Input to DTwinSeger

Step1: Determine the granularity

Step2: Use SAM everything mode

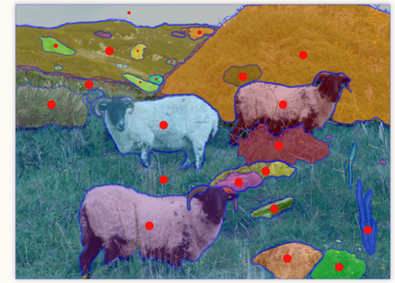
Step3: Compute the centroids

User Query: Segment the goat nearest to the bottom stone

User Image:

It's object level

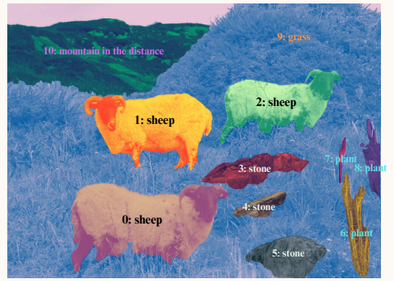
SAM everything



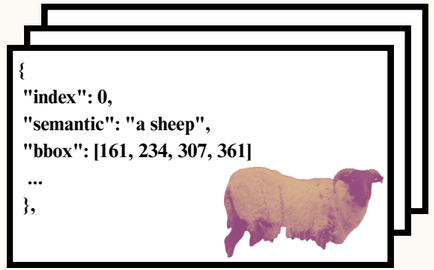
SAM multimask

Step4: Centroids as prompts and select the largest mask

Calculate the center point of each object and identify the bottom stone with index = 5. Then, calculate the distance d between the bottom stone and the sheep with indices 0, 1, and 2. The distances satisfy $d_0 < d_2 < d_1$, meaning that the sheep closest to the bottom stone is Sheep 0.



JSON



Step5: Semantic assignment

Step6: Generate DT representation

Step7: Decided by LLM