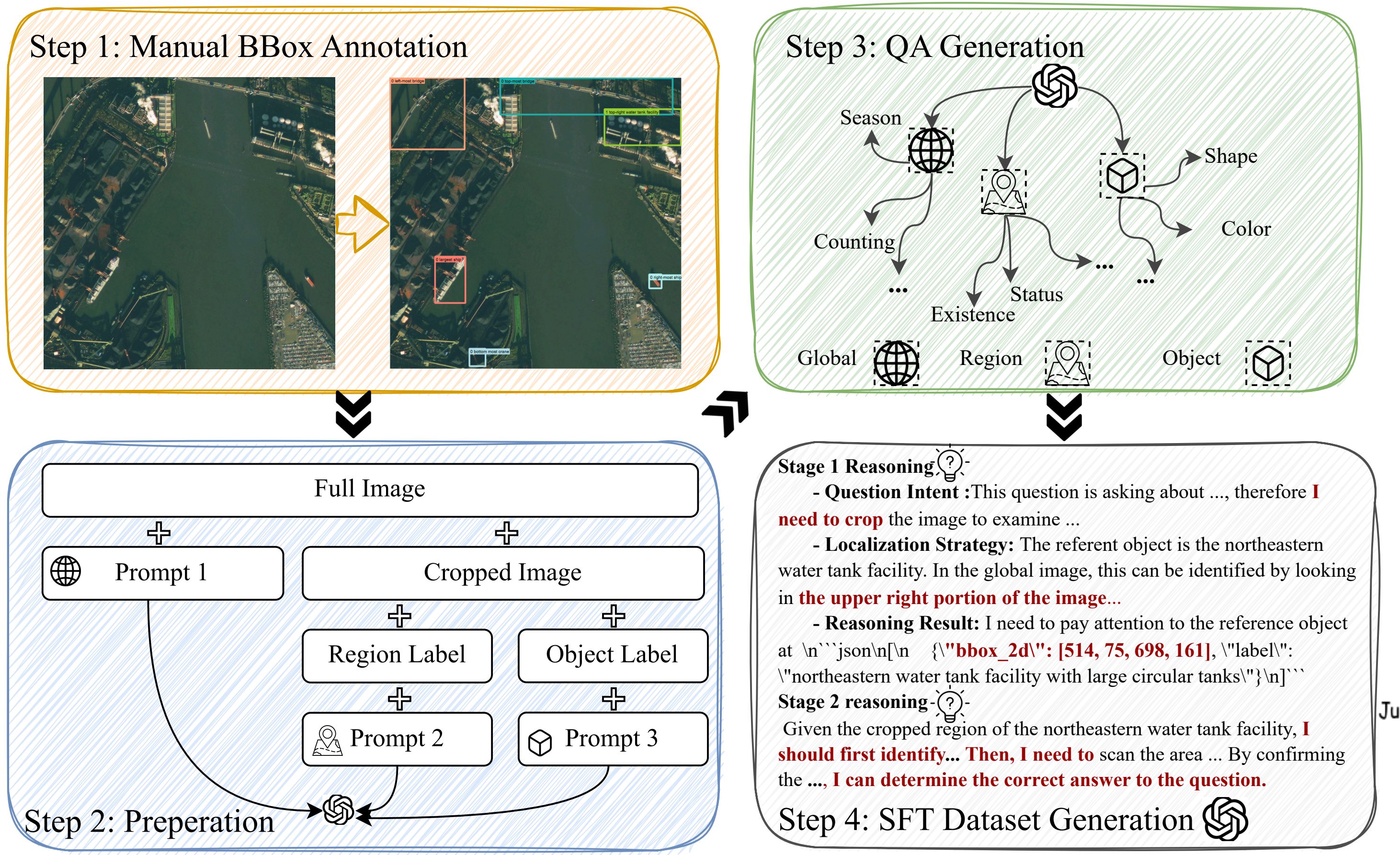


Stage 1 Reasoning

- Question Intent : This question is asking about ..., therefore **I need to crop** the image to examine ...

- Localization Strategy: The referent object is the northeastern water tank facility. In the global image, this can be identified by looking in **the upper right portion of the image**...

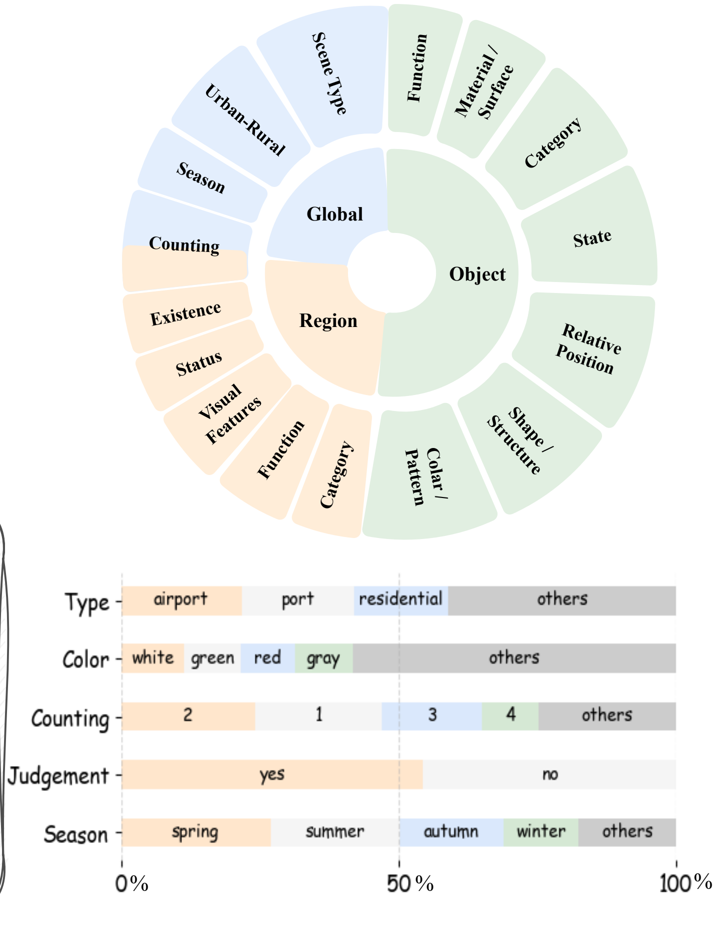
- Reasoning Result: I need to pay attention to the reference object at \n```json\n[\n {**"bbox_2d\": [514, 75, 698, 161]**, \"label\": \"northeastern water tank facility with large circular tanks\"}\n]```

Stage 2 reasoning

Given the cropped region of the northeastern water tank facility, **I should first identify... Then, I need to** scan the area ... By confirming the **...**, **I can determine the correct answer to the question.**

Step 4: SFT Dataset Generation

(a) Data Generation Pipeline



(b) Data Categories and Proportion