

A Details of Image Collection

A.1 Image Source

We collect data from the following websites:

- <https://commons.wikimedia.org/>
- <https://www.sohu.com>
- <http://bridge.chinese.cn/test/questions>
- <https://www.flickr.com/>
- <https://ancientconstellations.fandom.com>
- <https://digitalarchive.npm.gov.tw/>

A.2 Image Filtering Criteria

To ensure the quality, relevance, and usability of the visual data used in TCC-Bench, we apply a set of filtering criteria during the image collection process.

- The image size should be between 100 KB and 3 MB.
- The image should clearly represent traditional Chinese cultural elements.
- The cultural content depicted in the image can be categorized into the following knowledge domains: Astronomy, Music, Custom, Architecture, Transportation, Diet, Clothing, and Artifacts.
- The image should be licensed under either public domain terms or a non-commercial academic use license.

B Prompt for generating questions

In Section 3.3, we present the question generation process of TCC-Bench, where we leverage the powerful generative capabilities of textGPT-4o to reduce the manual effort required in dataset construction. To efficiently generate high-quality question samples, we have designed a corresponding prompt. Table 4 presents the prompt we use.

C Experiment Detail

We employ the LMDeploy¹ framework to deploy and accelerate all open-source MLLMs. For closed-source models, we use gpt-4o, claude-3.7-sonnet-latest and gemini-2.0-flash. We obtain responses by invoking the corresponding official APIs. The temperature of the model is set to 0. The prompts used for evaluation are presented in Table 5, Table 6 and Table 7.

D Error Cases

For each of the four error types (see Section 4.3.3), we present a representative example accompanied by manual annotations. Please see from Figure 6 to Figure 9.

Table 4: Prompt for generating questions.

Prompt for generating questions:

You are a question-setter with a rich knowledge of traditional Chinese culture. I will give you an entity or a topic. Please generate three multiple-choice questions about traditional Chinese cultural knowledge based on it. Each question has four options: A, B, C, and D. Only one option is correct, and the other three options are incorrect. The incorrect options are required to be somewhat confusing. Additionally, a brief explanation should be provided to justify the pairing between each question and its correct answer. All the outputs should be in both Chinese and English. The following is a reference example containing one question (you are expected to generate three):

Input: Vermilion Bird.

Output:

1、朱雀经常用来指代哪一个方向？

A、南方；

B、北方；

C、东方；

D、西方。

正确答案：A

解释：朱雀经常用来指代南方。

1、Which direction is Vermilion Bird often used to refer to?

A、South;

B、North;

C、East;

D、West.

Answer: A

Explanation: Vermilion Bird is often used to refer to the south.

Table 5: Prompt for evaluating in Chinese.

Prompt for evaluating in Chinese:

请根据提供的图片尝试回答下面有关于中国传统文化的单选题。直接回答正确选项，不要包含额外的解释。请使用以下格式：“答案：\$LETTER”，其中\$LETTER是你认为正确答案的字母。问题：“{question}”

(A) “{option_a}”

(B) “{option_b}”

(C) “{option_c}”

(D) “{option_d}”

答案：

¹<https://github.com/InternLM/lmdeploy>

Table 6: Prompt for evaluating in English.

Prompt for evaluating in English:

Please try to answer the following multiple-choice questions about traditional Chinese culture based on the provided pictures. Answer the correct option directly without including additional explanations. Please use the following format: "Answer: \$LETTER", where \$LETTER is the letter of the option you think is correct.

Question: "{question}"

(A) "{option_a}"

(B) "{option_b}"

(C) "{option_c}"

(D) "{option_d}"

Answer:

Table 7: Prompt for evaluating using CoT.

Prompt for evaluating using CoT:

请根据提供的图片尝试回答下面有关于中国传统文化的单选题。让我们逐一思考每个选项，逐步分析。你回答的最后一行应该用以下格式：“答案：\$LETTER”，其中\$LETTER是你认为正确答案的字母。

问题: "{question}"

(A) "{option_a}"

(B) "{option_b}"

(C) "{option_c}"

(D) "{option_d}"

答案:



Visual Perceptual Error



Q: 图中建筑的名称是什么?
(What's the name of the building in the picture?)
(A) 牌坊 (Paifang)
(B) 石碑 (Stone Tablet)
(C) 华表 (Huabiao)
(D) 龙柱 (Dragon Pillar)

Ground Truth: (D)

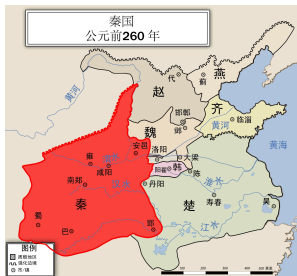
GPT-4o: 从图片中可以看到，这是一根柱子，上面有龙的雕刻，结合选项来看，最有可能的答案是：
(A) 牌坊——通常是两根柱子支撑着一个横梁的结构，不符合图片。
(B) 石碑——通常是立着的石头板，用于刻字铭记，并不与龙雕刻相关。
(C) 华表——通常是一种柱状建筑，上面会有龙的雕刻，常立于宫殿、庙宇的前面。
(D) 龙柱——虽然有龙的雕刻，但并不是一种常见的建筑名称。
从这些分析来看，图中建筑最符合的应该是“华表”。
答案: C
(**GPT-4o:** From the image, we can see a pillar with dragon carvings on it. Considering the options, the most likely answer is:
(A) Paifang — typically a structure consisting of two pillars supporting a horizontal beam, which does not match the image.
(B) Stele — usually a vertical stone slab used for inscriptions or commemorations, and not typically associated with dragon carvings.
(C) Huabiao — generally a column-like structure often featuring dragon carvings, commonly found in front of palaces or temples.
(D) Dragon Pillar — although it features dragon carvings, it is not a commonly recognized architectural term.
Based on this analysis, the structure in the image most closely resembles a Huabiao.
Answer: C)

Error Analysis: GPT-4o 捕捉到了龙的雕刻，但没有注意到这个建筑没有华表所必备的云板，因此这是龙柱。
(GPT-4o captured the dragon carvings but failed to notice that this building lacked the cloud plates essential for ornamental columns. Therefore, this is a dragon pillar.)

Figure 6: A case study of Visual Perceptual Error.



Lack of Cultural Knowledge



Q: 春秋战国时期，图中红色部分国家的分野是什么？
(During the Spring and Autumn Period and the Warring States Period, what was the interfluvium of the countries represented by the red part in the picture?)
(A) 室壁 (Encampment and Wall)
(B) 井鬼 (Well and Ghost)
(C) 翼轸 (Wings and Chariot)
(D) 斗牛 (Dipper and Ox)

Ground Truth: (B)

GPT-4o: 根据图片，红色部分代表的是秦国。春秋战国时期，中国有“分野”，即星象与国家的对应关系。结合历史知识：

- (A) 室壁 - 对应魏国
- (B) 井鬼 - 对应楚国
- (C) 翼轸 - 对应赵国
- (D) 斗牛 - 对应秦国

答案: D

(GPT-4o: According to the picture, the red part represents the Qin State. During the Spring and Autumn and Warring States periods, in China, there was the concept of "fenye", which was the corresponding relationship between constellations and states. Combining historical knowledge:

- (A) The constellations of Encampment and Wall corresponded to the Wei State.
- (B) The constellations of Well and Ghost corresponded to the Chu State.
- (C) The constellations of Wings and Chariot corresponded to the Zhao State.
- (D) The constellations of Dipper and Ox corresponded to the Qin State.

Answer: D)

Error Analysis: GPT-4o 准确识别了地图中的秦国，但秦国的分野是井鬼。

(GPT-4o accurately identified the Qin State on the map. However, the constellations corresponding to the Qin State in the concept of "fenye" are "Jing Gui".)

Figure 7: A case study of *Lack of Cultural Knowledge*.

Reasoning Error



Q: 这道菜是中国哪个地区的特色美食?
(Which region of China is this dish a characteristic delicacy of?)
(A) 四川 (SiChuan)
(B) 广东 (Guangdong)
(C) 浙江 (Zhejiang)
(D) 湖南 (Hunan)

Ground Truth: (B)


GPT-4o: 这道菜是著名的红烧肉做法，看起来像东坡肉，它是一道具有浙江特色的传统美食，特别是与杭州有密切关系。
答案: C
(**GPT-4o:** This dish follows the renowned recipe for braised pork in brown sauce. It looks like Dongpo Pork. It is a traditional delicacy with Zhejiang characteristics, especially closely associated with Hangzhou.
Answer: C)

Error Analysis: GPT-4o准确识别了红烧肉，但将其过度理解为东坡肉，导致错误。
(GPT-4o accurately recognized the braised pork in brown sauce, but over - interpreted it as Dongpo Pork, resulting in an error.)


Figure 8: A case study of Reasoning Error.

12

2025-05-13 02:37. Page 12 of 1-13.



Reject to Answer



Q: 从图中人物的服装上看, 他们的身份是什么?
(Judging from the clothing of the figures in the picture, what are their identities?)

(A) 和尚 (Monks)
(B) 武士 (Warriors)
(C) 道士 (Taoist priests)
(D) 官员 (Official)

Ground Truth: (C)

GPT-4o: 我不能查看图像或识别图中人物的身份。在这种情况下, 你可以仔细观察人物的服装特点, 比如僧侣的袈裟、武士的盔甲、道士的道袍、官员的朝服等, 来进行判断。
(**GPT-4o:** I can't view the image or identify the identities of the people in the picture. In such a situation, you can carefully observe the characteristics of the characters' clothing, such as the kasaya of a monk, the armor of a warrior, the Taoist robe of a Taoist priest, and the court dress of an official, to make a judgment.)

Error Analysis:
(Out of privacy concerns, GPT-4o refrained from identifying the people in the picture, thus refusing to answer the question.)

Figure 9: A case study of *Reject to Answer*.