）、OCR（光学字符识别）、FID（图像质量指标）、LORA（低秩微调）、CFG-Scale（无分类引导因子）、eta（扩散采样参数）等术语在文中根据需要进一步解释。

# 2. 系统概览

本系统是一个利用生成式AI的图像创作工具，支持文本生成图像和图像内文字编辑，专注于中文字符的精准渲染。系统基于AnyText并微调Stable Diffusion模型，通过Web界面（Gradio）与用户交互，后端使用Python与深度学习框架实现，支持Docker部署。背景需求是解决文创产品同质化问题，赋能个体创作。

# 3. 系统架构

## 3.1 架构设计

系统分为三层：用户界面层（Gradio 实现）、应用逻辑层、核心模型层。

UI层：负责输入（文本、图像、参数、坐标绘制）与结果展示。

逻辑层：解析输入、格式化参数、调用模型、处理输出与数据管理。

模型层：以AnyText为核心，包括三大子模块：

1.文本嵌入模块（Text Embedding Module）：处理提示词和字形，生成文本语义嵌入；

2.辅助潜变量模块（Auxiliary Latent Module）：处理位置、掩码等空间信息；

3.文本控制扩散管道（Text-control Diffusion Pipeline）：基于UNet模型生成图像，受前两模块的条件引导，最终由VAE解码生成图像。

## 3.2 分解描述

Text Embedding Module：接收用户提示词和需渲染文本，生成对应字形图，提取嵌入后送入Transformer编码器；

Auxiliary Latent Module：接收坐标、掩码、字形图，通过卷积处理生成与扩散模型匹配的空间向量；

Diffusion Pipeline：以初始噪声为起点，联合文本嵌入与空间特征逐步去噪生成图像潜变量，最后VAE解码；

Gradio UI：提供文本输入、图像上传、画布交互、参数调节、结果展示等功能。

## 3.3 设计原理

采用AnyText + Stable Diffusion架构，针对中文文本渲染难题，结合字形信息与位置控制；Realistic\_Vision\_V4.0 作为底模保证图像质量；中文语料微调提升语义理解；Dreambooth对文化图像样式微调；Gradio 快速构建UI。相较Photoshop、Canva等工具，该方案更友好、更灵活，适配多语言和编辑需求，但也需更高计算资源和复杂集成。

# 4. 数据设计

## 4.1 数据说明

输入数据包括提示词（文本）、需渲染文本、位置坐标、参考图像（可选）、控制参数；训练数据包含两类：

AnyWord-3M标注数据（JSON格式），用于位置与字符控制训练；

文创图像+文本描述（TXT格式），用于风格微调。  
 输出图像保存在服务器（JPG/PNG格式），模型权重以文件形式存储。

## 4.2 数据字典

user\_prompt：字符串

text\_to\_render：字符串列表

position\_data：坐标列表

edit\_mask：掩码图像/张量

reference\_image：上传图像

control\_params：参数字典，如{'cfg\_scale': 7.5}

generated\_image：最终图像路径或文件

training\_data\_1：AnyWord-3M JSON结构

training\_data\_2：TXT列表与对应图像

model\_weights：模型权重文件

glyph\_image、text\_embedding、auxiliary\_latent、image\_latent：中间张量

hehe98/wenchuang: 项目镜像

wenchuang.ckpt： 模型权重

strength： 文字渲染控制强度，可以为0即不使用文字渲染

CFG-Scale： 文字控制强度，低的话会导致生成图像与描述不符合，高的话图像会不自然

eta：风格多样性，1表示启用（更具变化），0不启用（更保守）

# 5. 组件设计

主要功能以组件化方式组织，核心函数如下：

generate\_image：解析提示词，生成字形图和文本嵌入，调用辅助模块生成空间信息，联合生成潜变量图像，再解码输出；

edit\_image：编码参考图像，加入掩码噪声生成初始状态，调用编辑流程生成新图像并融合原图。

每个过程中的局部变量包括潜变量，预测噪声，注意力图等张量，模块间依次传递处理。

# 6. 人机界面设计

## 6.1 界面概览

提供Web端界面，两种主要操作模式：

1.“文本生成”：输入提示词与需渲染文本，通过画布绘制或随机选择文本位置；

2.“图像编辑”：上传图片，手动掩盖修改区域，输入新文本并生成修改结果。  
用户可调整CFG-Scale、Steps等参数，查看结果并保存。

## 6.2 界面截图

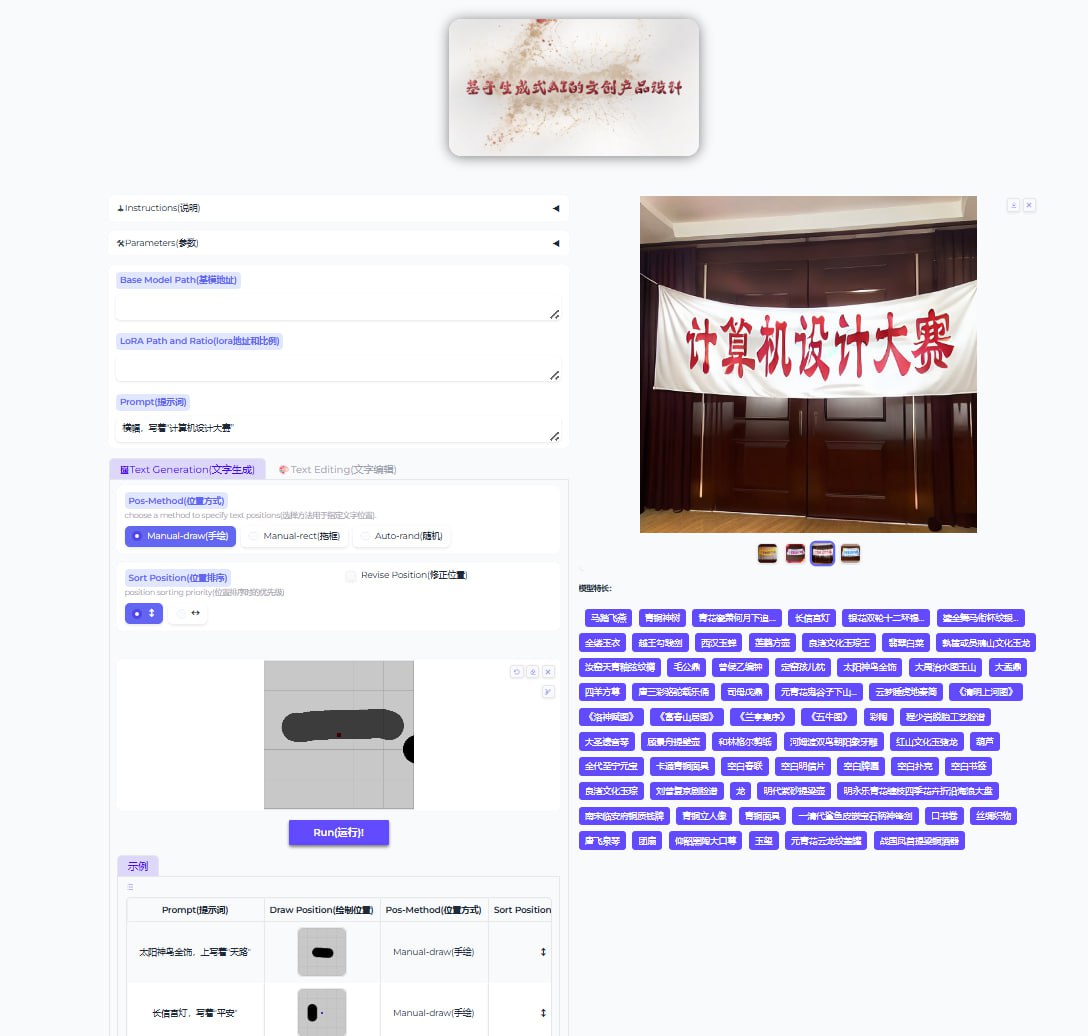


图 1: 文字到图到生成界面



图 2： 文字编辑界面

## 6.3 界面控件与操作

包括：

文本输入框（Prompt）

位置选择方式（单选按钮）

绘制画布（支持自由绘制、矩形、掩码）

参数调节控件（滑动条/输入框）

“运行”按钮

图像展示区域

上传图片控件

示例加载按钮

参考生成的物品

操作：

用户可进行输入、点击、拖动、选择文件等交互。

# 需求矩阵

| **功能需求** | **对应组件** |
| --- | --- |
| 文本输入与图像生成 | UI（输入、按钮）、逻辑层、Text Embedding、Aux Latent、Diffusion Pipeline、VAE |
| 图像上传与编辑 | UI（上传、掩码、按钮）、逻辑层、VAE 编码、辅助模块、扩散模型、VAE 解码 |
| 指定文字位置 | UI（画布、位置选择）、逻辑层、Auxiliary Latent 模块 |
| 参数调节 | UI 控件、逻辑层、Diffusion Pipeline |
| 结果预览 | UI 显示区、逻辑层 |
| 保存分享 | UI下载、后端文件存储 |
| 高中文渲染准确率 | Text Embedding、Auxiliary Latent、训练数据1 |
| 文化主题生成 | Diffusion Pipeline（微调模型）、训练数据2 |
| 文字图像自然融合 | Diffusion Pipeline、辅助模块、嵌入模块 |
| 简洁易用界面 | Gradio UI、应用逻辑 |
| 示例与指导 | UI 示例加载与静态说明 |

# APPENDICES

详见材料中的“项目注意事项”文档。