

 **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 郑仕博**

**学 号 2023151610075 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。在国家推动文化和旅游融合发展，以及《如果国宝会说话》等成功案例的启发下，本项目致力于通过创新方式“激活”文化遗产，赋能个体创造独特文创作品。核心目标是利用先进的生成式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，从而满足市场对个性化文创设计的需求，并探索相关技术在中文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。

本项目由四川大学计算机学院的杨波老师指导，隶属于计算机科学与技术专业的第一组，团队成员包括组长郑仕博（我本人），以及组员陈奕嘉和苏泳豪。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，并申请了软件著作权。

#### 2. 承担的主要工作情况

在本项目中，我主要承担了项目统筹与规划、模型训练与调优、核心代码编写与上传以及项目部署等关键职责。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）：** 作为组长，我首先对项目计划书进行了细致的修订，并明确了团队成员的具体分工。我安排组员负责搜集图片素材以构建数据集，并着手进行前后端开发工作，而我则专注于模型的训练环节，并深入研究应对过拟合问题的有效策略。在第五周，我组织组员开展了图片收集工作，并与组员共同学习扩散模型的结构。我还编写并修改了相关权重文件，对收集到的图片进行了初步处理和筛选，例如去除水印、优化图像质量等，以确保数据的可用性和一致性。
* **需求分析与数据集构建（Week 6-7）：** 在第六周，我与组员陈奕嘉一同完成了需求分析系统的初稿，并与组员一起收集数据集图片，寻找标注工具并指导组员进行标注。第七周，我与陈奕嘉一起完善了需求分析文档的细节，并与组员一起进行了AnyText框架的调节。同时，我也与苏泳豪一起进行了扩散模型的训练。
* **模型训练与优化（贯穿项目始终，尤其在Week 8）：** 模型训练与调优是我的核心工作。我专注于模型训练，研究并解决了过拟合问题。在第八周，我与组员一起完成了模型的训练，得到了最终版本，并由我上传至ModelScope。我还负责编写了模型测评文件代码并上传至GitHub。在项目总结报告中也提到，我负责了模型微调与优化，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。为了更贴合主题，在Realistic Vision V4.0 (基于sd1.5)权重基础上，我使用dreambooth的方法进行微调，并与微调后的AnyText模型的权重进行整合。
* **系统实现与部署（Week 9-12）：** 在第九周，我与组员一起完成了前后端的制作，并上传至GitHub，修改了标题并增加了一些示例。在第十周，我进行了系统概述文档和需求分析文档的修改，用LaTeX重新书写，使文档更加正式和规范。第十一周，我们完成了Dockerfile的编写，并将项目封装并上传至Docker Hub，解决了正常部署可能因网络问题导致文件下载不下来的问题。在第十二周，我负责编写了requirements.txt文件，并将其与代码部分一起上传至ModelScope，进行了部署。
* **文档完善与汇报准备（Week 13-14）：** 在项目后期，我持续参与了文档的完善工作。在第十三周，我们对两份文档进行了修饰，修改了需求分析并增加了技术难点。第十四周，我们完善了文档，完成了测试报告，并开始制作报告PPT。

总的来说，我不仅承担了项目管理和统筹的职责，确保项目按计划进行，还深入参与了核心技术部分的开发，特别是模型训练与部署，为项目的顺利完成做出了关键贡献。

#### 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。

* **前后端不完善与文字叙述问题（Week 4）：** 在项目初期，我们发现前后端并不完善，并且部分文字叙述不得体。
  + **处理结果：** 我立即督促组员进一步完善前后端开发，并对文字叙述进行了修改和润色，确保表达准确得体。在后续周报中，这些问题逐渐得到解决，例如在Week 9，我们完成了前后端的制作并上传至GitHub。
* **数据集不足与开源文件存在Bug（Week 5）：** 我们面临现有数据集不足的问题，并且使用的开源文件存在Bug。
  + **处理结果：** 我指导组员继续收集更多有用的数据集，例如通过网络爬虫采集与中华文化及文物相关的图片。同时，我们积极通过论坛等渠道寻求解决方案，以解决开源文件中的Bug。在后续的数据集制作部分，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与扩散模型结合不理想（Week 6）：** AnyText框架的效果并未达到预期，且扩散模型与框架结合的效果不理想。
  + **处理结果：** 我们考虑使用更新的权重，并深入研究框架对扩散模型生成的影响。在Week 7，我们与组员一起进行了AnyText框架的调节，并与苏泳豪进行了扩散模型的训练。最终，我们决定在中文训练集上对AnyText模型进行专门训练，以更好地适应中文应用场景。
* **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：** 我设定了一周内训练出模型的明确目标，并计划将其上传至ModelScope。在Week 8，我们成功完成了模型的训练，得到了最终版本，并由我上传至ModelScope。
* **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：** 在项目总结报告中，文字正确率测试结果也印证了这一点，训练前后文字生成和修改的正确率及编辑距离没有显著提升，这与算力、参数及训练批次有关。最终，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。
* **前后端美化与部署难度（Week 9）：** 前后端美化有待加强，且部署可能会存在一些难度。
  + **处理结果：** 我将模型的封装到Docker镜像中，以方便部署。在Week 11，我们完成了Dockerfile的书写，并将项目封装并上传至Docker Hub，解决了正常部署可能会因为网络问题导致文件下载不下来的问题。
* **需求分析文档有待改进，未进行Docker封装（Week 10）：** 需求分析文档有待改进，且当时尚未完成Docker封装。
  + **处理结果：** 我计划进一步改进需求分析文档，并用Docker进行封装。如前所述，Docker封装在Week 11成功完成。
* **模型未部署到服务器，无法演示（Week 11）：** 项目存在未将模型部署到服务器上，导致无法进行演示的问题。
  + **处理结果：** 我将模型部署到服务器上，方便后续进行演示，并进行对照实验。
* **ModelScope持久化存储问题导致启动慢（Week 12）：** 在ModelScope上部署后，发现无法持久化存储，导致每次启动非常慢。
  + **处理结果：** 我计划对部分文档按照老师的建议进行修正，并寻求ModelScope的持久化存储解决方案。
* **时间有限，工作量较大（Week 13）：** 本周时间有限，所做工作有限，有较多工作有待完成。
  + **处理结果：** 我计划未来进行测试，继续完善文档，并准备起草展示的PPT和测试报告。在Week 14，我们完善了文档，完成了测试报告，报告PPT正在制作中。
* **算力限制：** 在项目总结报告中明确指出，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。
  + **处理结果：** 尽管算力限制是一个难以立即解决的外部因素，但我们通过调整策略，最终采用了官方的AnyText权重并与训练好的扩散模型进行合并，以保证项目的最终效果。项目展望中也提到，未来需要寻求更强大的计算资源支持。

这些问题的解决，体现了团队在项目管理和技术攻关方面的能力，也为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。团队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。