

Automatic Spatial Annotation of Robot Trajectories

ACE BRAIN -0.6

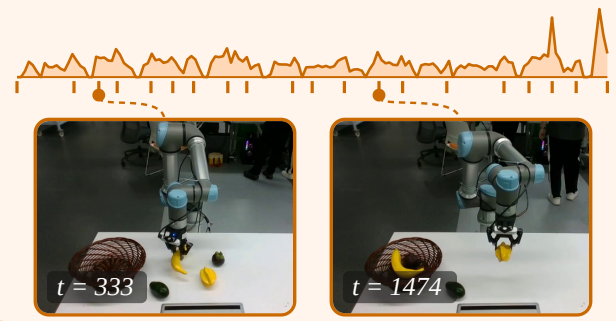
1 Robot Trajectories



INSTRUCTION "Place the four fruits into the nearby basket one by one."

2 Keyframe Selection

Activity score 20 keyframes



3 Object Vocabulary

● Basket ● Banana ● Starfruit
● Mangosteen ● Avocado

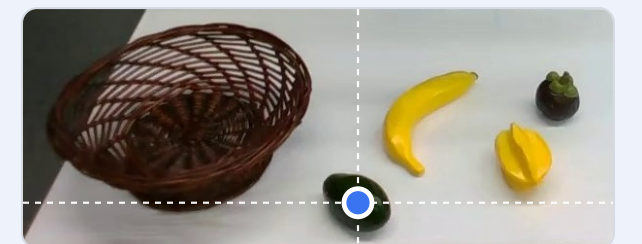
● Mangosteen
"A dark purple round fruit with green top"

4 Spatial Annotation



5 Spatial QA

Q Point to the avocado in the image.



A <point>[[368, 849]]</point>

Spatial Annotation

Keyframe & Descriptions



- **Basket** A brown woven basket sitting on the table
- **Banana** A yellow curved banana lying on the table
- **Starfruit** A yellow star-shaped fruit on the table
- **Mangosteen** A dark purple round fruit with green top
- **Avocado** A dark green oval avocado on the table

LOCALIZATION BRANCH

Grounding DINO



SAM 3



RECOGNITION BRANCH

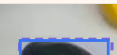
SAM 3

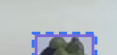
● Basket	0.98	● Banana	0.96
● Starfruit	0.82	● Avocado	0.80
● Mangosteen	0.55		



Semantic Relabeling

Adopt the SAM 3 label if $IoU \geq 0.5$, keep the box

 Mangosteen → Avocado
 $IoU = 0.96 \geq 0.5$

 Avocado → Mangosteen
 $IoU = 0.94 \geq 0.5$

Box count & geometry unchanged

Object Annotations



- 5 Boxes from DINO + SAM 3
- 2 Labels corrected by SAM 3
- 0 Boxes added by label fusion