Large Model Based Crossmodal Chinese Poetry Creation

*L. Yang, **Z. Zhang**, K. Niu, S. Pan, W. Zhu, C. Ma*

Wuhan University

IEEE Advancing Technology for Humanity

The professional home for the engineering and technology community worldwide

Introduction

Backgrounds

- **Natural Language Generation** is considered one of the most challenging fields in NLP, as it focuses on the "*creation*" of the models.
- **Chinese ancient poetry**, one of the most valuable heritages of human culture, conveys rich connotations through its concise form and elegant language.

Due to the importance of 1) further exploring language models' ability and 2) promoting Chinese traditional culture, the automatic generation of Chinese classical poetry has attracted much research interest.

However, there remain some challenges.

The professional home for the engineering and technology community worldwide

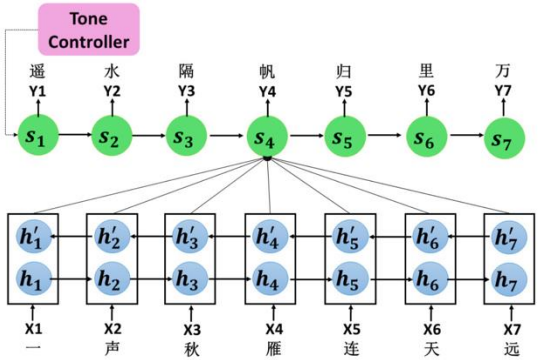
Introduction

Challenges

1) To achieve a deep understanding of *general* aspects. (e.g. imagery and cohesion)

- **Text2poem: RNN** Encoder-Decoder (X. Yi et al. 2017), which takes it as a *seq2seq* task.

  Traditional systems can produce good poems *in form* but may fail to reach a deeper content.



(X. Yi et al. 2017)

- **Text2poem:** “CharPoet” Based on **LLM** (C. Yu et al. 2024)

Fortunately, **large models** (LMs) have shown significant potential in this field, and there’ve been LMs with good capacity in Chinese.

The professional home for the engineering and technology community worldwide

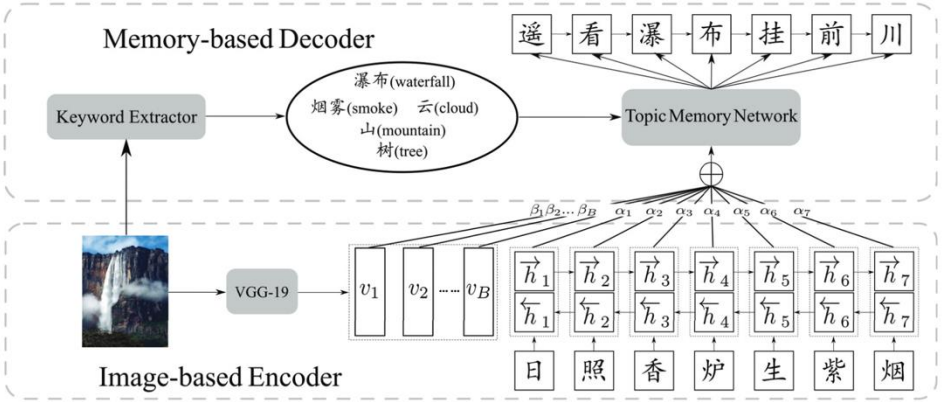
Introduction

Challenges

2) To achieve *cross-modal* generation across text and image

- **Image2poem: RNN** integrated with **CNN** and attentional structure. (L. Xu et al. 2024) (L. Liu et al. 2018)

Despite good performance, existing systems are limited to single-modal generation.



(L. Xu et al. 2024)

3) To improve the *interpretability* of outcome poems

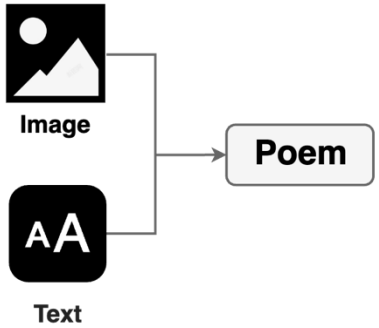
4) To achieve *multiple rounds* of generation

The professional home for the engineering and technology community worldwide

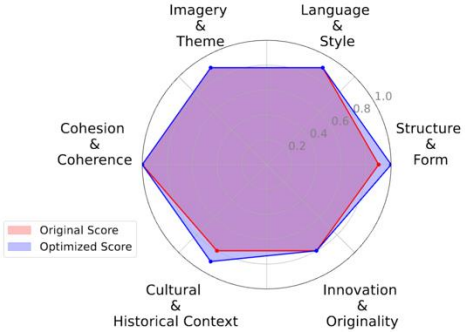
Introduction

The Proposed System

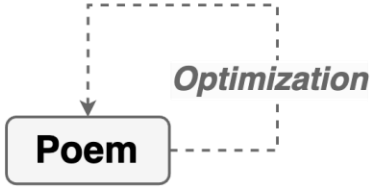
The primary contributions of our proposed system are:



Cross-modal



Interpretability



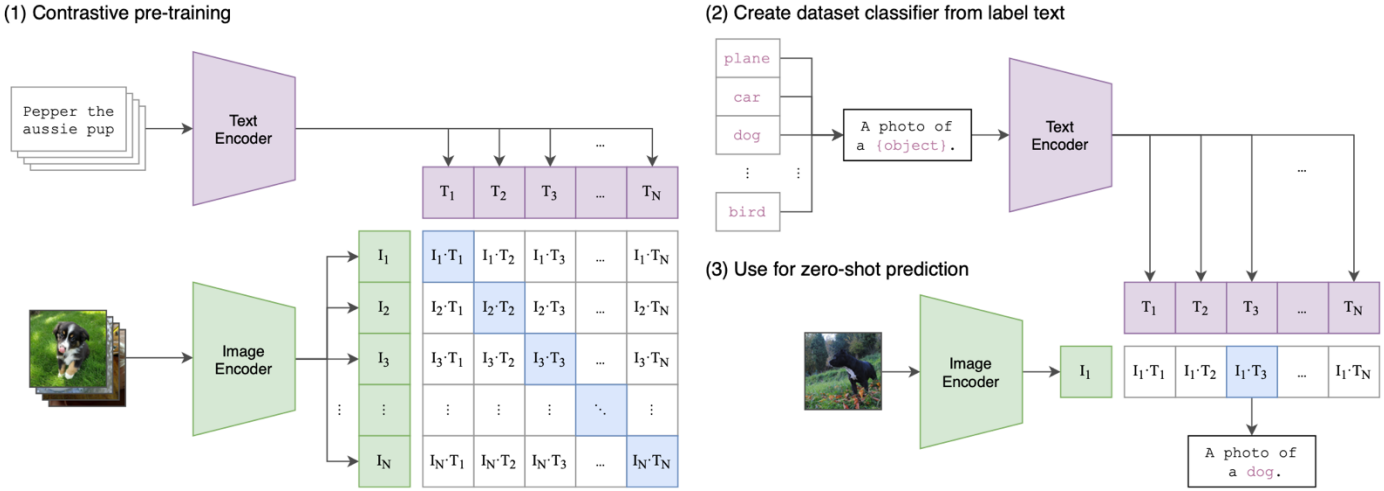
Multiple optimizations

The professional home for the engineering and technology community worldwide

Related Work

CLIP (To extract keywords of images)

- **CLIP** (**C**ontrastive **L**anguage–**I**mage **P**retraining) is a pre-trained large model that contains an **image encoder** and a **text encoder**.



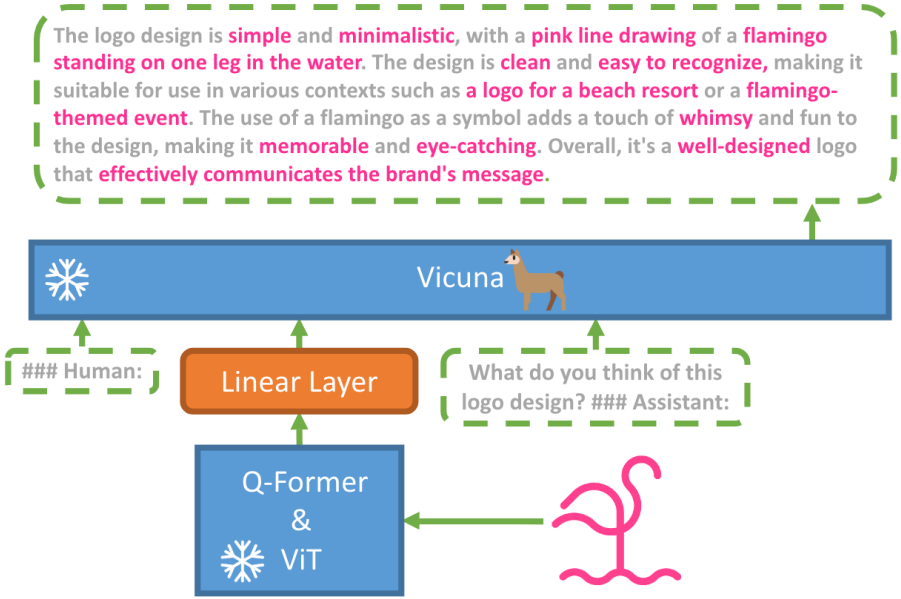
(Radford et al. 2021)

The professional home for the engineering and technology community worldwide

Related Work

MiniGPT4 (To generate descriptions of images)

- **MiniGPT4**, a model pairing **a visual encoder**, which adopts the same architecture as BLIP-2, with **a language model**, Vicuna, via **a single linear layer**.



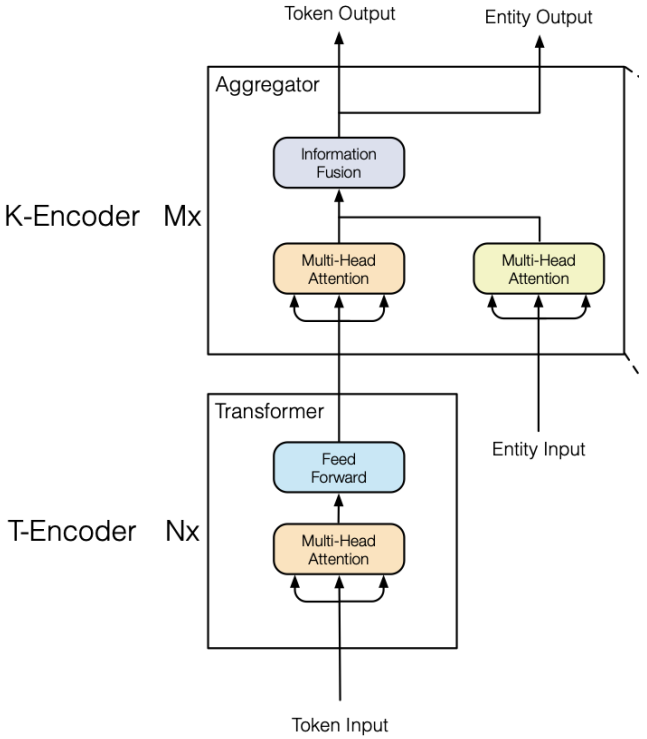
(Zhu et al. 2021)

The professional home for the engineering and technology community worldwide

Related Work

ERNIE (Poem Generation)

- **ERNIE** (**E**nhanced **R**epresentation through k**N**owledge **I**ntegration), a large language model developed by Baidu, contains two encoder:
  - **a Text-Encoder (T)**: basic lexical and syntactic information
  - **a Knowledge-Encoder (K)**: additional lexis-based knowledge information, incorporated into the text information.



(a) Model Achitecture

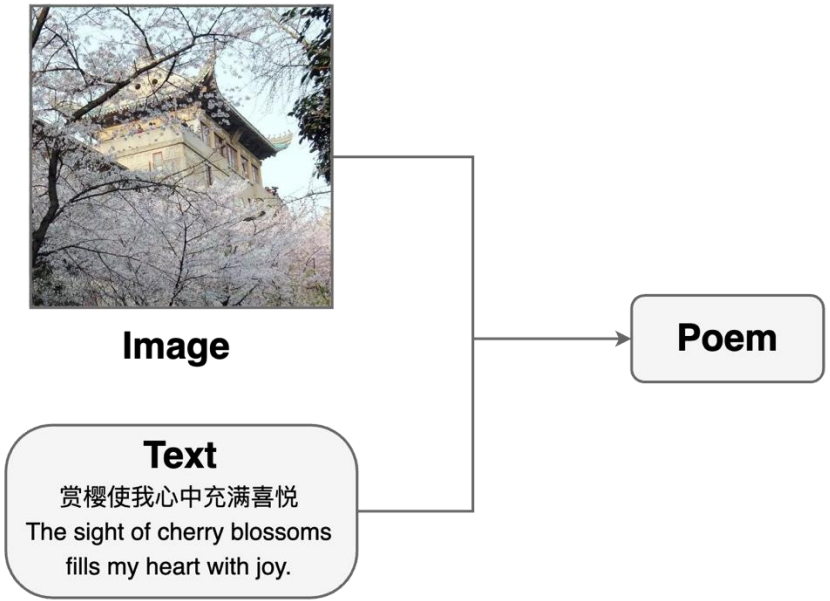
(Zhang et al. 2019)

The professional home for the engineering and technology community worldwide

The Proposed System

Example Scenario

Say we have a landscape **photo** and a sentence of **text**, then want to produce a Chinese poem according to both.

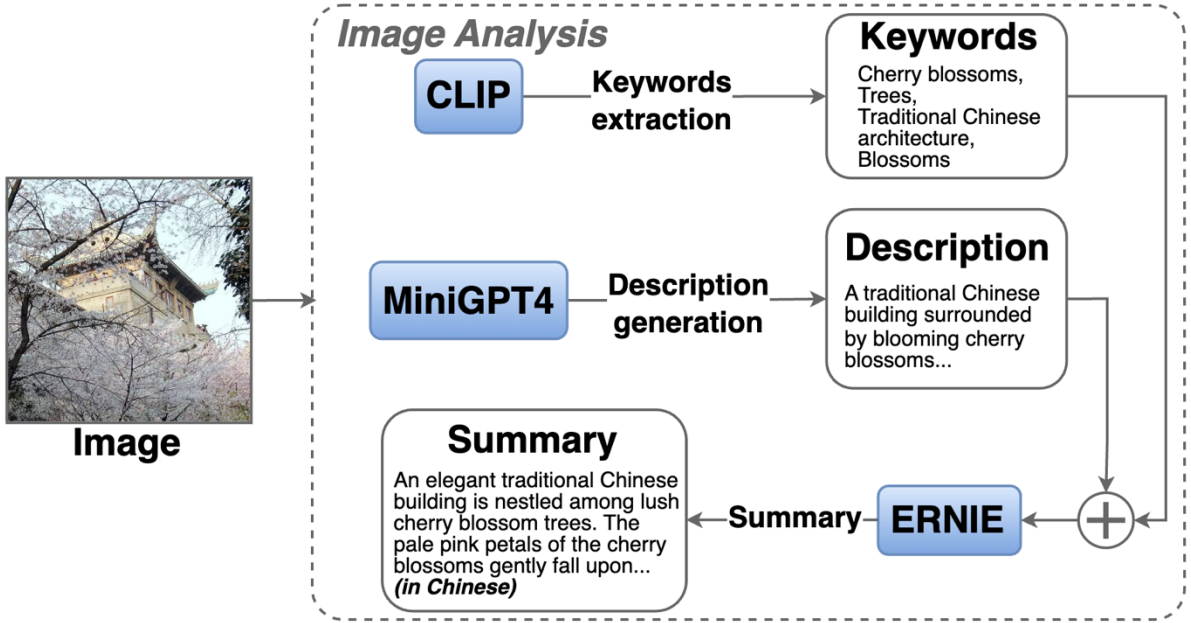


The professional home for the engineering and technology community worldwide

The Proposed System

Image Analysis

The first step is to derive a **description** from the image, where we took advantage of three models:



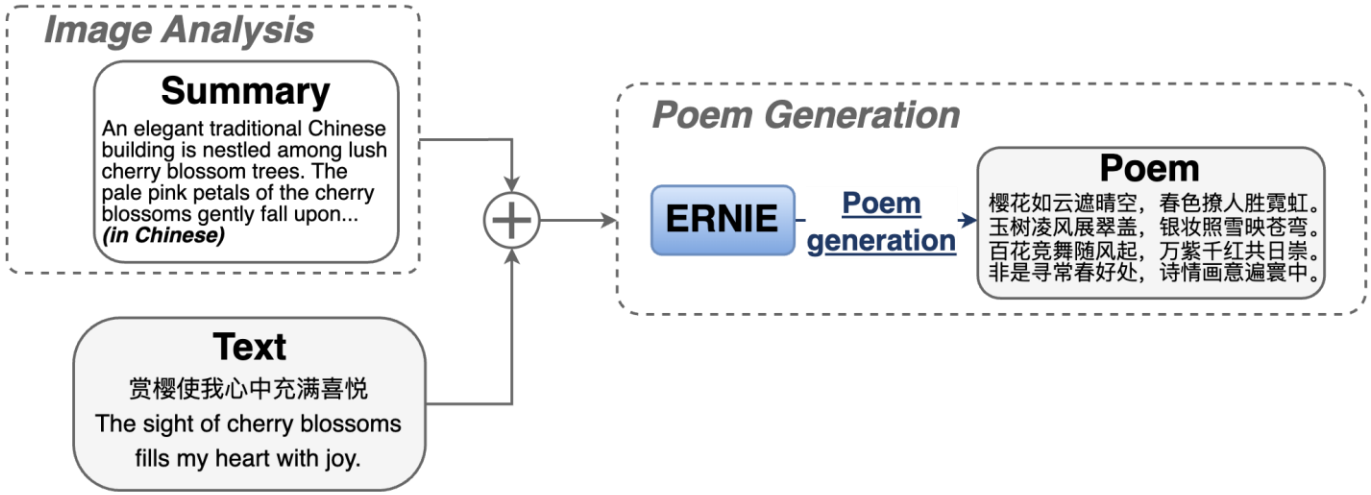
- CLIP
- MiniGPT4
- ERNIE

The professional home for the engineering and technology community worldwide

The Proposed System

Poem Generation

Then, **the image description** can be combined with the **input text** as the real input of the ERNIE model.

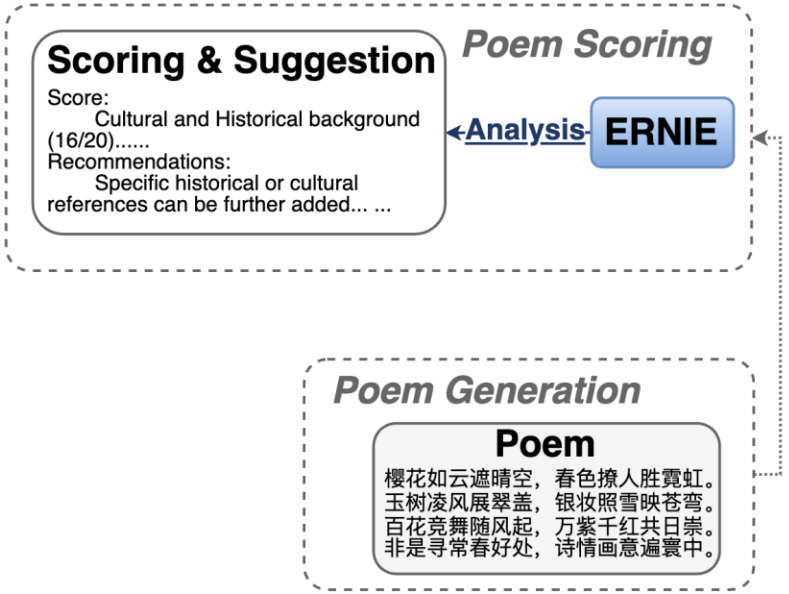


The professional home for the engineering and technology community worldwide

The Proposed System

Poem Scoring

The generated poem can be analyzed, producing **quantified scoring** and **suggestions** for further improvement.



Total Scoring: *89/100*

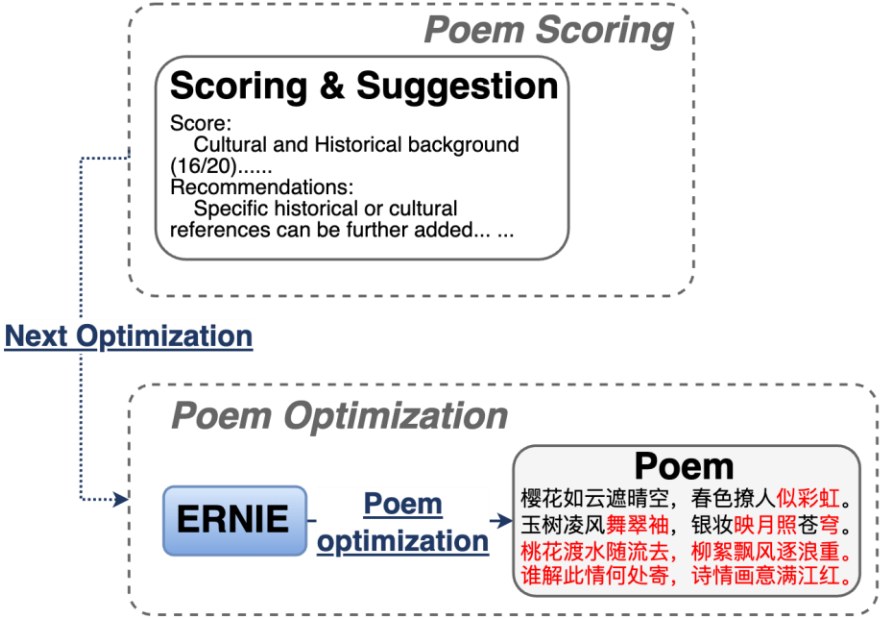
1. **Structure & Form:** *10/10*
The poem has a clear structure, follows the rhythm and format of traditional poetry, and shows a high degree of regularity. Without improvement, it has shown a high degree of regularity.
2. **Language & Style:** *18/20*
The words are precise and poetic, and the rhetoric is diverse, such as "cherry blossoms cover the clear sky like clouds", which vividly depicts the scene of spring, but the innovation can be further strengthened. Try to use more innovative figures of speech and expression to enhance the artistic charm of poetry.
3. **Imagery & Theme:** *27/30* ...
4. **Cohesion & Coherence:** *10/10* ...
5. **Cultural & Historical Context:** *16/20* ...
6. **Innovation & Originality:** *8/10* ...

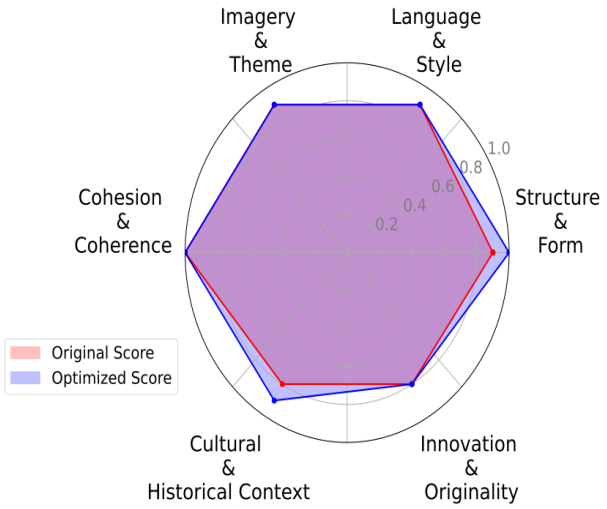
The professional home for the engineering and technology community worldwide

The Proposed System

Poem Optimization

Finally, the ERNIE model generates **an optimized version** of the previous poem according to the suggestions.





The professional home for the engineering and technology community worldwide

The Proposed System

Overall Structure

Image

Text
The sight of cherry blossoms fills my heart with joy. ***(in Chinese)***

Image Analysis

CLIP — Keywords extraction → **Keywords**: Cherry blossoms, Trees, Traditional Chinese architecture, Blossoms

MiniGPT4 — Description generation → **Description**: A traditional Chinese building surrounded by blooming cherry blossoms...

Summary: An elegant traditional Chinese building is nestled among lush cherry blossom trees. The pale pink petals of the cherry blossoms gently fall upon... ***(in Chinese)***

Summary — ERNIE

Poem Scoring

Scoring & Suggestion
Rating: Cultural & Historical Context (16/20)...
Suggestions: specific historical or cultural references can be further added... ***(in Chinese)***

Analysis — ERNIE

Next round of optimization

Poem Generation & Optimization

ERNIE — Poem generation → **Poem**
樱花如云遮晴空，春色撩人似彩虹。
玉树凌风舞翠袖，银妆映月照苍穹。
桃花渡水随流去，柳絮飘风逐浪重。
谁解此情何处寄，诗情画意满江红。

The professional home for the engineering and technology community worldwide

Prompt Engineering

Poem Generation & Optimization

The prompts used for poem generation & optimization follow most of the **CRISPE** framework.

CRISPE

Capacity & Role	You are the Ancient Rhyme creation platform, a multi-functional system focused on generating and appreciating ancient Chinese poetry. Your abilities include creating Tibetan poems, context-specific poems, and couplets according to user needs.
Insight	You understand the rich tradition and aesthetics of ancient Chinese poetry, including the structure, rhythm and language style of different poetic forms. You also understand how modern users appreciate and interact with ancient poetry, including their emotional reactions to poetry and expectations of the creative process.
Statement	**Poem Generation**: Your task is to generate ancient poems that match the user's needs. This involves taking specific instructions from the user regarding subject matter, style, length, etc., and producing poems that meet these requirements. Here's what the user wants: ***"{user_text}"***. Meanwhile, the user enters an image, which is converted into a text description based on the content of the image analysis: ***"{overall_description}"*** **Poem Optimization**: Your task is to modify the original poem based on the poem given by the user and suggestions for improvement. This is the user's poem ***"{poem}"*** and the improvement suggestion ***"{suggestion}"***.
Personality	Your style should be classical, reflecting the charm and depth of ancient Chinese poetry. At the same time, your response should be concise and direct to the user's needs and instructions.
Attention	The content you generate should strictly conform to the level and structure rules of Chinese poetry. The number of words in each line should be the same. You need to be able to adapt and optimize based on user feedback to deliver a product that is more relevant to your users' needs.

The professional home for the engineering and technology community worldwide

Prompt Engineering

Poem Scoring

The prompts used for poem scoring make use of the **Few-shot** framework, which contains detailed explanations of the **criteria**, **output format**, as well as two I/O **examples**.

Criteria

The following are the detailed scoring rules used to automatically evaluate Chinese classical poetry. Based on these rules, I will provide Chinese poetry texts for you to evaluate and score. After completing the scoring, I also need you to provide specific suggestions for improvement of these ancient poems, especially those with low scores. These suggestions should focus on how to improve the quality of Chinese poetry, including but not limited to improving the structure, the use of words, and the depth of themes.

...

Output Format

Please ensure that all output strictly follows the following formatting requirements to facilitate subsequent data analysis and processing.
Scoring format: Each evaluation category (structure and form, language and style, imagery and theme, overall coherence, cultural and historical context, innovation and originality) is scored separately. The score for each evaluation category should be given in the format *[evaluation category]: [score]/[full score]*. Finally, the total score of all categories is given.
Improvement suggestion format: Specific improvement suggestions are provided for each evaluation category. Suggestions should be given in the format *[Evaluation categories]: [specific suggestions]*.

Example 1

Input:
...
Output:
...

Example 2

Input:
...
Output:
...

The professional home for the engineering and technology community worldwide

Prompt Engineering

Poem Scoring

Criteria

...

1. **Structure and Form** - 10 points
   Type of poetry: up to 5 points.
   Evaluation of compliance with the basic structure and rules of the specified type (e.g., poems of four lines or eight lines).
   Prosodic rules: up to 5 points.
   Analyze whether the rhyme of the poem is regular and consistent with the traditional rhyme rules.
2. **Language and Style** - 20 points
   Word selection: up to 10 points.
   Assess the appropriateness, richness, and originality of words.
   Figure of speech: up to 10 points.
   Assess the appropriate use and creativity of rhetoric.
3. **Imagery and Themes** - 30 points
   Use of imagery: up to 15 points.
   Assess the originality, appropriateness, and expressiveness of the imagery.
   Topic depth: up to 15 points.
   Assess the depth and sentiment expression of the topic.
4. **Cohesion and Coherence** - 10 points
   Internal logic: up to 5 points.
   Assess the logical coherence between verses.
   Emotional coherence: up to 5 points.
   Assess the consistency and fluency of emotional expressions.
5. **Historical Context** - 20 points
   Cultural references: up to 10 points.
   Assess the accuracy and appropriateness of cultural and historical elements in the poem.
   Historical context adaptability: up to 10 points.
   Assess the fit of the content with the era context.
6. **Originality and Innovation** - 10 points
   Unique perspective: maximum 5 points. Evaluate the novel ideas or expressions offered.
   Creative approach: maximum 5 points. Evaluate innovations in structure, language, or subject matter.

The professional home for the engineering and technology community worldwide

Evaluation

Criteria & Optimization

* the 6th Chinese Traditional Creation Competition (www.shicizhongguo.cn)

To test the validity of **criteria**, we selected three types of poem sets, got the scorings as well as the optimized versions.

Poem set	Number	Quality Level
Famous ancient Tang poems	100	Highest
Awarded works *	20	Medium
Doggerel poems	20	Lowest

TABLE I

COMPARISON OF AVERAGE PERFORMANCE BETWEEN FIVE DIFFERENT POEM COLLECTIONS ON ALL SIX ASPECTS

	Famous	Opt_Awarded	Awarded	Opt_Doggerel	Doggerel
Structure & Form	**0.950**	0.885	0.860	0.850	0.770
Language & Style	**0.873**	0.845	0.843	0.818	0.720
Imagery & Themes	**0.888**	0.847	0.845	0.823	0.727
Cohesion & Coherence	**0.968**	0.915	0.905	0.880	0.825
Cultural & Historical Context	0.656	**0.750**	0.720	0.615	0.433
Originality & Innovation	**0.744**	0.710	0.735	0.685	0.645

The professional home for the engineering and technology community worldwide

Evaluation

Input & Generation

To test the efficacy of **cross-modal**, we further compared the scorings of three types of input, using the same content as in the example.

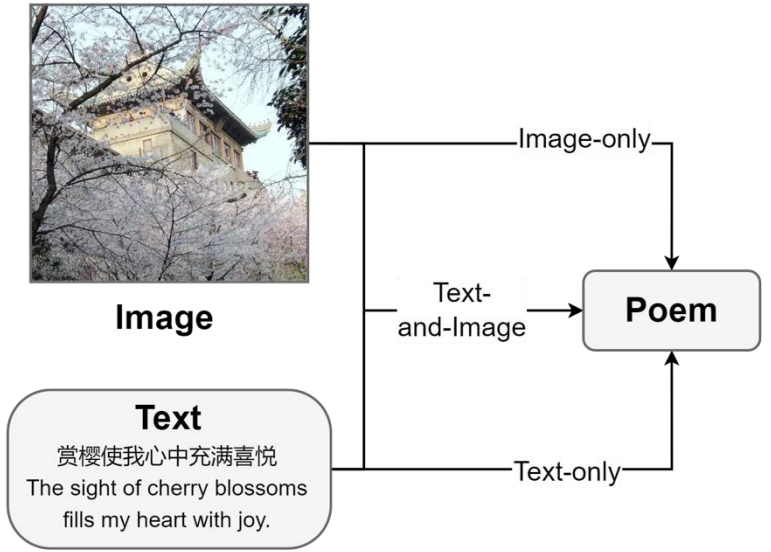


TABLE II

COMPARISON OF AVERAGE PERFORMANCE BETWEEN THREE DIFFERENT INPUT TYPES ON ALL SIX ASPECTS

	Text-and-Image	**Text-only**	**Image-only**
Structure & Form	**0.940**	0.910	0.880
Language & Style	**0.885**	0.870	0.850
Imagery & Themes	**0.897**	0.887	0.867
Cohesion & Coherence	**0.970**	0.940	0.920
Cultural & Historical Context	**0.755**	0.610	0.660
Originality & Innovation	**0.772**	0.750	0.710

The professional home for the engineering and technology community worldwide

Conclusion

Summary

- We realized the cross-modal generation of Chinese ancient poems combined with text and image,
  - **Image Analysis**: extracts keywords and produces a *description* of the uploaded image
  - **Poem Scoring**: produces *scoring*, *analysis*, and improvement *suggestions* on six aspects of generated poems
  - **Poem Generation and Optimization**: generate ancient poems given either image description and input text or the original poem and suggestion.
- We tested the system by
  - testing the **criteria** and the **optimization** on three poems set with different quality
  - comparing the quality of poems generated **under three input modalities**

IEEE
The professional home for the engineering and technology community worldwide

Conclusion

Limitations

- The **criteria** used for poem scoring and further improvement are mostly based on the *prompts and the ERNIE model* particularly.
- Similarly, it can further validate the reliability of the criteria by comparing the system's automatic scoring with **human expert scoring results** on the same poem set.
- Besides, this research may lack **quality comparisons with other existing platforms** or with this system, where fewer models are used for image analysis.

In the future, we aim to further develop our system, trying new models and technology to upgrade the effect of the system, expand functions, and improve user interaction experience.

IEEE
The professional home for the engineering and technology community worldwide

Thank you for listening

*L. Yang, **Z. Zhang**, K. Niu, S. Pan, W. Zhu, C. Ma*

Wuhan University

{zhidong.zhang, wpzhu}@whu.edu.cn