

STAR-R1: Multi-View Spatial TrAnsformation Reasoning by Reinforcing Multimodal LLMs

Supplementary Material

A. Comparison with other works

Compared with studies [6, 10], despite shared high-level findings on SFT versus RL, STAR-R1 investigates this phenomenon within the distinct domain of fine-grained multi-view spatial reasoning (focusing on object correspondence rather than navigation). Specifically, we identify a unique mechanism: RL drives an explicit object-anchoring behavior, identifying key objects for cross-view alignment to capture spatial semantics. In contrast, SFT fits superficial shortcut patterns tied to object attribute changes and degrades significantly under view shifts.

While Actial [66] and COMFORT [70] share the broad goal of spatial reasoning, STAR-R1 fundamentally differs in scope and methodology. Actial focuses on global viewpoint learning (camera movement) for macroscopic 3D consistency, whereas COMFORT is an evaluative work probing static biases in spatial frames of reference. In contrast, STAR-R1 is a methodological work targeting complex dynamic multi-view scenarios. Inspired by the TVR task, we identify the anchor object as decisive for cross-view reasoning. Unlike Actial’s general RL or COMFORT’s static evaluation, we propose a targeted RL framework (structured SFT + key object selection reward) that explicitly empowers models to lock onto and align objects across changing viewpoints, achieving robust fine-grained spatial reasoning beyond simple camera movement or static bias probing.

B. Dataset and experimental setup

For the TVR task, we first clean the TRANCE dataset by removing samples containing redundant transformations. Then, we randomly sample 9,000 and 4,500 samples as the training set and test set, respectively, ensuring that the number of samples corresponding to each transformation length (ranging from 1 to 4) is equally distributed in both the training and test sets.

For the real-world datasets, we first curate 2.4k and 1.7k samples from the MindCube and SPAR-7M datasets, respectively. We then use Gemini-2.5-Pro to generate high-quality, long CoT data for the first-stage supervised fine-tuning (SFT), as outlined in Sec. 3.2.1. Subsequently, we collect an additional 12k samples from MindCube and 7.2k from SPAR-7M for the second-stage reinforcement learning (RL) training. For evaluation, we employ the MMSI-Bench, MindCube-Tiny, and SPAR-Bench as our evaluation benchmarks.

For both the TVR task and the tasks involving real-world

datasets, we employ Qwen2.5-VL-7B as our base model and utilize vLLM as the training framework. All experiments (including RL and SFT) are conducted on a single node equipped with $8 \times$ H20 GPUs. For the exploratory experiments on the TVR task, we perform approximately 1,200 steps of single-stage RL training. For the real-world datasets, in contrast, the training involves a two-stage process: a cold-start SFT phase for 2 epochs, followed by an RL phase of approximately 1,200 training steps. For the RL experiments, 8 answers are generated for each question to compute the corresponding advantage function.

C. Training curves on the TVR task

We illustrate the evolution of *Total Accuracy* for the STAR-R1 and STAR-SFT models on OOD test data of the TVR task across training steps. The training dynamics plotted in Fig. 7 further show that SFT quickly plateaus around 1,200 steps on OOD data and eventually degrades with further training, whereas RL maintains steady performance growth throughout the training process, suggesting its unique ability to progressively unlock the model’s reasoning capabilities in complex scenarios.

Additionally, we plot further curves to analyze the training dynamics of STAR-R1 on TVR task. As shown in the right panel of Fig. 5, we present the format reward curve of the STAR-R1. It can be observed that for models with strong foundational capabilities, such as Qwen2.5-VL-7B, the required format is already followed by most responses from the beginning. Thus, no additional customization for format reward is necessary. Moreover, in the left panel of Fig. 5, we illustrate the accuracy reward curve. The model’s accuracy reward steadily increases with training progression, ultimately stabilizing at a high value. This indicates that our designed reward function effectively encourages the model to engage in accurate, proactive, and efficient exploration, progressively unlocking the multimodal reasoning capabilities of the base model.

As shown in Fig. 6, these five curves illustrate the dynamics of the *Total Accuracy* as well as the *Attribute Accuracy* for the four types throughout the training process. All accuracy curves exhibit a steady increase during training and eventually stabilize at a high level of prediction accuracy. This demonstrates that our STAR-R1 effectively guides the model to efficiently explore the solution space and gradually develop stronger reasoning capabilities.

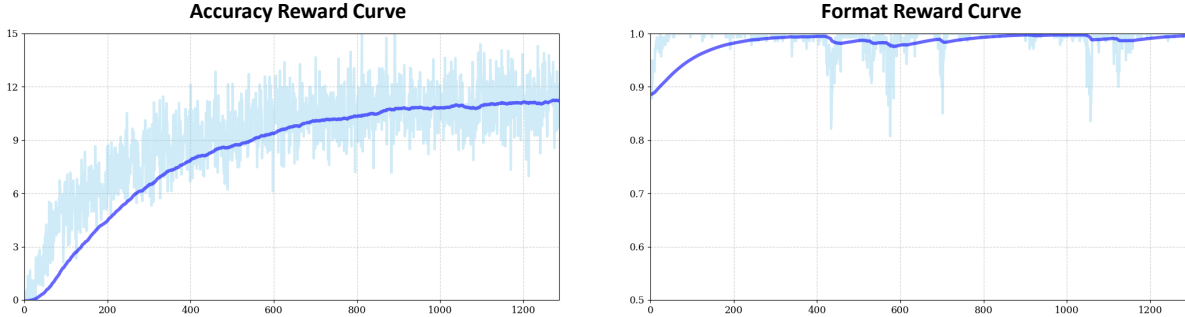


Figure 5. Training curves about accuracy reward and format reward.

Table 5. Comparison of RL vs. SFT performance on ID and OOD settings.

Method	ID			OOD		
	TAcc	Diff	NDiff	TAcc	Diff	NDiff
STAR-SFT	84.2	0.22	0.08	30.9	1.65	0.83
STAR-SFT&RL	87.7	0.19	0.07	36.5	1.53	0.71
STAR-R1	76.3	0.38	0.14	53.9	0.96	0.39

D. Training STAR-SFT with RL on TVR task

We additionally apply RL training to the SFT model on the TVR task. In Row 2 of Tab. 5, results demonstrate that further training the model with RL not only achieves continuous improvement in ID performance but, more importantly, exhibits significant progress on OOD tasks. This indicates that RL successfully breaks through the performance bottleneck encountered during SFT training on OOD data. Furthermore, we observed that the two-phase training approach of STAR-SFT&RL does not ultimately achieve the same performance on OOD data as pure RL training. We hypothesize that this may be due to imitation learning in the SFT phase locking the model’s reasoning patterns, such that the subsequent RL training phase using the same amount of data can only marginally adjust the model’s reasoning. This adjustment leads the model to attempt comparing more objects.

To validate this conjecture, we further analyze the difference between the number of objects compared in each response and the total number of objects in the scene. The average difference for STAR-SFT is -3.3, while for STAR-SFT&RL, it is -3.0. This indicates a slight increase in the number of objects compared during the reasoning process of the STAR-SFT&RL. Consequently, this results in a modest improvement in OOD accuracy, though it still falls short of the performance achieved by the pure RL training.

Table 6. Ablation studies on data volume.

Volume	TAcc	Diff	NDiff	Object Total Accuracy			
				Num3	Num6	Num8	Num10
1,000	13.8	2.32	0.95	29.0	16.2	9.0	4.9
3,000	23.7	1.97	0.88	49.0	25.3	17.2	10.3
5,000	40.7	1.49	0.65	75.1	47.0	30.1	19.5
7,000	52.6	1.08	0.46	82.7	59.6	45.0	31.2
9,000	61.4	0.77	0.31	91.0	70.7	54.2	37.5

E. Ablation studies on the TVR dataset

Training Data. To evaluate the effect of training data size, we randomly sample four subsets for training from the whole training set, containing 1000, 3000, 5000, and 7000 samples, respectively, and evaluate them on the original test set. The experimental results shown in Tab. 6 demonstrate that as the volume of training data increases, model performance improves continuously, but the growth rate of accuracy exhibits an initial increase followed by a subsequent decline. This suggests that insufficient data in the early stages constrains model learning, while expanded datasets facilitate a phase of rapid knowledge acquisition, leading to significantly accelerated performance gains.

F. Case studies on TVR task

In this section, we provide a detailed presentation of our problem prompt (shown in Fig. 8) along with additional case studies on TVR task. The **{ObjectFeature}** in Fig. 8 will be substituted by all features corresponding to each object in the initial image of every sample. An example is provided below: {idx: 0; color: yellow; material: metal; shape: cylinder; size: medium}, {idx: 1; color: gray; material: rubber; shape: sphere; size: medium}, {idx: 2; color: blue; material: rubber; shape: cylinder; size: medium}, {idx: 3; color: yellow; material: metal; shape: sphere; size: medium}, {idx: 4; color: blue; material: rubber; shape: cube; size: large}, {idx: 5; color: brown; material: rubber; shape: sphere; size: medium}, {idx: 6; color: brown; material: rubber; shape: sphere; size: large}.

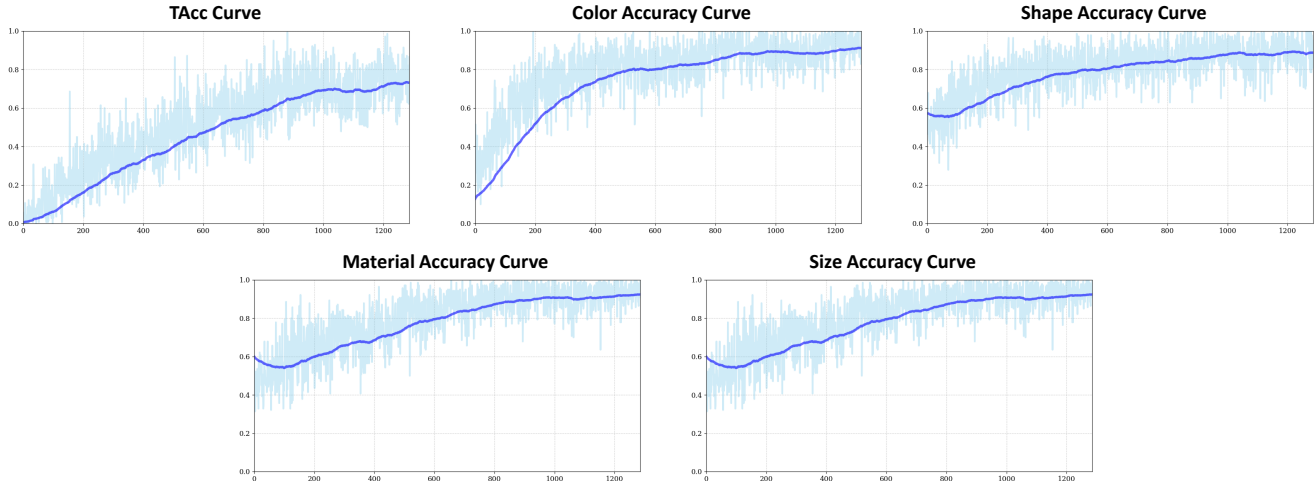


Figure 6. Dynamics of the total accuracy as well as the attribute accuracy for the four types throughout the training process.

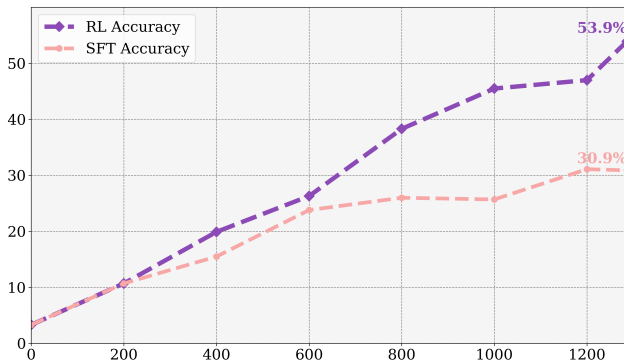


Figure 7. Accuracy curves of RL and SFT.

The case studies shown in Figs. 9 to 11 further substantiate the argument presented in Section 4.2 of the main paper, namely that STAR-SFT merely engages in imitation learning and does not perceive viewpoint changes. It erroneously associates objects located at the same positions in the initial and final images (referring to their positions within the images, not their positions in the real-world scene) as the same object, thereby significantly impairing the model’s reasoning performance.

In contrast, STAR-R1 demonstrates human-like behavior by systematically comparing the states of all objects between the initial and final images during its reasoning process to ascertain their correspondence. Consequently, without any explicit mention of viewpoint changes in the problem prompts, STAR-R1 successfully detects the perspective shift and achieves a significantly higher answer accuracy than STAR-SFT.

Since the STAR-SFT model fails to perceive changes in viewpoint, it mistakenly assumes that objects at the same position in two images are the same object. In Fig. 9, the

STAR-SFT model incorrectly identifies objects numbered 0, 3, and 5 as objects numbered 3, 5, and 6. In Fig. 10, it misidentifies objects numbered 1, 0, and 5 as objects numbered 0, 5, and 4. In Fig. 11, the model confuses objects numbered 3, 4, and 0 with objects numbered 7, 3, and 4. In contrast, by comparing the states of all objects in both images, STAR-R1 correctly matches the same objects across the two images and ultimately outputs the correct answer.

G. Case studies on MindCube-Tiny

In this section, we provide more case studies on MindCube-Tiny. Figs. 12 to 14 respectively illustrate the reasoning processes and final answers generated by our model STAR-R1, on the Rotation, Among, and Around subtasks of the MindCube-Tiny.

H. Case studies on SPAR-Bench

In this section, we provide more case studies on SPAR-Bench. Figs. 15 and 16 respectively illustrate the reasoning processes and final answers generated by our model STAR-R1, on the ObjRel-OC-MV and ObjRel-OO-MV subtasks of the SPAR-Bench.



Question

I am giving you two images containing a series of objects. Below, I will provide the attribute information describing each object. You need to carefully observe both images and identify which objects from the first image have changed and what specific transformation operations are performed.

For attribute information, there are four types: shape, size, color, and material. The possible values for each attribute are:

1. shape: "cube", "sphere", or "cylinder"
2. size: "large" (radius = 6), "medium" (radius = 4), or "small" (radius = 2)
3. color: "gray", "red", "blue", "green", "brown", "purple", "cyan", or "yellow"
4. material: "rubber", "metal", or "glass"

In the first image, information about each object is as follows, with object index starting from 0:

{ObjectFeature}.

Note that for each object, each attribute can only be transformed once.

The final answer should include all transformation operations, with each operation expressed in the following format:

- "idx.attribute.value" (e.g., "0.size.small", "2.color.red")

- where "idx" is the object index

- "attribute" is the attribute that was changed (shape, size, color, or material)

- "value" is the NEW value after transformation (e.g., if a cube changed to a sphere, the value would be "sphere")

List transformations in sequential order as the final answer, separated by semicolons (e.g., '3.shape.cube; 6.material.metal'). The number of transformations must range from 1 to 4.

Output your thinking process in <think>...</think> tags and your final answer in <answer>...</answer> tags.

Figure 8. The problem prompt.

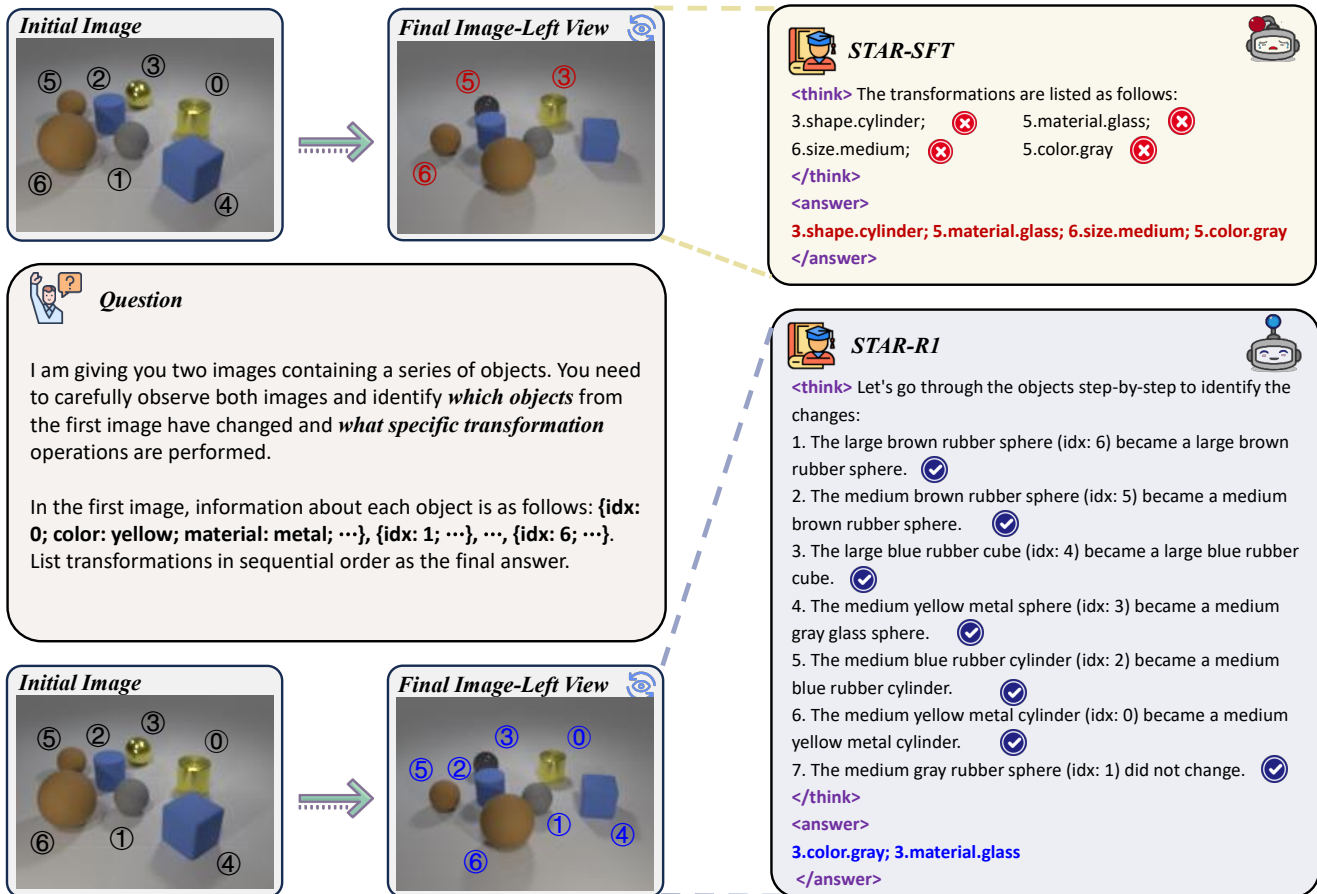


Figure 9. A case study comparing the reasoning processes of STAR-SFT and STAR-R1. Blue indicates correct answers while red denotes incorrect ones.



Figure 10. A case study comparing the reasoning processes of STAR-SFT and STAR-R1. Blue indicates correct answers while red denotes incorrect ones.

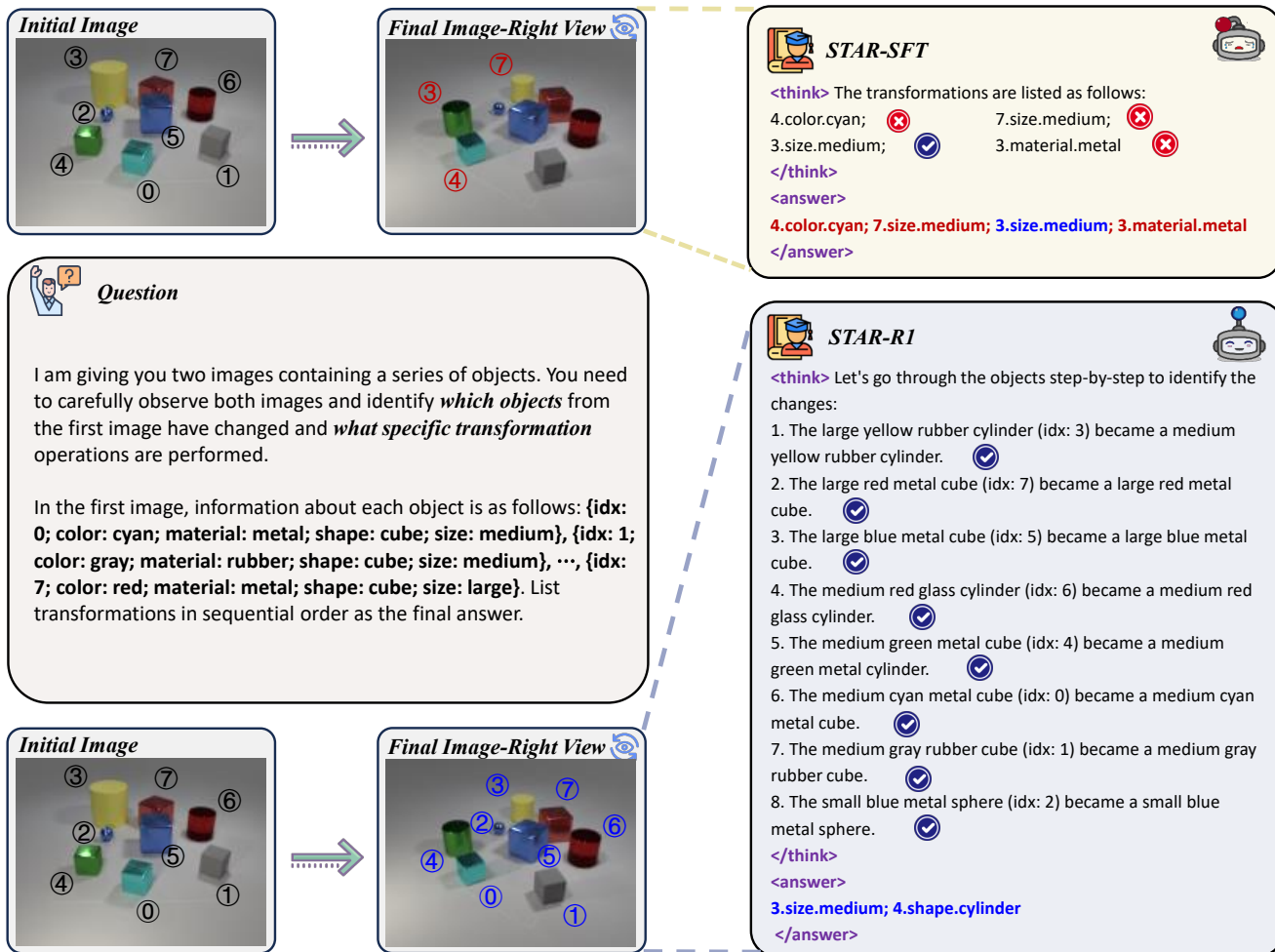
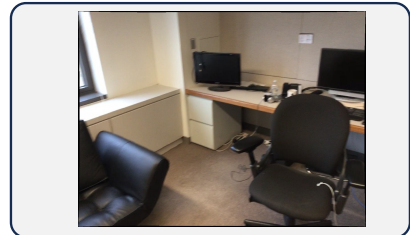
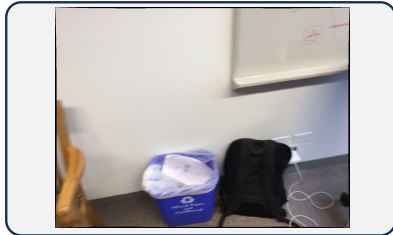


Figure 11. A case study comparing the reasoning processes of STAR-SFT and STAR-R1. Blue indicates correct answers while red denotes incorrect ones.



Question

These three images (image 1, 2, and 3) show the same scene from three different viewpoints. The image 2 was taken after turning the camera 90 degrees to the right (clockwise) from the position of image 1. For image 3, the camera was turned another 90 degrees right, so it's basically facing the opposite direction of image 1. Based on these three images: If I am standing at the same spot and facing the same direction as shown in image 2, then I turn 90 degrees to the left, what is to my right? A. Black sofa and window B. Blue bin and black bag C. Monitors on desk



STAR-RI



<think> Image 1:- Key objects: Blue bin, black bag, whiteboard, wall. - Spatial relations: [Blue bin, wall, against], [black bag, whiteboard, next], [blue bin, black bag, left]. Camera is facing wall/bin area.

Image 2:- Key objects: Whiteboard, black sofa, window. - Spatial relations: [whiteboard, black sofa, next], [black sofa, window, next]. Camera is facing sofa/window area.

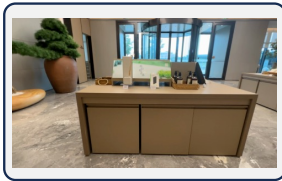
Image 3:- Key objects: Desk, black sofa, black office chair, window.- Spatial relations: [black sofa, desk, left], [desk, black office chair, next]. Camera is facing desk area.

Spatial Mapping: The three images represent a 360-degree layout of the room. Image 1 (bin/bag) is 90° to the left of Image 2 (sofa/window). Answer: The starting position is facing the direction of Image 2 (sofa/window). Turning 90 degrees to the left **means you are now facing the direction of Image 1 (bin/bag)**. From this new orientation, the area to your right is Image 2. Image 2 shows the black sofa and window. Therefore, the black sofa and window are to your right

</think>

<answer>A</answer>

Figure 12. A case study on Rotation subtask of MindCube-Tiny.



Question

Based on these four images (image 1, 2, 3, and 4) showing the sideboard table from different viewpoints (front, left, back, and right), with each camera aligned with room walls and partially capturing the surroundings: From the viewpoint presented in image 4, what is to the left of the sideboard table?
A. Gate B. Wooden door and yellow decoration C. Display shelves D. Sideboard table



STAR-R1

<think>

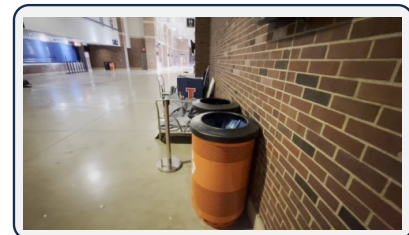
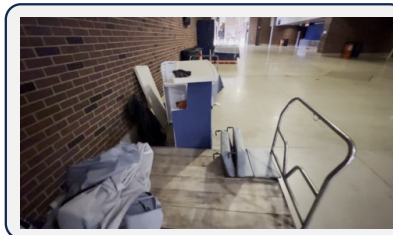
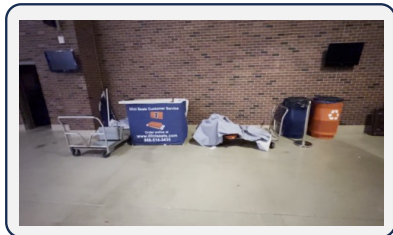
Image 1: - Key objects: Sideboard table, gate/glass doors. - Spatial relations: [Gate, sideboard table, behind]. This is the 'front' view.
Image 2: - Key objects: Sideboard table, display shelves. - Spatial relations: [display shelves, sideboard table, behind]. This is the 'left' view.
Image 3: - Key objects: Sideboard table, display shelves.- Spatial relations: [display shelves, sideboard table, behind]. This is the 'back' view.
Image 4: - Key objects: Sideboard table, wooden door.- Spatial relations: [wooden door, sideboard table, behind]. This is the 'right' view
Spatial Mapping: The four images provide a 360-degree layout of the area around the sideboard table. Image 1 (front) shows gate behind. Image 2 (left) shows display shelves behind. Image 3 (back) shows display shelves behind. Image 4 (right) shows wooden door behind. This creates a complete 360-degree 3D layout of the surrounding large features.

Answer: The question asks what is to the left from the perspective of Image 4. Image 4 is the 'right' view, where the camera is looking towards the right. **From this viewpoint, the area to the left is Image 3. Image 3 shows the display shelves behind the table.** Therefore, the display shelves are to the left.</think>

<answer>C</answer>



Figure 13. A case study on Among subtask of MindCube-Tiny.



Question

Based on these three images (image 1, 2, and 3) showing the same scene from different viewpoints (front, left, and right): From the viewpoint presented in image 3, what is the nearest object behind the orange bin? A. Blue-printed box B. Blue bin C. Metal trolley D. Wall

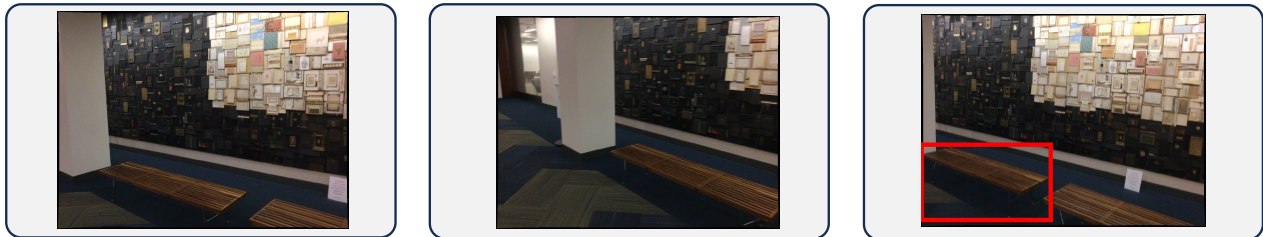


STAR-RI



<think> Image 1:- Key objects: Blue-printed box, Metal trolley, Orange bin, Wall. -Spatial relations: [Blue-printed box, Wall, against], [Metal trolley, Blue-printed box, to left], [Orange bin, Wall, against].
 Image 2: - Key objects: Metal trolley, Blue-printed box, Orange bin. - Spatial relations: [Metal trolley, Blue-printed box, closest], [Blue-printed box, Wall, against].
 Image 3:- Key objects: Orange bin, Blue bin, Metal trolley, Wall. - Spatial relations: [Orange bin, Wall, against], [**Blue bin, Orange bin, behind**], [Metal trolley, Blue bin, behind].
 Spatial Mapping: The objects are arranged along a wall. **From left to right, the order is Metal trolley, Blue-printed box, Blue bin, and Orange bin.** Image 1 shows them side-by-side. Image 2 (from the far left) shows Metal trolley closest. Image 3 (from the far right) shows Orange bin closest.
 Answer: Based on spatial mapping, Image 3 shows the orange bin closest to the wall. The nearest object located directly behind the orange bin is the blue bin. The blue-printed box is further to the far left. The metal trolley is further to the far left-left. Therefore, the blue bin is the nearest object behind the orange bin from this viewpoint*</think>*
<answer>B</answer>

Figure 14. A case study on Around subtask of MindCube-Tiny.



Question

In what direction does object bench (bbox) appear from the observer's perspective? Calculate or judge based on the 3D center points of these objects. The first image becomes the primary perspective, correlating with the observer's view. A. right, below, front B. left, below, front C. left, below, behind D. right, above, behind



STAR-R1



<think> Image 1: - Key objects: bench, wall, pillar, floor. - Spatial relations: [bench, observer, below], [bench, observer, front], [bench, observer, left]. [bench, wall, in front].

Image 2: - Key objects: bench, wall, pillar. - Spatial relations: [bench, observer, right], [bench, observer, below], [bench, observer, front]. [bench, pillar, next].

Image 3: - Key objects: bench (in bbox), wall, bench (partially). pillar. - Spatial relations: [bench, observer, front-left], [bench, observer, below]. [bench, wall, next].

Spatial Mapping: The bench is clearly located on the floor, below the observer's eye level. It is positioned **towards the left of the scene**. As it is closer to the foreground, it is in front. Combining these observations from the **front-left perspective, the bench is below and in front**. Therefore, its 3D center is below the observer's initial viewpoint.

Answer: Based on spatial mapping from the observer's perspective, the bench is located to the left. It is on the floor, below. It is clearly in front. Therefore, the bench is 'left, below, front'</think>

<answer>B</answer>

Figure 15. A case study on ObjRel-OC-MV subtask of SPAR-Bench.



Question

Where is the armchair (red bbox) found spatially in comparison to the table (blue bbox) in the two images, according to the observer? Calculate or judge based on the 3D center points of these objects. We establish the main perspective by aligning the first image with the observer's view. A. right, below, farther, , front B. right, above, farther, front, front C. right, below, farther, , behind D. left, , farther, front, front



STAR-R1



<think> Image 1: - Key objects: Armchair, Table, Wall. - Spatial relations: [Armchair, Table, left], [Table, Wall, against]. [Armchair, Floor, on]. Image 2: - Key objects: Armchair, Table, Second Armchair, Window. - Spatial relations: [Armchair, Table, left], [Table, Window, in front]. Image 3: - Key objects: Armchair, Table. - Spatial relations: [Armchair, Table, left], [Armchair, Table, closer to left]. Spatial Mapping: The main perspective is from the left, close to the armchair. The table is located to the right of the armchair. Therefore, **the armchair is to the left of the table**. Both objects are on the floor. However, **the armchair is closer to the left of the room**. Therefore, its center is likely slightly farther from the initial close-up area. Combining these observations, the armchair is closer to the far-left area of the space. Therefore, **the armchair's center is farther from the initial observer's close-up area than the table's center**. Therefore, the armchair is to the left and farther. Answer: Based on spatial mapping from the observer's perspective: 1. Left-Right: Armchair is to the left. 2. Above-Below: The resulting spatial description is [left, , farther, front, front]</think>
<answer>D</answer>

Figure 16. A case study on ObjRel-OO-MV subtask of SPAR-Bench.