

武汉大学

本科毕业论文（设计）

基于大模型的跨模态古诗创作

姓 名： 张志东
学 号： 2021302111480
专 业： 软件工程
学 院： 计算机学院
指 导 教 师： 朱卫平

二〇二五年四月

摘 要

[illegible]

请使用中文分号“;”分割关键词!

关键词: 关键词 1; 关键词 2; 关键词 3

ABSTRACT

This is abstract. This is abstract. This is abstract. This is abstract. This is abstract.
This is abstract. This is abstract. This is abstract. This is abstract. This is abstract. This
is abstract. This is abstract. This is abstract. This is abstract. This is abstract. This is
abstract. This is abstract. This is abstract. This is abstract.

Please use English semicolon and space to separate key words.

Keywords: Key1; Key2; Key3

目 录

摘要	I
ABSTRACT	II
1 绪论	1
1.1 研究背景和意义	1
1.2 研究现状	3
1.2.1 中文古诗生成	3
1.2.2 古诗质量评价	4
1.2.3 大模型技术	4
1.2.4 提示工程	5
1.3 研究思路和主要贡献	5
1.4 论文组织结构	6
2 古诗创作与生成	7
3 系统设计与实现	8
4 实验及结果分析	9
5 结语	10
参考文献	11
致谢	15

1 绪论

本章主要介绍研究背景与意义、中文古诗生成相关的研究现状，阐述本论文的研究思路与主要贡献，并介绍论文的组织架构。

1.1 研究背景和意义

文本生成任务（Text Generation）是自然语言处理（Natural Language Processing, NLP）的一个重要研究方向，需要在给定输入或上下文的条件下，输出符合要求的文本，涵盖机器翻译、文本摘要、对话生成、作品创作等多个应用方向。而在文本生成领域中，中国古代诗歌的生成更是一个困难的任务。

古诗是中华优秀传统文化的瑰宝，作为最早形成的中国古代文学作品体裁之一，其措辞简洁、内涵丰富且韵律整齐。中国古诗的典故意象运用极具文化特色、用词凝练优雅，这些独特的艺术特色都为古诗的机器创作带来巨大的挑战。如何让机器模型深入理解并创作出具有文化内涵和艺术价值的古诗，继而为中华优秀传统文化的再创造赋能，是一个富有挑战的有趣问题。

早期的古诗生成主要依赖于其他子领域的研究思路。例如，[1] 利用现有古诗作为模板，根据既定规则替换字词来生成新的古诗，这样生成的古诗虽在形式上合格，但表达力欠佳；[2] 将其看作是一个摘要生成任务，只是输入是作者的表达意图，且需要考虑中文古诗独特的韵律形式约束；[3] 则将其看作一个机器翻译任务，将古诗的上下句子分别看作是翻译的源语言和目标语言，利用统计机器翻译（Statistical Machine Translation, SMT）的方法来生成古诗。

随着深度学习技术的发展，近年的中文古诗生成大多将古诗生成视作是“从序列到序列”的预测任务（Sequence-to-Sequence），并由此出发训练循环神经网络（Recurrent Neuro Networks, RNNs），如编码器-解码器（Encoder-Decoder）模型 [4]，并以此为基础设置额外机制来增强语义表现 [5] 或韵律格式约束 [6][7]。然而，这些方法对古诗内涵的掌握往往停留在上下文语义或是韵律对仗，无法进一步触达诸如典故、意象和全诗连贯性等更复杂的方面。所幸，大模型（Large Model, LM）展现出强大的语义理解与创作能力，而在中文领域也出现了诸如 ERNIE[8]、DeepSeek-R1[9] 这样的大语言模型（Large Language Model, LLM）和

DeepSeek-VL[10] 等跨模态大模型，为这一领域注入全新的活力。[11]

除了文本输入外，中国古诗往往蕴含着丰富的场景意象，其对应的视觉信息难以通过用户输入来精确描述。目前也有方案直接使用图像作为输入，如在循环神经网络外增加卷积神经网络（Convolutional Neuro Networks, CNNs）以处理图像信息，捕捉图像关键主体并把握整体氛围，最终生成古诗。[12][13] 但相应地，这些方案放弃了文本输入能够具体描述要求的优势，输入图像所含信息的繁杂也导致生成古诗主题、内涵乃至风格的波动。现有的方案大都局限于或文本或图像的单一模态输入，要么局限于上下文语义或韵律的形式规律而无法触达更高的艺术层次，要么受制于图像信息的多变而无法实现更精细的输出控制。这种单一模态的输入限制了用户描述需求的可能，也使生成的古诗缺乏层次，这暗示着文本“跨模态”输入的研究方向，也是本选题希望探讨的内容。

另一个重要的问题是生成结果的“可解释性”（Interpretability），其指系统以人类可理解的术语解释或呈现模型行为的能力。[14] 深度学习技术在带来更高表现的同时，其“黑盒”的特性也使得可解释性问题愈发突出。在古诗生成任务中，可解释性可体现为三个方面：1）过程可解释性，即系统是如何从输入的文本和图像中提取信息，并进一步并生成古诗的，其中包含哪些中间步骤；2）结果可解释性，即系统为何生成这样的古诗结果，其中遵守怎样的韵律形式、又运用了哪些典故意象；3）反馈可解释性，即系统如何分析生成古诗的质量，如何让用户理解古诗“好坏在哪里”。现有的古诗生成系统大多缺乏对生成过程的可解释性，用户难以理解系统是如何从输入信息中提取出关键信息并生成古诗的；而在结果可解释性方面，现有的系统也往往只提供了生成古诗的文本，而没有进一步解释其韵律、意象等方面的内涵；在反馈可解释性方面，现有的系统也缺乏对古诗质量评价维度的详细说明，用户需要具备较高的文学素养以理解和甄别对古诗质量的评价。因此，提升古诗生成系统的可解释性，将有助于增强用户对系统的信任感和使用体验，使得用户能够更好地理解和利用系统输出的古诗和相应的质量判断。

本文旨在探讨文图跨模态的中文古诗生成，开发了一个基于大模型的古诗创作系统。在给定用户两种模态输入的条件下，其能够通过图像的分析描述来充分提取图像信息，结合用户输入的文本信息，生成符合古诗韵律和意象的古

诗。此外，系统还将提供对古诗的分析文本、量化评分以及改进意见，涵盖韵律对仗、典故意象、主题思想、语言用词等多个赏析方面，并支持多轮迭代优化。

1.2 研究现状

1.2.1 中文古诗生成

近年来，中文古诗生成领域引起了广泛的研究兴趣。2016 年，[15] 使用修改的注意力结构的编码器-解码器模型来解决古诗生成过程中主题漂移的问题，但限制关键词的数量和顺序，降低了系统的灵活性。这一问题在 2018 年由 [13] 通过记忆网络（增加记忆组块的 RNNs）基于图像生成古诗解决。

古诗的生成可能会出现多方面的质量波动，包括主题、语言风格、字数格式、韵律对仗等等，需要在研究中纳入考量。2020 年，[6] 设计了一个基于 Transformer 的自回归模型，改进注意力机制并进一步收紧了包括中文古诗、歌词、英文十四行诗等特殊文体生成的格式要求。2021 年，[16] 从图像中提取物体关键词、情感和风格，以生成古诗；[17] 将数十万首古诗按照风格、情感、格式与主要关键词分类，并利用掩蔽自注意力机制来建立标签到诗句的关联，以此来生成情感与风格均可控的古诗。2023 年，[18] 构建了一个古诗图像数据集，精确标注了其中的诗歌元素，并利用 GRU 网络来增强生成古诗的上下文关联。2024 年，[7] 首次使用扩散模型来生成古诗，以实现语义与韵律的同时控制。

后来逐渐出现使用 LLM 生成古诗的方案。2024 年，[11] 修正了 LLM 以 token 计数会导致输出格式错误的问题，改以字符计数，达到了极高的格式精度。[19] 则提出了一种图片输入的三段式绝句生成方法，将短语特征纳入考量，并构建了一个图像主题数据集。其将输入的图片映射到一个主题词，再随机选择与该主题词相关的短语，再通过一正一反两种方向的 LLM 来依次生成古诗的首行、标题和其他主体内容。

值得注意的是，鲜有研究关注文图跨模态的古诗生成。据调研，目前只有 [20] 一项研究同时包含文本和图像两种模态的输入，其通过 Clarifai 来将图像映射到两个具体的主题词，根据这两个主题词在已有短语库中进行检索拓展，再进一步将得到的短语通过一个自注意网络来生成古诗。这一系统允许用户来限制古诗的诗句前缀词，因此可广义地认为实现了图文的跨模态输入。

1.2.2 古诗质量评价

如何评价生成古诗的质量是一个难题，考虑到古诗体裁的艺术性，其质量评价往往依赖于人工评估。除此之外，也可使用以往文本生成的自动度量方法，如源自机器翻译领域基于 n -gram^①的翻译文本评估方法 BLEU[21] 和 ROUGE[22]。其中，BLEU 指标计算生成文本中有多少 n -gram 出现在参考文本中，即使用精确率（Precision）来评估生成文本有多接近参考文本；相反，ROUGE 指标计算参考文本中有多少 n -gram 能够被生成文本包含，即使用召回率（Recall）来评估生成文本能否完整地覆盖参考文本。可见，BLEU 和 ROUGE 指标均依赖于与高质量参考文本的对比。此外，Distinct 指标基于互信息来衡量字符的多样性，以脱离参考文本独立地评价文本质量。[23]

2020 年，[24] 提出一个质量感知的掩蔽语言模型，实现一个可迭代的古诗优化框架，可用于判断古诗是否需要优化，并在润色时定位不恰当部分。2023 年，[25] 提出一种人机协作的古诗创作系统，能够在不同的约束条件下对古诗进行润色。2024 年，[26] 又提出了一个可迭代的古诗优化框架，基于 BiLSTM 和 CRF 构建用于检测低质量用词的检测网络、基于 BERT 模型构建用于修正用词的校正网络。这些研究均实现了“评价 + 优化”的流程功能，且效果良好。美中不足的是未提供对“评价”本身的解释，即“被选中的词为何是低质量的，而优化又依据着什么”的问题。

1.2.3 大模型技术

近年来，大模型技术取得了显著的进步，其中以大语言模型最为典型，例如 OpenAI 开发的生成预训练转换器（GPT）系列模型、百度开发的知识整合增强表示（ERNIE）系列模型 [8]。DeepSeek 开发了首个完全通过强化学习训练的大语言模型 DeepSeek-R1[9]，其采用混合专家模型（MoE）架构，实现在相同计算成本下大幅提升模型的参数规模与推理能力。而与 LLM 擅长自然语言类似，还有许多跨模态大模型适用于不同模态间数据的信息表征处理，如文本和图像。由 OpenAI 开发的 CLIP 模型包含一个文本解码器和一个图像解码器，在大量图像

^①表示文本中连续的 n 个词或字符。如“我爱你”中基于字符的 2-gram 集合为[“我爱”、“爱你”]

及其对应的文本描述的数据集上进行预训练，因而能够把握视觉表征与文本之间的关联。[27]MiniGPT-4 是 GPT-4 模型的缩小版，它将一个与 BLIP-2 架构相同的视觉编码器与语言模型 Vicuna 通过一个单一的投影层链接起来，在图像描述生成方面展现出卓越的能力。[28] DeepSeek 开发的 DeepSeek-VL 的训练数据来自广泛的现实场景，其采用了一个混合视觉编码器，可在固定的 token 预算内有效处理高分辨率图像（1024 x 1024），同时保持相对较低的计算开销，这一设计选择确保了模型在各种视觉任务中捕捉关键语义和详细信息的能力。有趣的是，DeepSeek-VL 在一开始便整合了 LLM 的能力训练，促进视觉和语言两种模态能力的平衡整合，使其在能够捕捉视觉语义信息的同时，仍然具有强大的语言能力。[10]

1.2.4 提示工程

提示工程逐渐作为一种通过自然语言来调整和控制 LLM 行为的技术。目前人们已提出多种提示词的设计原则与策略。Few-shot 框架为模型提供少量示例样本，以有效地指示模型完成指定任务。[29] 思维链（CoT）框架指导模型将任务分解为若干子任务来逐步完成，使模型能在无其他修改的情况下完成复杂的因果推理任务。[30] 自洽性（Self-Consistency）旨在小样本 CoT 中对多种推理路径进行采样，并在尝试生成之后选择最一致的答案，其有助于提高 CoT 提示在涉及算术推理和常识推理的任务中的性能。[31] 此外，开发者们也提出了十分多样的提示词框架，例如 CRISPE、ICIO、BROKE，均经过了开源社区的检验，令人目不暇接。

1.3 研究思路 and 主要贡献

本选题旨在利用大模型的通用能力，设计一个文图跨模态的中文古诗创作系统：1）分析输入的图像，生成易使用的描述文本，兼顾关键物体与整体氛围；2）设计评判标准，设计提示词分别用于古诗生成和分析，调用模型生成古诗、分析结果与改进意见；3）结合已有的质量评价维度和其他系统，设计实验验证古诗分析分数、跨模态生成对古诗质量影响的效度，并探索系统改进方向。

TODO: 详细阐述技术路线

本选题的工作将有助于推动人工智能技术在文化艺术领域的广泛应用，特

别是为中华优秀传统文化的传承和创新提供新的思路。通过跨模态输入的结合与大模型的创新应用，系统不仅能够生成具备深厚文化底蕴的古诗作品，还能够提升对生成结果的解释性，使得用户更易理解和信任机器生成的创作过程及其背后的文化内涵。此外，该系统的多轮迭代优化能力也将推动生成模型在艺术创作中的精细化与多样化应用，提升古诗创作的质量和艺术价值。最终，本研究有助于拓展和弘扬中华优秀传统文化的表达形式，推动诗词创作与人工智能技术的深度融合，具有重要的理论意义和实践价值。

.....

1.4 论文组织结构

第一章是概论部分。

第二章是对古诗生成的相关研究现状进行综述，介绍古诗生成的研究背景、研究现状和发展趋势。

第三章是系统设计与实现，介绍系统的整体架构、各个模块的功能和实现方法。

第四章是实验与结果分析，介绍系统的实验设计、数据集构建、实验结果和分析。

第五章是结语，对本文的工作内容进行总结，并探讨局限性与改进空间。

2 古诗创作与生成

3 系统设计与实现

4 实验及结果分析

5 结语

参考文献

- [1] OLIVEIRA H G. PoeTryMe: a versatile platform for poetry generation[J]. Computational Creativity, Concept Invention, and General Intelligence, 2012, 1 : 21.
- [2] YAN R, JIANG H, LAPATA M, et al. I, Poet: Automatic Chinese Poetry Composition through a Generative Summarization Framework under Constrained Optimization[J]. IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence, 2013 : 2197 – 2203.
- [3] HE J, ZHOU M, JIANG L. Generating Chinese Classical Poems with Statistical Machine Translation Models[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2012, 26(1) : 1650 – 1656.
- [4] YI X, LI R, SUN M. Generating Chinese Classical Poems with RNN Encoder-Decoder[A]. SUN M, WANG X, CHANG B, et al. Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data[C], Cham : Springer International Publishing, 2017 : 211 – 223.
- [5] ZHANG X, LAPATA M. Chinese Poetry Generation with Recurrent Neural Networks[A]. MOSCHITTI A, PANG B, DAELEMANS W. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)[C], Doha, Qatar : Association for Computational Linguistics, 2014 : 670 – 680.
- [6] LI P, ZHANG H, LIU X, et al. Rigid Formats Controlled Text Generation[A]. JURAFSKY D, CHAI J, SCHLUTER N, et al. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics[C], Online : Association for Computational Linguistics, 2020 : 742 – 751.
- [7] HU Z, LIU C, FENG Y, et al. PoetryDiffusion: Towards Joint Semantic and Metrical Manipulation in Poetry Generation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2024, 38(16) : 18279 – 18288.

- [8] ZHANG Z, HAN X, LIU Z, et al. ERNIE: Enhanced Language Representation with Informative Entities[J], 2019(arXiv:1905.07129).
- [9] DeepSeek-AI. DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning[J], 2025(arXiv:2501.12948).
- [10] LU H, LIU W, ZHANG B, et al. DeepSeek-VL: Towards Real-World Vision-Language Understanding[J], 2024(arXiv:2403.05525).
- [11] YU C, ZANG L, WANG J, et al. CharPoet: A Chinese Classical Poetry Generation System Based on Token-free LLM[A]. CAO Y, FENG Y, XIONG D. Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)[C], Bangkok, Thailand : Association for Computational Linguistics, 2024 : 315 – 325.
- [12] LIU L, WAN X, GUO Z. Images2Poem: Generating Chinese Poetry from Image Streams[A]. MM '18: Proceedings of the 26th ACM International Conference on Multimedia[C], New York, NY, USA : Association for Computing Machinery, 2018 : 1967 – 1975.
- [13] XU L, JIANG L, QIN C, et al. How Images Inspire Poems: Generating Classical Chinese Poetry from Images with Memory Networks[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2018, 32(1).
- [14] Doshi-Velez F, KIM B. Towards A Rigorous Science of Interpretable Machine Learning[J], 2017(arXiv:1702.08608).
- [15] WANG Z, HE W, WU H, et al. Chinese Poetry Generation with Planning Based Neural Network[A]. MATSUMOTO Y, PRASAD R. Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers[C], Osaka, Japan : The COLING 2016 Organizing Committee, 2016 : 1051 – 1060.
- [16] WU C, WANG J, YUAN S, et al. Generate Classical Chinese Poems with Theme-Style from Images[J]. Pattern Recognition Letters, 2021, 149 : 75 – 82.

- [17] SHAO Y, SHAO T, WANG M, et al. A Sentiment and Style Controllable Approach for Chinese Poetry Generation[A]. CIKM '21 : Proceedings of the 30th ACM International Conference on Information & Knowledge Management[C], New York, NY, USA : Association for Computing Machinery, 2021 : 4784–4788.
- [18] REN X, CHAI X, MAO M. Generating Chinese Poetry from Images Based on Deep Learning[J]. Proceedings of the 2023 3rd International Conference on Big Data, Artificial Intelligence and Risk Management, 2023 : 134–138.
- [19] LIU D, GUO Q, LI W, et al. A Multi-Modal Chinese Poetry Generation Model[A]. 2018 International Joint Conference on Neural Networks (IJCNN)[C], 2018 : 1–8.
- [20] LIU Y, LIU D, LV J. Deep Poetry: A Chinese Classical Poetry Generation System[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(09): 13626–13627.
- [21] PAPINENI K, ROUKOS S, WARD T, et al. BLEU: A Method for Automatic Evaluation of Machine Translation[A]. ISABELLE P, CHARNIAK E, LIN D. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics[C], Philadelphia, Pennsylvania, USA : Association for Computational Linguistics, 2002 : 311–318.
- [22] LIN C-Y. ROUGE: A Package for Automatic Evaluation of Summaries[A]. Text Summarization Branches Out[C], Barcelona, Spain : Association for Computational Linguistics, 2004 : 74–81.
- [23] LI J, GALLEY M, BROCKETT C, et al. A Diversity-Promoting Objective Function for Neural Conversation Models[A]. KNIGHT K, NENKOVA A, RAMBOW O. Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies[C], San Diego, California : Association for Computational Linguistics, 2016 : 110–119.
- [24] DENG L, WANG J, LIANG H, et al. An Iterative Polishing Framework Based on

- Quality Aware Masked Language Model for Chinese Poetry Generation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(05): 7643 – 7650.
- [25] MA J, ZHAN R, WONG D F. Yu Sheng: Human-in-Loop Classical Chinese Poetry Generation System[A]. CROCE D, SOLDAINI L. Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations[C], Dubrovnik, Croatia : Association for Computational Linguistics, 2023 : 57 – 66.
- [26] CHEN Z, CAO Y. A Polishing Model for Machine-Generated Ancient Chinese Poetry[J]. Neural Processing Letters, 2024, 56(2): 77.
- [27] RADFORD A, KIM J W, HALLACY C, et al. Learning Transferable Visual Models From Natural Language Supervision[A]. Proceedings of the 38th International Conference on Machine Learning[C], [S.l.] : PMLR, 2021 : 8748 – 8763.
- [28] ZHU D, CHEN J, SHEN X, et al. MiniGPT-4: Enhancing Vision-Language Understanding with Advanced Large Language Models[J], 2023(arXiv:2304.10592).
- [29] BROWN T, MANN B, RYDER N, et al. Language Models Are Few-Shot Learners[A]. Advances in Neural Information Processing Systems : Vol 33[C], [S.l.] : Curran Associates, Inc., 2020 : 1877 – 1901.
- [30] WEI J, WANG X, SCHUURMANS D, et al. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models[J]. Advances in Neural Information Processing Systems, 2022, 35 : 24824 – 24837.
- [31] WANG X, WEI J, SCHUURMANS D, et al. Self-Consistency Improves Chain of Thought Reasoning in Language Models[J], 2023(arXiv:2203.11171).

致谢

以简短的文字表达作者对完成论文和学业提供帮助的老师、同学、领导、同事及亲属的感激之情。