

Phase 1: Source Filtering

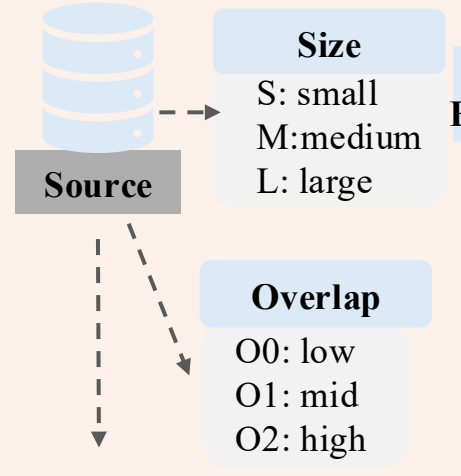
Uniqueness
U1: unique category in image
U2: non-unique

Clutter
C1: low (1-4)
C2: medium (5-9)
C3: high (≥ 10)

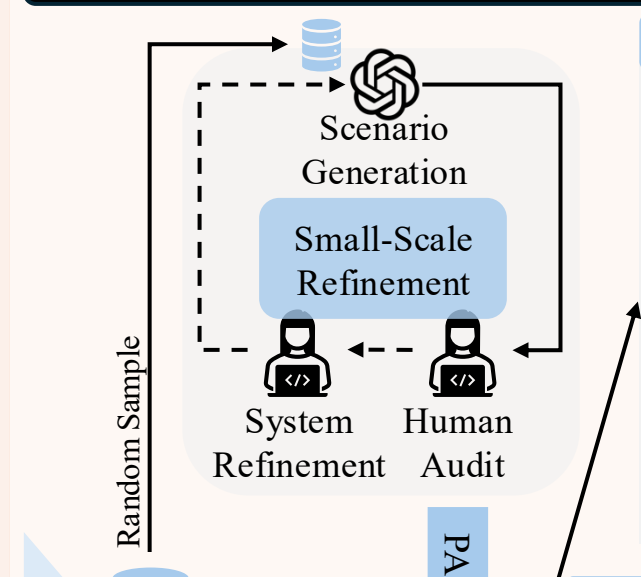
Size
S: small
M: medium
L: large

Overlap
O0: low
O1: mid
O2: high

Position
P0: near center
P1: off-center

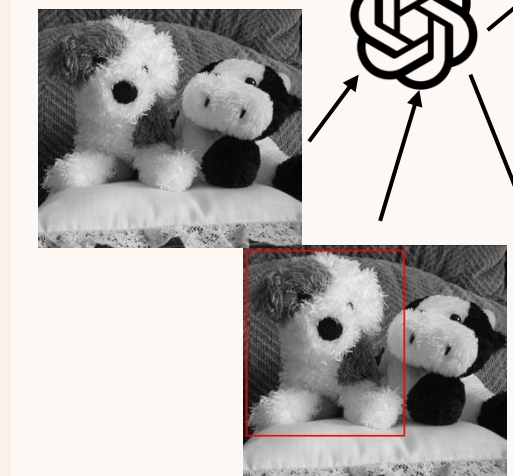


Phase 2: Annotation



Target Object Description
Object Description: plush toy with floppy ears and dark patch over one eye
Attributes:
color: white with a dark patch over one eye and ear
material: plush fabric
shape: rounded, animal-like with floppy ears
texture: soft, fuzzy
size: medium, similar to the other plush toy
position: left side of the pillow, foreground
affordance: cuddling, display
relation: beside a plush toy with a wide snout and black spots

Scenario Query & Reasoning Traces
Scenario: As a parent organizing a child's toys, you want to find the soft, fuzzy friend with a dark patch over one eye, sitting to the left of the plush with a wide snout and black spots. Your goal is to separate it from the other plush for cleaning, using its unique markings, ear shape, and position as cues.
Thought: The scenario highlights the need to distinguish between two plush toys. The target is identified by its left position, distinctive dark patch over one eye, and floppy ears, contrasting with the adjacent plush that has a wide snout and black spots. These visual and relational cues make the correct object clear.
Referring Expression Cues: The soft, fuzzy toy on the left with a dark patch over one eye, sitting beside the spotted plush with a wide snout.



Labels and Prediction for Filtering
Acceptable Names: dog, puppy, stuffed dog, toy dog
LLM Predicted BBox (norm): 0.014, 0.024, 0.594, 0.764

Ground Truth Bounding Box & Category

Phase 3: Audit & Filter

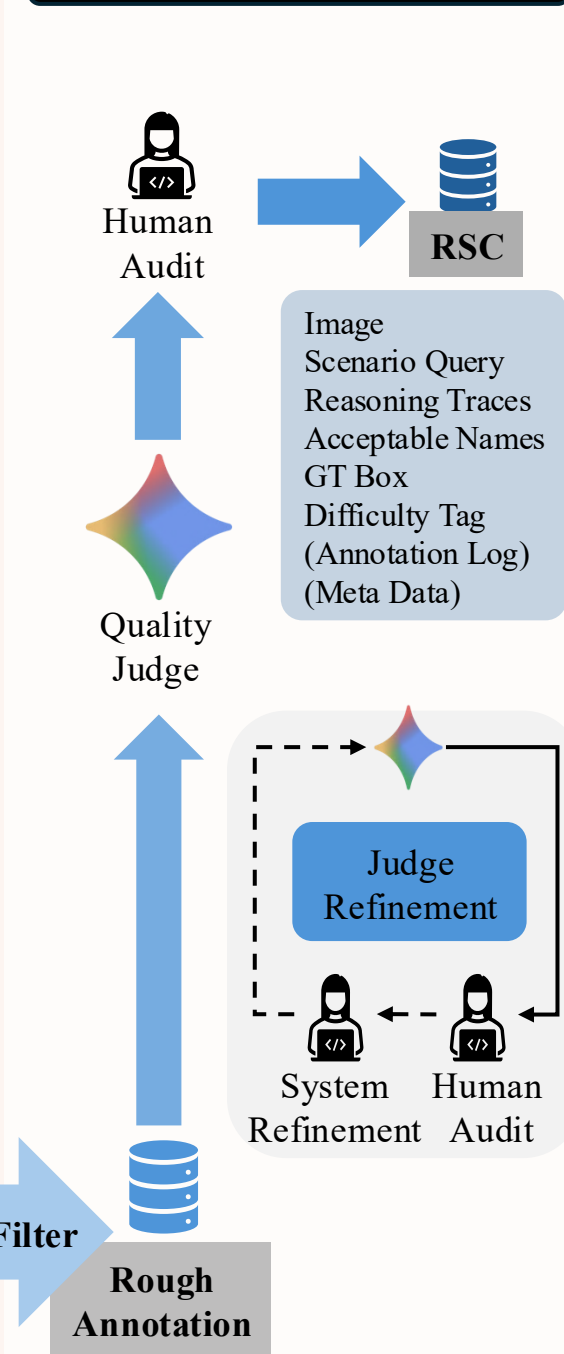


Image
Scenario Query
Reasoning Traces
Acceptable Names
GT Box
Difficulty Tag
(Annotation Log)
(Meta Data)

Judge Refinement
System Refinement
Human Audit

Rough Annotation