

Level 1 - Global Occlusion 80%

Question: You are presented with one main image in which a substantial part of the scene is occluded or masked. Alongside it, you are shown four complete candidate images, labeled A, B, C, and D.

Exactly one candidate represents the original, unoccluded version of the main image.

Task: 1. Focus on the visible structures and contextual cues within the occluded image — such as object outlines, edges, colors, textures, and spatial layout; 2. Compare these visible regions with each of the four candidates to determine which one aligns most closely in terms of composition, geometry, and contextual meaning; 3. Choose the candidate that best matches the occluded image and restores the original scene.

In your reply, follow this format exactly:

Reasoning: Describe the main rule or pattern and explain why your answer fits it.

Answer: [A, B, C, or D]



GPT-5-Thinking

Reasoning: The occluded image shows a brick house with a white arched doorway, a white car in its driveway, and a neighboring tan house with a hedge and black car to the right. The left neighboring house is clearly visible.

Candidate B matches this composition and perspective best—the amount of the left house shown, the alignment of the cars/hedge, and the street view all line up. A shows less of the left house, while C and D are taken from farther right.

Model Answer: B ✗



Ground Truth: A ✓

B

C

D

Perceptual failure: the model relies on coarse semantic similarity (houses, cars, street) but misjudges the exact viewpoint and spatial layout.