

 **个人报告**

**课 程 名 问题求解实战**

**任课教师 杨波**

**学 院 计算机学院 专 业 计算机科学与技术**

**学生姓名 陈奕嘉**

**学 号 2023141460019 年 级 2023**

**2025 年 6 月 27 日**

#### 1. 项目基本情况

本项目名为“基于生成式AI的个性化文创图像作品设计系统”，旨在解决当前文创产品同质化严重、难以满足日益增长的个性化需求的痛点。以国家推动文化和旅游融合发展，以及《如果国宝会说话》等成功案例为灵感，本项目致力于通过创新方式“激活”文化遗产，探索个体创造独特文创作品的新方式，使得个体创造文创具有现实可行性。核心目标是利用生成式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，从而满足市场对个性化文创设计的需求，并探索相关技术在文创领域的应用潜力。

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。其次是数据集构建，制作了两份数据集，一份是基于AnyWord-3M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。最后是系统实现与部署，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。

项目的主要说明性与成果性文档包含：项目计划书、报告PPT、软件著作权说明书、系统设计概述、需求分析。

本项目由四川大学计算机学院的杨波老师指导，项目组别为第一组，团队成员除我外还有组长郑仕博和组员苏泳豪。项目周期从最初的策划到最终的部署和文档完善，历时数月，期间经历了多次迭代和问题解决，最终完成了模型的训练、功能实现、系统部署。

#### 2. 承担的主要工作情况

在本项目中，我主要承担了较多的文案起草与修改、训练集搜集、模型训练、模型权重合并、项目部署、软件测试筹划执行、PPT制作等关键职责。以下是我在项目不同阶段的具体工作及完成情况：

**项目初期（Week 4-5）：** 项目计划书对未来项目规划十分重要，在这段时间，在组长的整体把关下，我起草并完成了项目计划书。同时，项目的正常开展需要大量的有效数据，由于本项目主要面向对文创产品感兴趣的群体，因而为了确保准确的生成，本项目的部分数据集需要与中华传统文化相关，我负责了相关的数据集搜集（利用爬虫）与预处理（主要是去水印）。同时我成功撰写了合并两个模型的权重并上传相关代码。

**需求分析与数据集构建（Week 6-7）：** 在这段时间，我与组长共同完成了需求分析的初稿，同时，由于我们项目需要的数据较多，我们依然在继续收集图片并处理。同时，为了使协作更规范，我优化了我们项目的文件体系结构，使其更加合理。我也与组长郑仕博共同完成了AnyText框架的微调以解决AnyText框架效果未能达到预期的问题。这一阶段结束后，我们几乎完成了项目的准备工作。

**模型训练与优化及软件著作权说明书撰写（Week 8）：**模型训练调优是本阶段的核心工作，在组长的明确分工下，我与组长郑仕博共同完成了模型部分的训练，最终在完成模型调试后将项目上传至ModelScope。同时，由于项目日益成熟，功能日益完善，出于防御性目的，我起草了我组项目的软件著作权说明书，以保护我组项目的版权。

**系统实现与部署及文档修正（Week 9-12）：** 我成功完成了软件著作权说明书全部内容的撰写，并完成了软件著作权的提交，同时我使用latex重写并修正了系统设计文档，以规范化文档，同时由于需求分析文档存在一些问题，我参与了需求分析文档的重写。此时我们的项目已经可以用于封装，所以我们共同完成了Dockerfile的编写，由于经验缺乏，我们在部署的过程中遇到了许多挑战。在克服相关挑战后，我们成功封装项目，最终实现部署。

**软件测试与最终文档准备（Week 13-14）：** 在这一阶段，我负责了软件测试的规划执行以及汇报PPT的部分内容制作。最终，软件测试通过对比训练前与训练后的相关指标完成（详细测试结果见汇报PPT15页），PPT可以清楚地介绍本小组完成的项目。

总的来说，在组长的合理规划下，我参与了本项目的几乎所有方面，承担了较多的职能，为项目的顺利完成做出了较大贡献。

#### 3. 项目实践过程中遇到的问题及处理结果

在项目实践过程中，我们遇到了一系列挑战，主要集中在技术实现和资源限制方面。

**文字叙述问题（Week 4）：** 在项目初期的文档中，由于缺乏相关的软件开发经验，我撰写的文档中部分文字叙述不得体。

**处理结果：** 我对文字叙述进行了修改和润色，确保表达准确得体。在后续文档中，这些问题逐渐得到解决，同时我们也在后续不断修正前面文档不得体的内容。

**数据集不足与开源文件存在Bug（Week 5）：** 我们面临现有数据集数量严重不足的问题，并且我们选择使用的开源文件存在Bug。

**处理结果：** 我继续收集更多有用的数据集，通过网络爬虫采集与中华文化及文物相关的图片。同时，我们不断通过各种各样的渠道寻求解决方案，以修复开源文件中的Bug。最终，我们成功制作了两份高质量的数据集。

**AnyText框架效果未达到预期（Week 6）：** AnyText框架的效果并未达到预期。

**处理结果：** 考虑使用更新的权重，之后我辅助组长进行了AnyText框架的调节。

**模型效果未达到预期（Week 7）：** 尽管进行了调节和训练，模型效果仍未达到预期。

**处理结果：** 我辅助组长对扩散模型进行了针对性的训练，得到了模型训练的最终版本。

**文字生成正确率未达到预期（Week 8）：** 尽管模型效果显著提升，但文字生成正确率并未达到预期。

**处理结果：** 我们采用了官方的AnyText权重并与训练好的扩散模型进行合并。未来可以考虑使用更新的模型，并且增加改变字体的功能。项目展望中也提及，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。

**文档规范性与周报统一性欠佳（Week 10）**：在老师的提醒下，我们发现我们的文档存在不规范问题。

**处理结果：**于是，为了确保开发过程的严谨性，我们考虑并最终使用latex重写了相关文档。同时我们发现我们的周报存在格式不统一，难以统一阅看的问题，于是我们经过商讨，最终确认了一个统一格式并以该格式重写之前的所有周报

**ModelScope持久化存储问题导致启动慢（Week 12）：** 在ModelScope上部署后，我们发现项目无法持久化存储，每次启动都非常慢。（huggingface存储空间有限）

**处理结果：**后来我不断查找启动速度慢的原因，与组长共同进行了测试。

这些问题的解决，体现了我们小组在项目管理和技术攻关方面的能力，同时说明我们的小组配合默契，分工明确，善于发现并解决问题，这也为我们项目的最终成功奠定了基础。

#### 4. 参与项目过程的体会与项目评价

**参与项目过程的体会：**

参与“基于生成式AI的个性化文创图像作品设计系统”项目是一次宝贵的学习和成长经历。作为项目组长，我不仅在技术层面得到了显著提升，更在项目管理、团队协作和问题解决能力方面积累了丰富的经验。

首先，在技术方面，我深入了解并实践了生成式AI模型，特别是AnyText模型和扩散模型的训练与优化。从最初的AnyText模型微调、数据集构建，到最终的模型部署和性能测试，每一个环节都让我对深度学习模型有了更深刻的理解。我学会了如何处理和筛选大规模数据集，如何针对特定应用场景（如中文文创）对模型进行定制化训练，以及如何评估模型的性能并解决过拟合等问题。特别是对文字控制框架的训练和扩散模型的合并，让我对文本与图像融合的复杂性有了更直观的认识。尽管遇到了算力限制的挑战，但我们通过研究和调整，最终选择了最合适的模型权重，这让我体会到在实际项目中，权衡技术理想与现实条件的重要性。

其次，在项目管理和团队协作方面，我作为组长，从项目计划书的修订、团队成员的分工、任务进度的督促，到文档的规范化编写和汇报材料的准备，全程参与并主导了各项工作。这让我深刻体会到清晰的项目规划、有效的沟通机制和明确的职责分工对于项目成功的关键作用。每周的周报制度和定期的团队会议确保了我们能够及时发现并解决问题，例如前后端开发进度、数据集收集困难、模型效果不理想等。通过与陈奕嘉和苏泳豪两位组员的紧密合作，我学会了如何发挥每个人的长处，共同克服困难。尤其是在Docker封装和ModelScope部署等复杂环节，团队成员的协同努力是不可或缺的。

最后，在问题解决方面，面对数据集不足、开源文件Bug、模型效果不佳以及算力限制等一系列挑战，我们没有退缩。我积极带领团队分析问题根源，寻找解决方案，无论是通过论坛寻求技术支持，还是调整训练策略，都体现了我们解决问题的能力。特别是ModelScope持久化存储的问题，让我认识到在实际部署中，除了技术实现，还需考虑平台特性和用户体验。这些经历让我学会了如何在不确定性中保持灵活性和适应性，并不断探索新的方法。

**项目评价：**

我认为本项目取得了显著的成果，并在多个方面展现出其价值和创新性。

**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，展现了团队在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，并成功申请了软件著作权。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。这些展望表明团队对项目的未来发展充满信心，并已明确了后续的优化方向。

总而言之，本次项目经历让我对AI技术在实际应用中的潜力和挑战有了深刻的认识。我不仅在技术能力上得到了锤炼，更在团队协作、问题解决和项目管理方面积累了宝贵经验，为我未来的学习和职业发展奠定了坚实基础。