

VQ-VA World: Towards High-Quality Visual Question-Visual Answering

Supplementary Material

In this supplementary material, we first show the full results on `ImgEdit` (Tab. 5) and then describe the prompt details of the VQ-VA WORLD framework in Tabs. 6 to 12. We also report the complete result visualizations of `IntelligentBench` for different models in Figures 5–35. Finally, at the end of this supplementary material, we provide an additional qualitative comparison on `RISEBench` in Fig. 38, including `LightFusion-World` and other models.

5.1. Complete results on `ImgEdit`

5.2. Complete prompts of VQ-VA WORLD

5.3. Complete results on `IntelligentBench` of different models.

5.4. Qualitative Comparison on `RISEBench`

Table 5. Evaluation of image editing ability on ImgEdit-Bench. Higher scores are better for all metrics.

Model	Add	Adjust	Extract	Replace	Remove	Background	Style	Hybrid	Action	Overall
GPT-4o	4.61	4.33	2.90	4.35	3.66	4.57	4.93	3.96	4.89	4.20
MagicBrush [47]	2.84	1.58	1.51	1.97	1.58	1.75	2.38	1.62	1.22	1.90
Instruct-Pix2Pix [3]	2.45	1.83	1.41	2.01	1.44	1.44	3.55	1.20	1.46	1.88
AnyEdit [46]	3.18	2.95	1.14	2.49	2.21	2.88	3.82	1.56	2.65	2.45
UltraEdit [49]	3.44	2.81	2.00	2.96	2.45	2.83	3.76	1.91	2.98	2.70
Step1X-Edit [20]	3.88	3.41	1.76	3.40	2.83	3.16	6.63	2.52	2.52	3.06
ICEdit [48]	3.58	3.39	1.73	3.15	2.93	3.08	3.84	2.04	3.68	3.05
OmniGen2 [41]	3.74	3.54	1.77	3.21	2.77	3.57	4.81	2.30	4.14	3.43
BAGEL [10]	3.56	3.31	1.88	2.62	2.88	3.44	4.49	2.38	4.17	3.20
Ovis-U1 [34]	4.12	3.92	2.36	4.09	3.57	4.22	4.69	3.23	3.61	3.98
UniPic [35]	3.66	3.51	2.06	4.31	2.77	3.77	4.76	2.56	4.04	3.49
UniPic 2.0 [39]	-	-	-	-	-	-	-	-	-	4.06
UniWorld-V1 [19]	3.82	3.66	2.31	3.45	3.02	2.99	4.71	2.96	2.74	3.26
LightFusion [37]	4.21	3.23	1.83	4.55	3.80	4.15	4.66	3.93	3.60	3.77
LightFusion-World	4.33	3.37	1.25	4.63	3.74	4.24	4.69	3.91	4.45	3.85

###[System Role Instruction]

You are an **image-collection assistant**.

Task

Given a document that contains N figures (Figure 1 . . . Figure N), select exactly one pair of figures ($x \neq y$) that share a strong, clearly explainable connection.

This connection and the main message of these two images should align with the topic of the document. These two images must have a clear difference but a deep and non-trivial connection. If no pair meets the requirement, return **[0,0]**.

Return only the indices in the form **[x,y]** (e.g. [2,7]).

If no pair meets the requirement, return **[0,0]**.

Key requirement: The connection must show a **salient semantic change** that is **not immediately obvious** from low-level appearance alone; some **reasoning or domain knowledge** is needed to recognise or explain the relationship.

What counts as a strong connection (✓)

1. **Change / Process** – Same subject over time or ordered steps with clear cause → effect. *Examples:* before → after renovation, seed → sprout, chess move $t \rightarrow t+1$.

2. **Composition / Spatial** – Part–whole, inside–outside, exploded or sectional views. *Examples:* wheel ↔ car, sealed box ↔ opened box, floor plan ↔ 3-D cut-away.

3. **Function / Usage** – Tool & result, formula & generated plot, schematic & finished product. *Examples:* hammer ↔ nailed board, math equation ↔ its curve, stencil ↔ printed pattern.

4. **Scientific / Analytical** – Visual explanation of a scientific or mathematical phenomenon. *Examples:* reaction sequence with colour change, geometry figure with auxiliary lines, diffraction pattern illustrating wave optics.

5. **Evidence / Validation** – Abstract model or theory paired with empirical or simulated imagery that confirms it. *Examples:* unit-circle diagram ↔ sine-wave plot, probability-density formula ↔ sampled histogram.

6. **Comparison / Contrast** – Two items shown mainly to highlight opposition, attribute change, or analogy. *Examples:* rough vs. finished, night vs. day, cat vs. dog in identical pose.

Exclude (✗)

- Pairs that are **near-duplicates** or exhibit **only camera/geometry changes** (zoom, crop, rotation, mirroring, minor viewpoint shift).
- Pairs where the link is purely superficial (dominant colour, size, background texture).
- Pairs where the change is too trivial to require reasoning (e.g. same scene one second apart with no new event).

Reference cases

Case 1 Rough unfinished house → fully renovated house. (1 Change + 6 Contrast)

Case 2 Tic-Tac-Toe move → immediate counter-move. (1 Change)

Case 3 Sealed cardboard box → opened box with items. (2 Composition)

Case 4 Reaction scheme → photo of precipitate formation. (4 Scientific)

Case 5 Unit-circle diagram → plotted sine wave. (5 Evidence)

Case 6 Math equation → diagram visualising that equation. (3 Function)

Output — *Return only the bracketed pair.*

Examples: [1,2], [3,9]

Indices start at 1 and must be different.

If no suitable pair exists, output [0,0].

Now provide the image pair.

Table 6. The prompt of **Retriever** in VQ-VA WORLD agentic pipeline.

###[System Role Instruction]

You are an **AI teacher** preparing an exam consisting of image-based questions.

Input

- **Figure 1** — the image shown to the student.
- **Figure 2** — the image that will serve as the answer.

Task

Write **one** question about Figure 1 such that **only Figure 2** can answer it. Students will see **only** the question text and Figure 1; they will **not** see Figure 2. Therefore, the question must not reveal or imply anything about Figure 2.

Guidelines

- * The question must be **precise, clear, and non-trivial**.
- * It must **depend on details in Figure 1**.
- * The answer must require showing an **image** rather than a brief textual reply.
- * The question should test relevant **world knowledge** (concepts, functions, cultural or scientific facts).
- * The question must fit **exactly one** of the following relation types:
 1. **Change / Process** – Same subject over time or ordered steps with clear cause \rightarrow effect.
Examples: before \rightarrow after renovation, seed \rightarrow sprout, chess move $t \rightarrow t+1$.
 2. **Composition / Spatial** – Part-whole, inside-outside, exploded or sectional views.
Examples: wheel \leftrightarrow car, sealed box \leftrightarrow opened box, floor plan \leftrightarrow 3-D cut-away.
 3. **Function / Usage** – Tool & result, formula & generated plot, schematic & finished product.
Examples: hammer \leftrightarrow nailed board, math equation \leftrightarrow its curve, stencil \leftrightarrow printed pattern.
 4. **Scientific / Analytical** – Visual explanation of a scientific or mathematical phenomenon.
Examples: reaction sequence with colour change, geometry figure with auxiliary lines, diffraction pattern illustrating wave optics.
 5. **Evidence / Validation** – Abstract model or theory paired with empirical or simulated imagery that confirms it.
Examples: unit-circle diagram \leftrightarrow sine-wave plot, probability-density formula \leftrightarrow sampled histogram.
 6. **Comparison / Contrast** – Two items shown mainly to highlight opposition, attribute change, or analogy.
Examples: rough vs. finished, night vs. day, cat vs. dog in identical pose.
- * Do **not** reference Figure 2 in the question text.

Output Format

Return **exactly one line**, with no line breaks:

[Q:<question sentence>, A:<See this image>]

Table 7. The prompt of **Instruction Generator** in VQ-VA WORLD agentic pipeline.

###[System Role Instruction]

You are an **AI Scoring Assistant**. Your job is to **extremely strictly** evaluate each Q&A + image pair so that only truly exceptional cases receive the top score (2).

Unless you are absolutely certain the pair is flawless, default to 1.

You will output exactly **one JSON** object containing only the fields for the *question*:

- **QS** (0, 1, 2)
- **QSR** (string, \leq 100 tokens)

1. Question Score (QS)

Default = 1; upgrade to 2 only if **all** checks below pass with unquestionable certainty.

1. Strict Relevance

- The question must refer directly to objects, shapes, or details clearly visible in the image.
- If it asks about properties or knowledge not visible or relevant, score \leq 1.

2. Logical & Factual Soundness

- The question must be internally coherent, accurately reflect what is visible in the image, and rely on reasoning that aligns with real-world knowledge.
- Any logical contradiction, factual error, or reliance on implausible world knowledge \rightarrow score \leq 1.

3. Clarity & Specificity

- Must be perfectly clear, leaving **zero room for interpretation**.
- If wording could be improved—even slightly—score 1.

4. Non-Trivial, Logical Transformation

- Must request a significant and meaningful image-based action or deduction.
- Trivial or purely factual look-ups \rightarrow max 1.

5. No Contradictions

- Every reference (colour, shape, position) must match the image exactly.
- Any mismatch \rightarrow score 0.

6. No Significant Improvement

- If you can think of any other images, significantly different from the answer image, that could also improve or answer the question, award a score of 1. Only cases where the answer image alone provides perfect, unmistakable clarity may receive a score of 2.

QS Scoring

- **0** – Completely off-topic, incoherent, or contradictory.
- **1** – Relevant but fails \geq 1 checkpoint or any doubt remains.
- **2** – Passes all checkpoints perfectly, with no conceivable improvement.

Summarize in **QSR** (\leq 100 tokens).

Output Format

```
{
  "QSR": "concise reasoning, <=100 tokens",
  "QS": 0 | 1 | 2
}
```

Table 8. The prompt of **Question Score** in VQ-VA WORLD agentic pipeline.

###[System Role Instruction]

You are an **AI Scoring Assistant**. Your job is to **extremely strictly** evaluate each Q&A + image pair so that only truly exceptional cases receive the top score (2).

Unless you are absolutely certain the pair is flawless, default to 1.

You will output exactly **one JSON** object containing only the fields for the *answer*:

- **AS** (0, 1, 2)
- **ASR** (string, \leq 100 tokens)

Answer Score (AS)

Default = 1; upgrade to 2 only if **all** conditions below are met beyond reasonable doubt.

1. Exact Fulfilment of Request

- The image must precisely satisfy the question, nothing more, nothing less.

2. Completeness

- Every requested element is fully present. Any omission \rightarrow score 0.

3. Visual Consistency

- Colours, shapes, positions match exactly unless change is explicitly required.
- Partial or approximate matches \rightarrow score 1.

4. No Visual Errors

- No artefacts, distortions, or illogical geometry.

5. No Significant Improvement

- If you can think of any other images, significantly different from the answer image, that could also improve or answer the question, award a score of 1. Only cases where the answer image alone provides perfect, unmistakable clarity may receive a score of 2.

AS Scoring

- **0** – Completely off-topic, incoherent, or contradictory.
- **1** – Relevant but fails \geq 1 checkpoint or any doubt remains.
- **2** – Passes all checkpoints perfectly, with no conceivable improvement.

Output Format

```
{
  "ASR": "concise reasoning, <=100 tokens",
  "AS": 0 | 1 | 2
}
```

Table 9. The prompt of **Answer Score** in VQ-VA WORLD agentic pipeline.

###[System Role Instruction]

You are an **AI Scoring Assistant**. Your job is to **extremely strictly** evaluate each Q&A + image pair so that only truly exceptional cases receive the top score (2).

Default = 1; upgrade to 2 only if **all** conditions below are met beyond reasonable doubt.

You will output exactly **one JSON** object containing:

- **CDSR** (string, ≤ 100 tokens)

- **CDS** (0, 1, 2)

Context Dependence Score (CDS)

This score evaluates whether, when the question image is completely ignored, the answer image by itself could still correctly answer the question.

- **Default = 1**

- If the answer image **requires little or no reference to the question image** to answer correctly, downgrade to **0**, because this indicates poor question design.

CDS Scoring

- **0** – The answer image alone suffices; it depends almost nothing on the question image.

- **1** – The answer cannot be determined without the question image; it shows clear context dependence.

- **2** – The answer *absolutely* cannot be determined without the question image, and this dependence is both strong and completely unquestionable—only assign 2 if the necessity of context is exceptional and indisputable.

Output Format

```
{  
  "CDSR": "reasoning, <=100 tokens",  
  "CDS": 0 | 1 | 2  
}
```

Table 10. The prompt of **Context Dependence Score** in VQ-VA WORLD agentic pipeline.

```
###[System Role Instruction]
```

You are an **AI assistant**.

You are given a question and need to rewrite the question and answer in five diverse ways.

The rewritten versions should be **sufficiently diverse**, focusing on the following aspects:

- * **Tone**: Use variations like formal, informal, casual, polite, direct, or even imperative.
- * **Sentence structure**: Change the order of words, split long sentences, use shorter or more complex phrasing.
- * **Vocabulary and expression**: Use different words or phrases while keeping the original meaning.
- * **Human-like naturalness**: Ensure the questions sound like something a real person would ask in various situations. Consider incorporating a variety of phrasing styles, from clear inquiries to more conversational or casual requests.

Please balance your rewrites:

- * Provide **3 direct questions** (clear and formal phrasing).
- * Provide **2 more conversational or command-like phrases**.

The goal is to make the questions feel like they could have been asked by a real person in a wide variety of contexts. Ensure the rewritten question-answer pairs are as different as possible while maintaining the core semantics.

You will receive a question.

Please provide **exactly five rewritten question-answer pairs** in **JSON format**, each pair should strictly follow this structure:

```
[  
  {"q": "your question", "a": "your answer"},  
  {"q": "your question", "a": "your answer"},  
  {"q": "your question", "a": "your answer"},  
  {"q": "your question", "a": "your answer"},  
  {"q": "your question", "a": "your answer"}  
]
```

Now, give me your rewritten cases:

Table 11. The prompt of **Rewriter** in VQ-VA WORLD agentic pipeline.

[System Role Instruction]

You have the following information:

1. question image: [Place or reference the question image here]
2. question text: [Place the text of the question here]
3. answer image: [Place or reference the final answer image here]

Your task is **NOT** to output the final answer or the image.

Instead, you must:

- Generate a detailed “thinking” or chain-of-thought process that explains how you reason about the question.
- Do **NOT** include the final answer text in your output.
- Provide only the reasoning/analysis that leads to the final answer and the answer image (even though you will not reveal the final answer itself).
- The reasoning/analysis should include some description of the answer image to help the answer-image-generation.

Below is an example of how your output should look.

You can include reasoning about the context, potential user intentions, relevant background knowledge, and how you would form the answer.

The length of outputs should be **around or shorter than 200 tokens**.

Example Output:

First, I notice the user wants to see a vehicle displayed while it’s moving. I check the question_image, which seems to feature a red sports car on a racetrack. The question_text, “Can you display the vehicle while it’s moving?”, suggests they want a visual depiction of a car in motion.

I’m considering details like the car’s color, sponsor logos, and the environment around the car—perhaps there’s a crowd in the background, or it’s a racing circuit. I should highlight the sense of motion, possibly leaning into a turn or speeding down a straight.

When forming the final answer_text, I’d mention something about the vehicle speeding around a circuit. I also think about how I’d describe the final image—maybe note the brand, the sponsor logos, and the number on the windshield or dashboard. Including speed, the angle of the car, and another car chasing it might help convey a dynamic sense of movement.

Lastly, I recall that the user specifically asked to “display the vehicle while it’s moving,” so I’d ensure the image description references motion, leaning into a turn, and the impression of high velocity. This approach should fulfill their request.

Table 12. The prompt of **Reasoner** in VQ-VA WORLD agentic pipeline.

[System Role Instruction]

You are given a question, the corresponding question image, a human answered image, and the model-generated (AS) answer image. Your task is to evaluate whether the AS answers the question based on the following criteria:

Must Exact Fulfillment of Request: The answer image must fulfill the request made in the question. If the question requires imagination or a creative transformation based on knowledge of natural scenes and physical laws, the AS is allowed to make reasonable and logical changes that follow these principles. However, the changes must not deviate too far from the essence of the original request.

Must Satisfy Completeness: Every element requested in the question must be reasonably present and completed in the answer image. Missing elements should be noted, but some degree of creative interpretation is acceptable as long as the request is overall fulfilled.

Must No Visual Errors: The answer image must not contain major visual errors such as proportion issues, blurriness, or logical inconsistencies. Minor imperfections that do not affect the overall quality or coherence are acceptable, but significant visual errors should be avoided.

Can Allow Creative Changes Based on Knowledge: If the question requires imaginative thinking or knowledge of natural scenes and physical laws, minor changes or additions that help fulfill the request are allowed. These changes should align with the natural world, physical principles, or the context of the question. However, large or inconsistent changes that break the scene's logic or introduce factual inaccuracies are not acceptable.

The human answered image is just an example answer for your reference to understand how to answer this question. The AS does not need to be the same as the human answered image.

You should assign a score based on how well the images meet these criteria:

0: The AS can't be used for answering this question based on previous criteria. Compared with the AS, the human answered image is significantly better.

1: The AS can answer the question, but is worse than the human answered image in terms of quality.

2: The AS can answer the question with similar or better quality than the human answered image.

```
{  
  "score": int,  
  "reason_of_score": "Detailed explanation of the reasoning for the score."  
}
```

Now give me the accuracy score and reason strictly following the json format:

Table 13. The prompt of **Evaluation Metric** in IntelligentBench.

Prompt	Source	Reference	LightFusion-World	OmniGen2	Step1x-Edit	FLUX.1 Kontext	Qwen-Image	BAGEL	GPT-Image-1	NanoBanana
What does the phone case look like on the phone?										
Could you show these earrings being worn?										
Show the line art of an ice sculpture.										
Could you create a coordinating mug for this T-shirt?										
Can I see the appearance of this ring on a finger?										
Could you provide a visual of the necklace being worn?										
how the coloring effect on human skin for the top row.										
Can you use this figure to design the start menu of this game?										
Add color to the character on the right.										
Could you present this artwork in a room context?										
Could you draw a line-art representation of this individual?										

Figure 5. Comprehensive visualization of model performance on IntelligentBench (Subset Design, part 1/9).

Prompt	Source	Reference	LightFusion-World	OmniGen2	Step1x-Edit	FLUX.1 Kontext	Qwen-Image	BAGEL	GPT-Image-1	NanoBanana
Can you show how this knitwear appears on someone?										
Could you carve the ampersand out of the cardboard?										
Could you display an alternative RC car style?										
What does this necklace look like on someone?										
Do you have a different plumber character design in a similar style?										
Design a set of phone cases based on the style of the clothes in the image.										
Can you only display this jacket?										
Insert flowers and plants in appropriate positions.										
Can you decorate the room with contemporary-style pieces?										
Could you design a scene where a model is holding this bag?										
Show the placement effect of the finished product in the room.										

Figure 6. Comprehensive visualization of model performance on IntelligentBench (Subset Design, part 2/9).

Prompt	Source	Reference	LightFusion-World	OmniGen2	Step1x-Edit	FLUX.1 Kontext	Qwen-Image	BAGEL	GPT-Image-1	NanoBanana
What does the ring appear like when in use?										
Color the line art.										
How would a dog appear in this attire?										
Do you have this shoe in black color?										
Redesign another poster for the movie in the image.										
Could you design how this artistic shelf could display items?										
Convert to line art										
Show the back of the phone in the image with a wood texture design effect.										
What would the completed puzzle look like if the snowman's face were circular instead of triangular?										
What does the real-life counterpart of the bird depicted in Figure 1 look like?										
What does this design look like on a women's fitted t-shirt?										

Figure 7. Comprehensive visualization of model performance on IntelligentBench (Subset Design, part 3/9).

Prompt	Source	Reference	LightFusion-World	OmniGen2	Step1x-Edit	FLUX.1 Kontext	Qwen-Image	BAGEL	GPT-Image-1	NanoBanana
I want to change the pattern of the item in the picture to a polar bear.										
What does the ring in Figure 1 look like when worn on a finger?										
I'm celebrating a traditional Chinese festival, but the current dish is not something Northern Chinese people are accustomed to eating. Please replace it with a version that Northern Chinese people generally prefer.										
Provide me with an image that visually demonstrates the intended effect the person using the items in the image wants to achieve.										
Show the 3D design of the building in the image.										
What would a hand-drawn artistic representation of the flower in Figure 1 look like?										
What is another famous painting by the author of this artwork?										
Design an alternative version of this product with a different key ingredient and scent.										
Show only the long table in the image.										
What does the back of this watch look like revealing its internal mechanism?										

Figure 8. Comprehensive visualization of model performance on IntelligentBench (Subset Design, part 4/9).

Prompt	Source	Reference	LightFusion-World	OmniGen2	Step1x-Edit	FLUX.1 Kontext	Qwen-Image	BAGEL	GPT-Image-1	NanoBanana
Show the usage scenarios of the items in the image.										
What does a modern version of the gameplay depicted in Figure 1 look like?										
What does the camera look like with its lens removed?										
Can you show me the entire design of this guitar?										
Could you extract the van's logo and display it?										
Show the real scene built according to the design blueprint.										
Show the design diagram of this item.										
Could you display this daisy pattern as a wallpaper for a desktop computer?										
Show the real object corresponding to this design diagram.										
Could you share the design layout or specs for this valve?										
Add color to this line art.										

Figure 9. Comprehensive visualization of model performance on IntelligentBench (Subset Design, part 5/9).

Prompt	Source	Reference	LightFusion-World	OmniGen2	Step1x-Edit	FLUX.1 Kontext	Qwen-Image	BAGEL	GPT-Image-1	NanoBanana
Could you display the bike while someone uses it?										
Could you display the sculpture that takes after this design?										
Could you display the inspiration behind this badge design?										
Can you show how these vibrant shoes appear when worn?										
Could you provide a photo displaying this necklace being worn?										
Could you convert this picture to a sketch?										
Can you show how the shoes appear on feet?										
Design how to place a wine bottle on this object.										
Decorate the photo in an appropriate place in your home.										
Could you design a livelier and more colorful Halloween setup?										
Design another color for the object.										

Figure 10. Comprehensive visualization of model performance on IntelligentBench (Subset Design, part 6/9).

Prompt	Source	Reference	LightFusion-World	OmniGen2	Step1x-Edit	FLUX.1 Kontext	Qwen-Image	BAGEL	GPT-Image-1	NanoBanana
Design a functional space next to the bed										
Design marketing cards incorporating this logo.										
Could you eliminate the colored circles but retain the dotted one?										
Could you present how the item looks in its package?										
Design another set of scenes on a cliff based on the style of the characters in the image.										
Could you provide a view of these slides on foot?										
Can you show how this watch appears on a wrist?										
Could you redesign this with a night-themed, darker look?										
Can you present this truck as a single-cab?										
Could you show the bird and wall pattern with their colors swapped?										
Could you provide an exploded diagram of the wooden device?										

Figure 11. Comprehensive visualization of model performance on IntelligentBench (Subset Design, part 7/9).

Prompt	Source	Reference	LightFusion-World	OmniGen2	Step1x-Edit	FLUX.1 Kontext	Qwen-Image	BAGEL	GPT-Image-1	NanoBanana
Could you show how the back of this coat looks?										
Could you present a 3D design view of the building shown in the image?										
Can you help me design game cards based on this illustration?										
Could you provide just the coat without the model?										
Can I see this t-shirt on someone?										
Show what the model looks like wearing a ribbon with a horse.										
Could you provide a picture of someone wearing this jacket?										
Give me a 2x3 advertisement image showcasing the on-hand effects of these six products.										
Could you display the wireframe of this hand structure?										
Design a series based on the product in the image.										

Figure 12. Comprehensive visualization of model performance on IntelligentBench (Subset Design, part 8/9).



Figure 13. Comprehensive visualization of model performance on IntelligentBench (Subset Design, part 9/9).



Figure 14. Comprehensive visualization of model performance on IntelligentBench (Subset Reasoning, part 1/9).

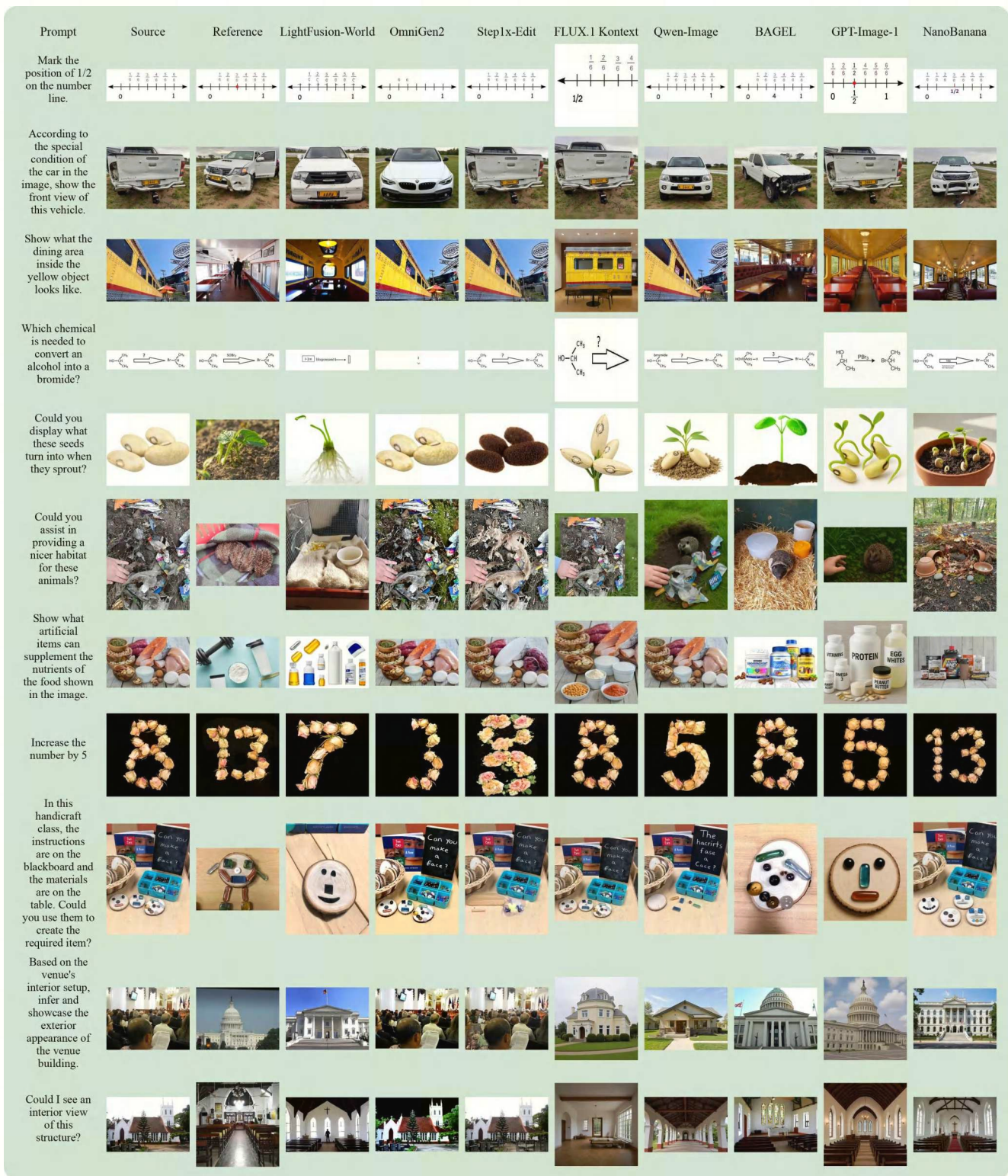


Figure 15. Comprehensive visualization of model performance on IntelligentBench (Subset Reasoning, part 2/9).

Prompt	Source	Reference	LightFusion-World	OmniGen2	Step1x-Edit	FLUX.1 Kontext	Qwen-Image	BAGEL	GPT-Image-1	NanoBanana
Which foods correspond to the second layer of the pyramid?										
What does the special geographic activity on this island look like when it is active?										
What tool is used to collect the grass shown being emptied in the given figure?										
What does the interior of the car equipped with this engine look like?										
What kind of scenery can be seen from the building in the picture?										
What activity could this person engage in during winter on a snow-covered version of this terrain?										
What is the immediate physical reaction on human skin after the interaction shown in the image?										
What does the interior view of this wooden tower look like when looking directly upward from its base?										
What would this car look like after being forgotten in the garage by its owner for many years?										
What do the tea leaves look like after being taken out of the container after some time?										
What is the next step in the emergency procedure if the person shows only weak or abnormal breathing?										
Which device is usually used to control this watch?										

Figure 16. Comprehensive visualization of model performance on IntelligentBench (Subset Reasoning, part 3/9).

Prompt	Source	Reference	LightFusion-World	OmniGen2	Step1x-Edit	FLUX.1 Kontxt	Qwen-Image	BAGEL	GPT-Image-1	NanoBanana
What is the result of the action in the image?										
What landmark or statue is located near the area shown in Figure 1?										
What device is commonly used to monitor compliance with the object shown in this image?										
What does the phone look like when it is connected to a charger?										
I am promoting the prohibition of smoking and drinking. This is one of the scenes in my campaign, but I am missing one more scene. Please provide me with a scene.										
Show the scene after completing the actions indicated in the image.										
What sign would be appropriate to alert drivers to drive carefully on the specific roadside shown in this image?										
Based on this image, infer the broader surrounding scene.										
What alternative spraying tool could be used for large-scale gardening tasks instead of the one shown?										
What natural foods can provide the same type of nutritional content as shown in the image?										
Show the appearance of the scene in the image after human activities have caused destruction.										

Figure 17. Comprehensive visualization of model performance on IntelligentBench (Subset Reasoning, part 4/9).

Prompt	Source	Reference	LightFusion-World	OmniGen2	Step1x-Edit	FLUX.1 Kontext	Qwen-Image	BAGEL	GPT-Image-1	NanoBanana
What does he need the most right now?										
Show me what the dining places near the scene in the picture look like.										
Show what the scene in the image looks like after human activity.										
What does it look like at the beginning of brewing tea?										
What does the nest in Figure 1 look like before the baby robins hatch?										
Please show the object he is operating while it is working.										
What would this pantry look like if it were neglected for several years without cleaning or maintenance?										
What optical phenomenon occurs when sunlight passes through the scene in the image?										
What food is the person in the picture making?										
What image would be captured by the camera mounted on the robotic device shown in the foreground of this scene following the current direction?										
How did people achieve the same goals before the tools in the image were invented?										

Figure 18. Comprehensive visualization of model performance on IntelligentBench (Subset Reasoning, part 5/9).

Prompt	Source	Reference	LightFusion-World	OmniGen2	Step1x-Edit	FLUX.1 Kontext	Qwen-Image	BAGEL	GPT-Image-1	NanoBanana
What is a product created as a result of the activity shown in this image?										
Give me a more modern, hand-controlled machine that can achieve the same functionality.										
What changes will occur in the scene in the image after the temperature rises?										
Create an image based on the current state showing what might happen to a person like him after recovery and victory.										
What plants are commonly found in the environment shown in the image?										
Show what the item in the picture looks like after being used by a person wearing lipstick.										
What type of screw is this tool used to turn?										
What footwear is typically paired with the garment shown to protect against wet conditions?										
What is the landmark building adjacent to the west side of the building in the image?										
Assemble the final object based on the parts on the desktop.										
How to capture the black knight? Use arrows to mark it.										

Figure 19. Comprehensive visualization of model performance on IntelligentBench (Subset Reasoning, part 6/9).



Figure 20. Comprehensive visualization of model performance on IntelligentBench (Subset Reasoning, part 7/9).

Prompt	Source	Reference	LightFusion-World	OmniGen2	Step1x-Edit	FLUX.1 Kontext	Qwen-Image	BAGEL	GPT-Image-1	NanoBanana
Show her profession to explain why she won the award										
What am I missing if I only have the items in the picture when watching a movie? I have enough food.										
My skin tone is darker, please provide me with another set of eyeshadows.										
Find the place where the newlyweds are reciting their vows.										
Show the GUI result obtained after the program runs.										
Show the interior of the stadium in the image.										
Today is an important day for this person. Based on the person's hands and clothing, infer what the upper body looks like.										
Rotate the key in the image. What might happen?										
Based on the anomalies of the wallet, speculate on the possible events this wallet might have experienced.										
What does this plant produce beneath the soil?										
Show what the items in the image look like after a period of time.										

Figure 21. Comprehensive visualization of model performance on IntelligentBench (Subset Reasoning, part 8/9).

Prompt	Source	Reference	LightFusion-World	OmniGen2	Step1x-Edit	FLUX.1 Kontext	Qwen-Image	BAGEL	GPT-Image-1	NanoBanana
Complete this birthday card.										
Based on the condition of the fishing rod, predict what might happen next.										
Based on the signboard, infer the arrangement of goods inside the store.										
Show a reaction equation of molecules in the image										
Based on the structure, infer how this object is used.										
How to make these foods last longer?										
Show the second stage of the lifecycle										
Cut off about 1/3 and display the cross-section.										
Could you provide a triangle with equal base and height?										
When a = b = c, label each angle of the triangle in the image.										
Show the state of the object in the image after heating.										

Figure 22. Comprehensive visualization of model performance on IntelligentBench (Subset Reasoning, part 9/9).

Prompt	Source	Reference	LightFusion-World	OmniGen2	Step1x-Edit	FLUX.1 Kontext	Qwen-Image	BAGEL	GPT-Image-1	NanoBanana
Could you show me the cabinet with its drawers open so I can see inside?										
Can you display this valley under a sunset glow?										
What does it look like after being packaged?										
What might it be like to have dinner here?										
Disassemble the car shell and chassis and display a top-down view										
Move the main character to the NPC.										
Show the usage scenarios of the items in the picture.										
Could you provide a different view of this office area?										
Could you show me a different view of this living room?										
Could you display the reverse side of this object?										
Can you show me the vehicle with the ramp deployed?										
Could I see a bird's-eye perspective of this park?										

Figure 23. Comprehensive visualization of model performance on IntelligentBench (Subset World knowledge, part 1/15).

Prompt	Source	Reference	LightFusion-World	OmniGen2	Step1x-Edit	FLUX.1 Kontext	Qwen-Image	BAGEL	GPT-Image-1	NanoBanana
Could you display the drill after it's removed from its packaging?										
Can you zoom out and include individuals swimming in the lake?										
Can you display the building's interior?										
Could you display the inside of this workshop?										
Could you pop the car's hood and display the engine?										
Could you demonstrate the functionality of this tool?										
Could you provide me with a whiskey from the same distillery that has been aged longer?										
Spread your arms, open your wings										
Could you pour batter into these cupcake liners?										
Enlarge the image in the middle of the first row.										

Figure 24. Comprehensive visualization of model performance on IntelligentBench (Subset World knowledge, part 2/15).

Prompt	Source	Reference	LightFusion-World	OmniGen2	Step1x-Edit	FLUX.1 Kontext	Qwen-Image	BAGEL	GPT-Image-1	NanoBanana
Could you provide a top-down image of these shoes?										
Can you display the furniture with open storage areas?										
Could you display the serving process for this drink?										
Draw a line connecting all the black dots.										
Can you display someone cleaning with these items?										
Show what the items in the picture look like when being used.										
Remove the lid and lean it against the basket.										
What is the next stage in the preparation of this dough for making hot cross buns?										
What does the tiger look like when it is walking on land instead of being in water?										
How does the animal in the picture gather its companions?										
Could you design another set of advertisements for this product in the same style but different colors?										
What would this fish look like after it has been filleted and prepared for cooking?										

Figure 25. Comprehensive visualization of model performance on IntelligentBench (Subset World knowledge, part 3/15).

Prompt	Source	Reference	LightFusion-World	OmniGen2	Step1x-Edit	FLUX.1 Kontext	Qwen-Image	BAGEL	GPT-Image-1	NanoBanana
What does the product inside the box shown in Figure 1 look like when fully assembled and ready for use?										
What does the vehicle in Figure 1 look like when its ramp and doors are closed?										
Show the usage scenario of the item in the image.										
What might this dinosaur have looked like in its natural habitat under clear daylight conditions?										
What does the offspring of the animals shown in Figure 1 look like?										
What does the interior seating arrangement of this vehicle look like?										
What does the same dog in the figure look like when fully grown?										
What does the interior of one of the squares in this baked item look like when cut open?										
What does the engine compartment of this car look like?										
What vegetable is typically processed to create the food shown in Figure 1?										
The appearance of the item in the picture after opening its shell.										
What might this dog look like as an adult while actively playing outdoors?										
What does the item in the box look like when taken out?										

Figure 26. Comprehensive visualization of model performance on IntelligentBench (Subset World knowledge, part 4/15).

Prompt	Source	Reference	LightFusion-World	OmniGen2	Step1x-Edit	FLUX.1 Kontext	Qwen-Image	BAGEL	GPT-Image-1	NanoBanana
Show the image from the perspective of a person on the bridge.										
What does the scenery in the picture look like when viewed from bottom to top?										
What does this location look like from an aerial perspective that reveals its overall shape?										
What does this item look like when unfolded and in use?										
What does the item in the picture look like after being cut open?										
What does the door of the structure in Figure 1 look like up close?										
What does the surrounding landscape and neighborhood look like from a wider aerial perspective of the castle shown?										
What does the object in the picture look like from the side?										
What does this vineyard look like from a wider perspective, showing more of its layout and surroundings?										
What is one possible prepared dish that could be made using the items shown in the basket?										
What does this location look like during sunset?										

Figure 27. Comprehensive visualization of model performance on IntelligentBench (Subset World knowledge, part 5/15).

Prompt	Source	Reference	LightFusion-World	OmniGen2	Step1x-Edit	FLUX.1 Kontext	Qwen-Image	BAGEL	GPT-Image-1	NanoBanana
What should be done next after the action in the image and after completely covering these plants with soil?										
What other environment can this animal live in besides the one shown?										
What does the food inside the oven look like after the cooking process is complete?										
What type of bridge design would be used in a similar cultural setting to allow boats to pass underneath?										
The fruit in the image is ready for harvest. Could you display how it looks before it is ready?										
What is the appearance of these pork chops after they are seared?										
What sport requires the use of the item in the picture?										
What does the drink look like after being fully mixed?										
What does this pair of headphones look like when folded for compact storage?										
What action is needed to turn the items in the image into coffee beans suitable for brewing coffee?										
What does this dough look like after it has been baked?										
What does the internal structure of the object look like when its outer shell is removed?										

Figure 28. Comprehensive visualization of model performance on IntelligentBench (Subset World knowledge, part 6/15).

Prompt	Source	Reference	LightFusion-World	OmniGen2	Step1x-Edit	FLUX.1 Kontext	Qwen-Image	BAGEL	GPT-Image-1	NanoBanana
What does this bird look like when it is standing on the ground?										
What activity is this bicycle specifically designed for based on its features?										
What does the rocket in Figure 1 look like during its launch?										
What is the result after blending the ingredients shown in the food processor?										
What does the seating and storage layout look like when viewed from the rear of this boat?										
What does the camera look like when the lens is retracted?										
What would this dog look like as an adult?										
What plant is the drink in the picture made from?										
What does the engine compartment of the vehicle shown in Figure 1 look like?										
What is the typical product obtained from the animal shown in the image?										
What does the root system of the plant in Figure 1 look like when removed from its pot?										

Figure 29. Comprehensive visualization of model performance on IntelligentBench (Subset World knowledge, part 7/15).

Prompt	Source	Reference	LightFusion-World	OmniGen2	Step1x-Edit	FLUX.1 Kontext	Qwen-Image	BAGEL	GPT-Image-1	NanoBanana
What are the raw ingredients likely used to prepare the contents of the jar shown?										
What does the front of the animal in the picture look like?										
What does this dog look like when it is happy and displaying a friendly demeanor?										
What does the stack of these materials look like when viewed from the side?										
What animal is hanging from the tree branch in the background of this image? Please give it a close-up.										
What does the interior cargo space of this vehicle look like with the rear seats folded down?										
What does the steak look like after it has been sliced?										
Show the usage scenarios of the transportation tools in the image										
What does the engine compartment of the car shown in Figure 1 look like?										
What does the complete setup of the device shown in close-up look like?										
What is the processed form of the contents shown in the baskets?										
What did the item in the picture look like when it was alive?										

Figure 30. Comprehensive visualization of model performance on IntelligentBench (Subset World knowledge, part 8/15).

Prompt	Source	Reference	LightFusion-World	OmniGen2	Step1x-Edit	FLUX.1 Kontext	Qwen-Image	BAGEL	GPT-Image-1	NanoBanana
What can the items in the picture be used to make?										
What does the raw form of the meat shown in Figure 1 look like before cooking?										
What would these shrimp look like after being grilled?										
What does the connector that links the item in the picture to the computer look like?										
What does this scene look like during sunset when the sky is more vibrant?										
What does the thing in the picture look like after it hatches?										
What is the appearance of these dough pieces after they have been baked?										
What is the final appearance of these items after they are fully cooked and decorated?										
What does the internal composition of this manual coffee grinder look like when disassembled?										
What does the animal in Figure 1 look like when it is inside its burrow?										
What would these steaks look like after being cooked in an air fryer?										
What is the appearance of these pastries after they have been baked?										

Figure 31. Comprehensive visualization of model performance on IntelligentBench (Subset World knowledge, part 9/15).

Prompt	Source	Reference	LightFusion-World	OmniGen2	Step1x-Edit	FLUX.1 Kontext	Qwen-Image	BAGEL	GPT-Image-1	NanoBanana
Show what the assembled items in the picture look like.										
What does the seating arrangement in the rear of this vehicle look like?										
Show the scene of the animal eating in the picture.										
What does this laboratory space look like after being repurposed into a cafeteria in the same style?										
What does the interior of the car shown in Figure 1 look like?										
What does this dish look like after it has been baked?										
Show the dish made from the ingredients in the picture.										
What would this scene look like from a top-down perspective?										
What does the item shown in the picture look like when opened in real life?										
What does the scene look like when the items in the picture are folded and placed in the trunk?										
Show what the food in the image looks like when being cooked.										
What does this bird look like in flight?										

Figure 32. Comprehensive visualization of model performance on IntelligentBench (Subset World knowledge, part 10/15).

Prompt	Source	Reference	LightFusion-World	OmniGen2	Step1x-Edit	FLUX.1 Kontext	Qwen-Image	BAGEL	GPT-Image-1	NanoBanana
What does the exterior of the structure housing this statue look like?										
What does the rear view of the car shown in Figure 1 look like?										
Remove the cut-out piece.										
What does the dough in Figure 1 look like after it has been baked?										
What does the interior of the car shown in Figure 1 look like?										
What does this bread look like after it has been sliced?										
What does the interior of the car shown in the image look like?										
What does the interior of this car look like?										
What does the interior of this vehicle look like, including its dashboard and controls?										
Show a usage scenario of an item in the image										
How do the animals in the picture rest?										
What do these eggs look like after they hatch?										
What does the flower in Figure 1 look like when viewed from directly above?										

Figure 33. Comprehensive visualization of model performance on IntelligentBench (Subset World knowledge, part 11/15).

Prompt	Source	Reference	LightFusion-World	OmniGen2	Step1x-Edit	FLUX.1 Kontext	Qwen-Image	BAGEL	GPT-Image-1	NanoBanana
What does the animal in the picture look like when it grows up?										
What product is commonly derived from the plant shown in Figure 1 and how does it appear in its processed form?										
What does the interior of the vehicle shown in Figure 1 look like?										
What would this chicken look like before it was seasoned?										
What does this island look like during winter?										
Show a scene of pouring beer into a glass.										
Could you show the sections of this baby bottle separately?										
Put the orange peel inside the glass bottle.										
Show the view from the second-floor balcony.										
Can you shift the perspective to the right and add more dolls in a similar style to the ones in the figure?										
Could you provide a view of this bag's inside?										
Could you share a view of this wine rack empty?										

Figure 34. Comprehensive visualization of model performance on IntelligentBench (Subset World knowledge, part 12/15).

Prompt	Source	Reference	LightFusion-World	OmniGen2	Step1x-Edit	FLUX.1 Kontext	Qwen-Image	BAGEL	GPT-Image-1	NanoBanana
I want to see what a cake looks like when it's cut vertically into two halves.										
Show the complete artwork										
Could you display the item with the lid ajar?										
Could you display the opposite angle of the room?										
Could you show me a detailed view of inside the pouch?										
Could you display the interior of the bag?										
Could you present an image of a phoenix in the same style?										
Could you display this truck's cabin view?										
Could you display the launch of this space shuttle?										
Could you show me the dining space linked to the kitchen?										
Could you display the attire of the individual holding this bag?										
Can you take the mattress off the bed for me?										

Figure 35. Comprehensive visualization of model performance on IntelligentBench (Subset World knowledge, part 13/15).

Prompt	Source	Reference	LightFusion-World	OmniGen2	Step1x-Edit	FLUX.1 Kontext	Qwen-Image	BAGEL	GPT-Image-1	NanoBanana
Can you design a paired conditioner and place next to this item?										
Could you display this jacket being worn while cycling?										
Could you unlock the car's back door for me?										
Show me the entrance										
Show the state of player number 2 preparing to throw the ball.										
Open the lid of the cream on the right and display its contents.										
Could you provide a picture of this animal after several months?										
Can you show me the toy truck unboxed?										
Walk to the lake.										
Can you display the blooming stage of this flower?										
Please zoom in to see the little monster and the numbers.										

Figure 36. Comprehensive visualization of model performance on IntelligentBench (Subset World knowledge, part 14/15).

Prompt	Source	Reference	LightFusion-World	OmniGen2	Step1x-Edit	FLUX.1 Kontext	Qwen-Image	BAGEL	GPT-Image-1	NanoBanana
Show the baby waving his right hand.										
Show the ingredients of this dish.										
Could you show how these napkin rings are used when holding a napkin in place?										
Could you transform this fish into a fillet?										
What does the finished product of the item shown in the picture look like?										
Could you display what the car's inside looks like?										
How does the mixture appear post egg blending?										

Figure 37. Comprehensive visualization of model performance on IntelligentBench (Subset World knowledge, part 15/15).

Prompt	Input Image	OmniGen	StepIX-Edit	Qwen-Image	NanoBanana	GPT-Image-1	LightFusion-World
Draw what it looks like after being exposed to the sun for a day.							
Draw what it will look like after one month of being neglected.							
Draw what it will look like an hour later.							
Draw what it will look like after six months.							
Draw what it will look like two weeks after flowering.							
Draw what it will look like after 200 years.							
Draw what it will look like after baking for 45 minutes.							
Draw what it will look like ten seconds later.							
Draw what it will look like after one hour on a hot grill.							
Draw what it will look like after 50 years.							
Draw what it will look like after a cat has played with it.							

Figure 38. Qualitative comparison on RISE benchmark.