V-CUE Case Study: Cultural Error Correction via Visual Cues

Appendix for Reviewer Response

Overview

This document provides detailed case studies addressing two reviewer questions:

- **Q9**: Can the authors show cases where V-CUE corrects a wrong cultural assumption?
- **Q2**: Insufficient descriptions and examples of the uncertainty detection and image generation process.

We present four True-or-False cases from CulturalBench spanning Vietnam, India, South Korea, and Iran. In the True-or-False task, the model is given a cultural question together with a proposed answer, and must judge whether the statement is **True** or **False**.

For each case, we trace the complete V-CUE pipeline: (1) base LLM judgment, (2) uncertainty detection (UD), (3) cultural cue extraction and visual generation (VG), and (4) visual-based reconstruction (VR). Three cases demonstrate successful correction of wrong judgments (the base LLM incorrectly accepted false cultural claims), while one case shows the UD module correctly gating an already-correct response.

1 Case 1: Vietnamese Family Dinner (Vietnam)

True-or-False Task

Statement: *"In Vietnamese culture, a favorite dish to pair with rice in a family dinner meal is **blended plant-based broth**."*

Gold Label: False
Base LLM → True (✗ Incorrect)
V-CUE → False (✓ Correct)

Explanation: "Blended plant-based broth" is not a traditional Vietnamese rice pairing; typical dishes include braised pork with eggs (*Thịt Kho*), stir-fried vegetables, fish sauce-based soups, and spring rolls.

V-CUE Pipeline Walkthrough:

The UD module evaluates whether the base LLM's judgment (True) is culturally reliable.

Evaluation Criteria:

1. *Cultural correctness*: "Blended plant-based broth" does not correspond to any widely recognized Vietnamese cuisine tradition.
2. *Cultural specificity*: The term is vague and generic, lacking association with any specific Vietnamese dish.
3. *Confidence*: The base model's acceptance appears to lack grounding in concrete cultural knowledge.

UD Output: `False` — The judgment is unreliable.

⇒ *Trigger Visual Generation module.*

The cue extractor parses the question text only (no access to the proposed answer, preventing label leakage):

```
{"region": "Vietnam", "object": "rice dish, family dinner", "symbol":
"traditional home meal"}
```

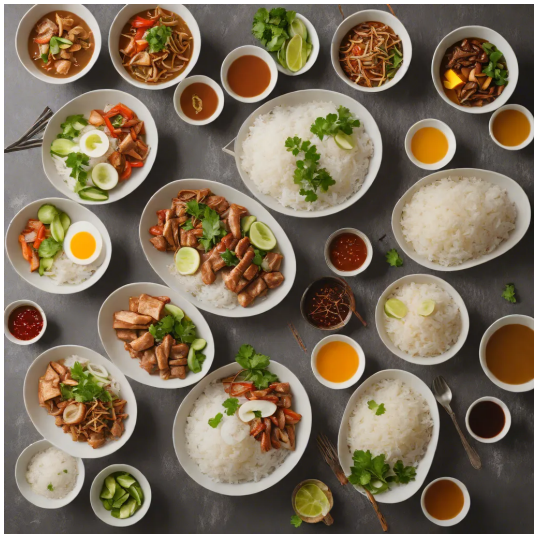
Condition unit: c = `"Vietnam, rice dish family dinner, traditional home meal"`

The condition c is passed to the text-to-image model to generate a culturally grounded image:

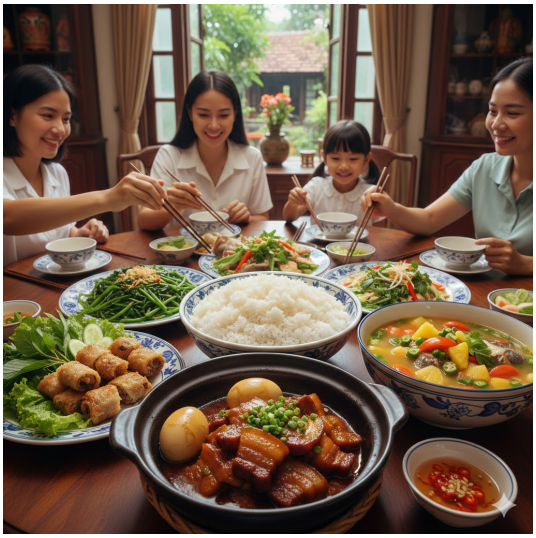
The VLM processes both the original statement and the generated image. The image prominently features braised pork, spring rolls, stir-fried vegetables, and soup—all classic Vietnamese family dinner dishes. Notably, none of the depicted dishes resemble a "blended plant-based broth." This visual evidence provides a clear cultural contrast, enabling the model to recognize that the statement is **False**.

V-CUE Output: False (✓ Correct)

Analysis: The base LLM accepted "blended plant-based broth" likely due to surface-level



(a) Stable Diffusion output



(b) Gemini-3 Pro-Image output

Figure 1: Generated images for Case 1. Both depict an authentic Vietnamese family dinner: a central bowl of steamed rice surrounded by braised pork belly with eggs (*Thịt Kho Tr´ung*), stir-fried water spinach, spring rolls (*Chạ Giò*), and a meat-based soup. No "blended plant-based broth" appears.

plausibility—broth and rice are linguistically compatible. However, the generated image grounded the model in actual Vietnamese cuisine, where rice is paired with richly flavored meat dishes and stir-fries, not blended broths. The visual cue made the cultural mismatch salient, enabling self-correction.

2 Case 2: Indian Wedding Traditions (India)

True-or-False Task

Statement: *"Indians typically prefer **lavish and large weddings**."*

Gold Label: **True**

Base LLM → **True** (✓ Correct)

V-CUE → **True** (✓ Correct, UD-gated)

Explanation: Indian weddings are widely known for their grand scale, multi-day celebrations, and elaborate rituals.

V-CUE Pipeline Walkthrough:

The UD module evaluates the base LLM's judgment (True).

Evaluation Criteria:

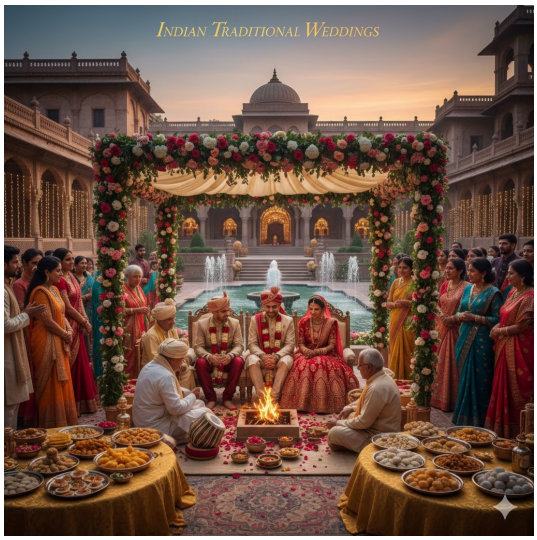
1. *Cultural correctness*: "Lavish and large weddings" is strongly aligned with widely documented Indian wedding traditions.
2. *Cultural specificity*: The answer reflects a specific, well-known cultural characteristic.
3. *Confidence*: The base model's judgment is decisive and internally consistent.

UD Output: `True` — The judgment is reliable.

⇒ *No visual enhancement needed. Return base judgment directly.*

Although UD gates this instance (saving inference cost), we show what VG would produce for completeness:

(a) Stable Diffusion output



(b) Gemini-3 Pro-Image output

Figure 2: Generated images for Case 2 (shown for illustration). Both depict a grand Indian wedding: a *mandap* canopy with floral arches, a sacred fire (*Agni*) ceremony, many guests in colorful saris and sherwanis, and lavish decorations—all consistent with the "lavish and large" characterization.

Analysis: This case demonstrates two strengths of V-CUE. First, the UD module correctly identifies a reliable cultural judgment and skips visual enhancement, reducing token cost by

$\sim$46% and latency by $\sim$43% (as quantified in Section 6.7). Second, even if VG were triggered, the generated images would further confirm the statement's correctness—the visual evidence of grand venues, hundreds of guests, and multi-ritual ceremonies is fully consistent with "lavish and large weddings."

3 Case 3: Korean Pre-Wedding Customs (South Korea)

True-or-False Task

Statement: *"In Korean culture, the first step a couple takes before deciding the date and location for their wedding is* ***throwing a pre-wedding party for friends****."*

Gold Label: **False**
Base LLM → **True** (✗ Incorrect)
V-CUE → **False** (✓ Correct)

Explanation: The actual first step is *Sanggyeonrye*, a formal meeting between both families to discuss and approve the marriage, not an informal party.

V-CUE Pipeline Walkthrough:

The UD module evaluates the base LLM's judgment (True).

Evaluation Criteria:

1. *Cultural correctness*: "Pre-wedding party for friends" reflects a Western engagement practice, not a Korean tradition.
2. *Cultural specificity*: The answer lacks any Korea-specific cultural markers (e.g., family involvement, formality).
3. *Confidence*: The vague phrasing suggests limited grounding in Korean wedding customs.

UD Output: `False` — The judgment is unreliable.
⇒ *Trigger Visual Generation module.*

Extracted cues from the question:

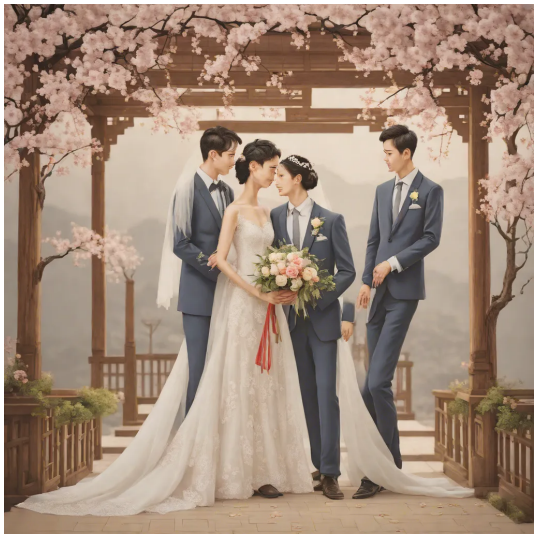
```
{"region": "South Korea", "object": "wedding preparation, couple",
 "symbol": "pre-wedding cultural tradition"}
```

Condition unit: c = `"South Korea, wedding preparation couple, pre-wedding cultural tradition"`

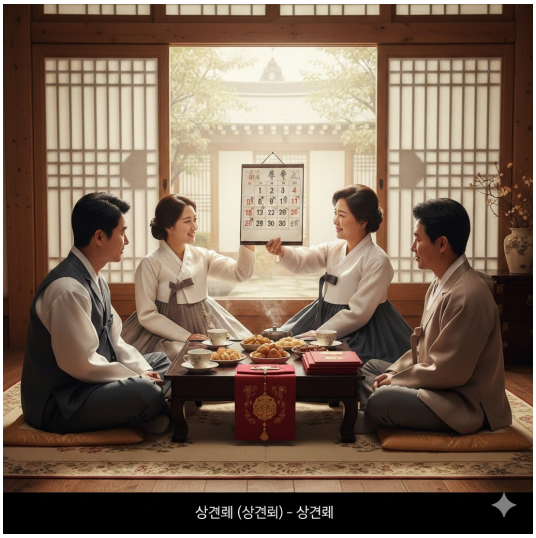
The generated images present a scene strikingly different from "throwing a pre-wedding party for friends": they show **two families** (not friends) in **formal attire** (*hanbok*, not casual wear) in a **traditional Korean room** (not a party venue), engaged in a **structured, seated meeting** (not a social gathering). This strong visual contrast exposes the cultural mismatch—Korean pre-wedding tradition centers on formal family approval, not informal friend celebrations. The model thus judges the statement as **False**.

V-CUE Output: **False** (✓ Correct)

Analysis: This is a clear example of Western-centric bias. The base LLM projected Western engagement customs (informal parties with friends) onto Korean culture, where the first pre-wedding step is *Sanggyeonrye*—a deeply formal, family-centered ritual. The visual cues provided decisive evidence of this cultural difference: the contrast between "friends at a party" and "families in hanbok at a formal table" enabled the model to reject the false claim.



(a) Stable Diffusion output



(b) Gemini-3 Pro-Image output

Figure 3: Generated images for Case 3. Both depict *Sanggyeonrye*: two families in traditional Korean *hanbok* seated formally around a low table with tea, traditional sweets, and ceremonial items in a traditional Korean room with paper-screen doors. Notably, this is a **family gathering**, not a "party for friends."

4 Case 4: Nowruz Celebration Date (Iran)

True-or-False Task

Statement: *"Families in Iran gather to celebrate Nowruz on **April 1st**."*

Gold Label: **False**

Base LLM → **True** (✗ Incorrect)

V-CUE → **False** (✓ Correct)

Explanation: Nowruz is celebrated on the first day of spring—the vernal equinox (around March 20–21), not April 1st. The base LLM likely confused Nowruz with *Sizdah Bedar* (the 13th day of Nowruz, which falls around April 1–2).

V-CUE Pipeline Walkthrough:

The UD module evaluates the base LLM's judgment (True).

Evaluation Criteria:

1. *Cultural correctness*: "April 1st" does not correspond to the established date of Nowruz, which is tied to the astronomical spring equinox.
2. *Cultural specificity*: A specific date was provided, but it is factually incorrect.
3. *Confidence*: The model's acceptance of a precise but wrong date suggests shallow knowledge about Nowruz timing.

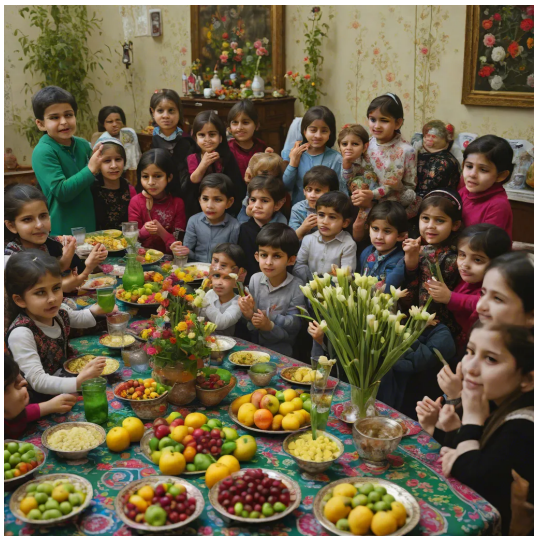
UD Output: `False` — The judgment is unreliable.

⇒ *Trigger Visual Generation module.*

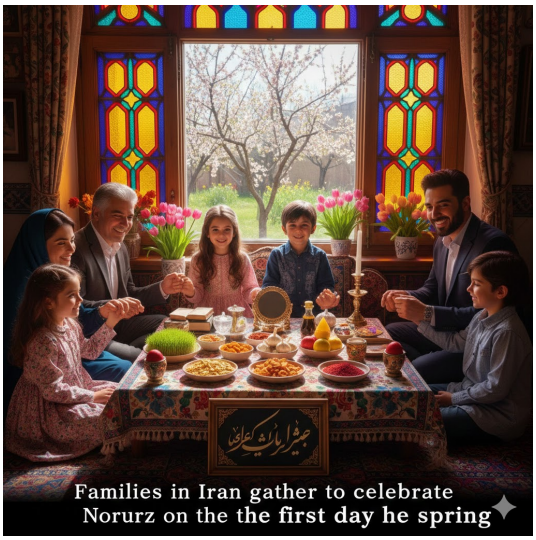
Extracted cues from the question:

```
{"region": "Iran", "object": "family gathering, Nowruz celebration",
                "symbol": "Nowruz, spring festival"}
```

Condition unit: c = `"Iran, family gathering Nowruz celebration, Nowruz spring festival"`



(a) Stable Diffusion output



(b) Gemini-3 Pro-Image output

Figure 4: Generated images for Case 4. Both depict a family gathered around the *Haft-sin* table—the centerpiece of Nowruz celebrations. Key spring equinox symbols are visible: sprouted wheat grass (*sabzeh*), blooming cherry trees through the window, fresh tulips, a mirror, colored eggs, and traditional stained-glass windows. All elements signify the **arrival of spring**, not early April.

The generated images are saturated with spring equinox symbolism: the *Haft-sin* table (seven items beginning with the Persian letter "S," each symbolizing a spring-related concept), sprouted wheat grass (*sabzeh*, symbolizing rebirth), fresh blossoms visible through the windows, and bright tulips. These visual elements unmistakably signal the **beginning of spring**—around March 20–21—rather than April 1st. The strong seasonal cues contradict "April 1st" and enable the model to judge the statement as **False**.

V-CUE Output: False (✓ Correct)

Analysis: The base LLM likely conflated Nowruz (March 20–21) with *Sizdah Bedar* (April 1–2), the 13th day of Nowruz festivities when families picnic outdoors. This is a subtle but significant cultural error—the model "knew" Nowruz involved spring, but accepted a plausible nearby date. The visual cues, rich with early-spring imagery (blooming trees, fresh sprouts), provided the seasonal grounding needed to reject "April 1st" as the celebration date.