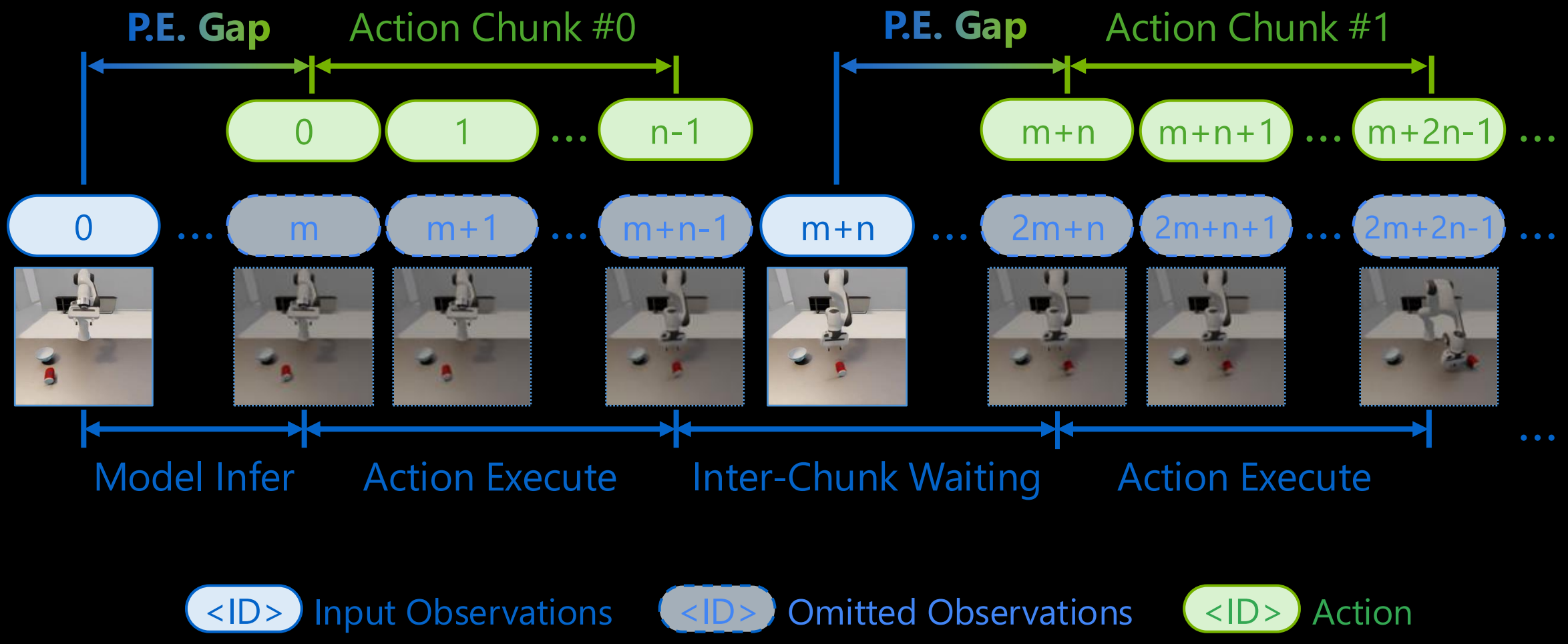
# DynamicVLA

## A Vision-Language-Action Model for Dynamic Object Manipulation

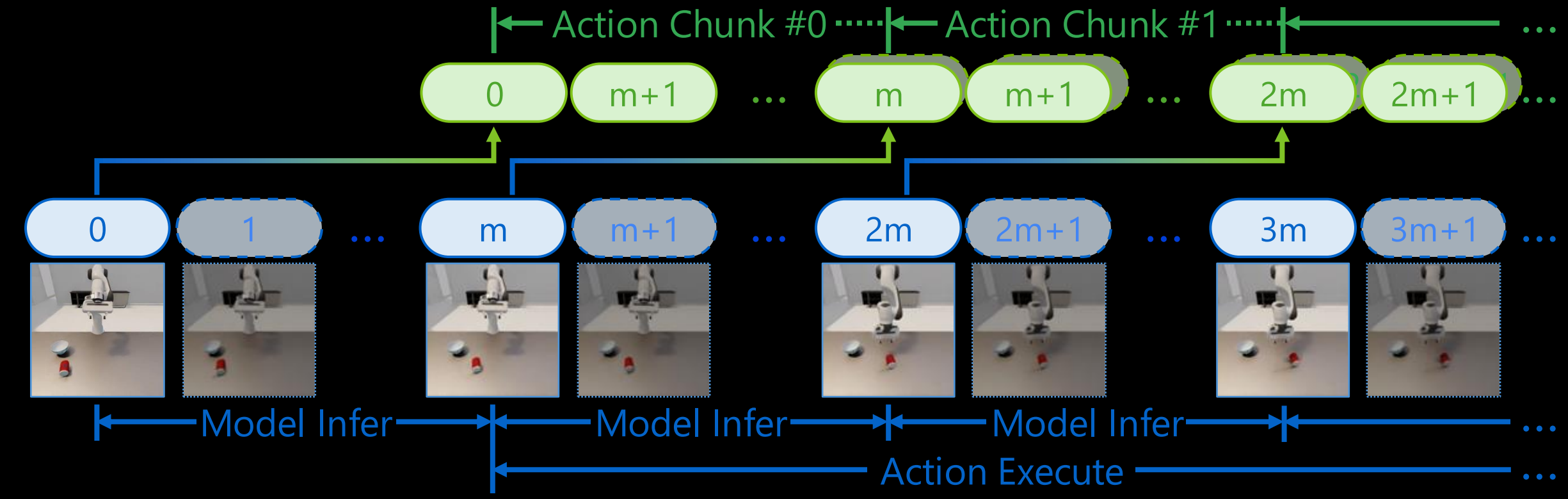
Haozhe Xie  Beichen Wen  Jiarui Zheng  Zhaoxi Chen
Fangzhou Hong  Haiwen Diao  Ziwei Liu

S-Lab, Nanyang Technological University, Singapore

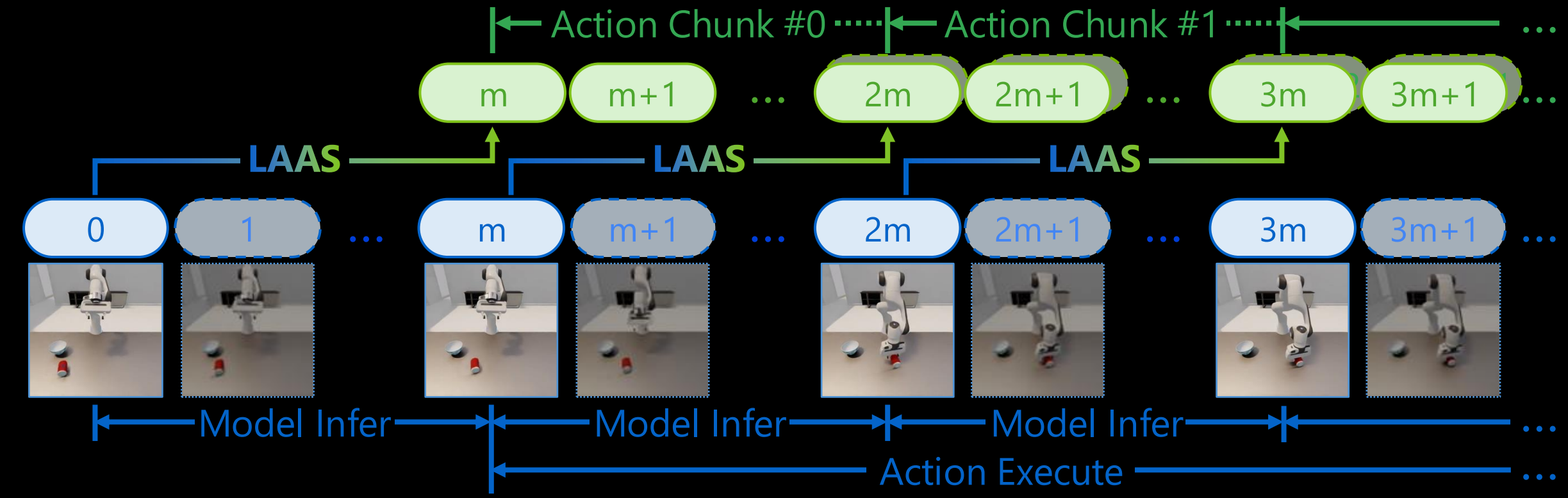
# Latent Analysis for Current VLAs



# Contiguous Inference in DynamicVLA

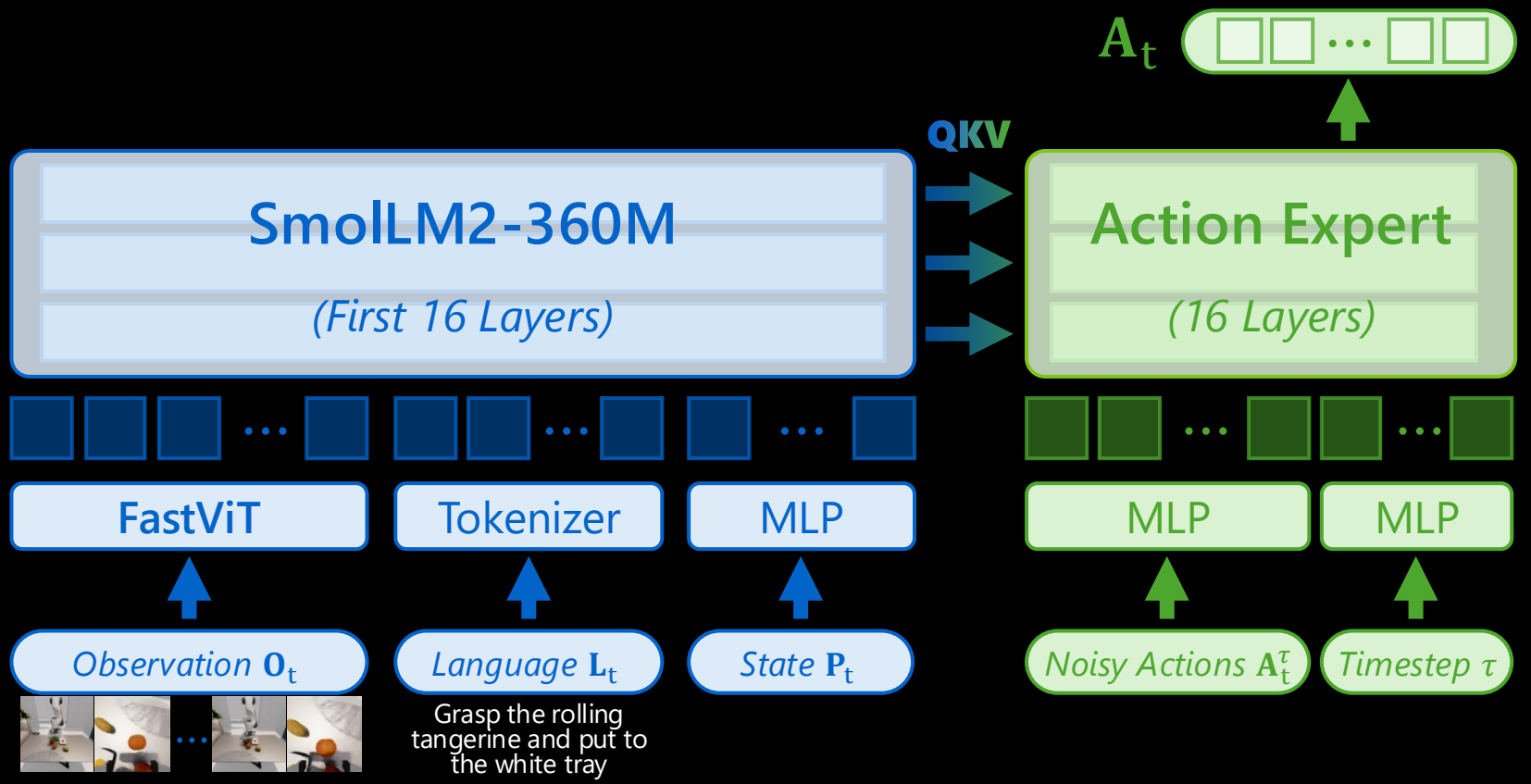


# Latent-aware Action Streaming in DynamicVLA




<ID> Input Observations <ID> Omitted Observations <ID> Action <ID> Omitted Action

# Lightweight VLA Model



# Automatic Data Collection

Simulation

Real-world

**Objects and Dynamics**

#Objects
*206*

Speed
*0–1 m/s*

Friction Coefficient
*0.5-1.5*

#Scenes
*2824*

Lighting
*4000-8000K; 150-750 lm*

Cameras
*f = 2.3mm*
*25 FPS @ 480x360*
*front / side / wrist views*

**Scenes and Sensors**

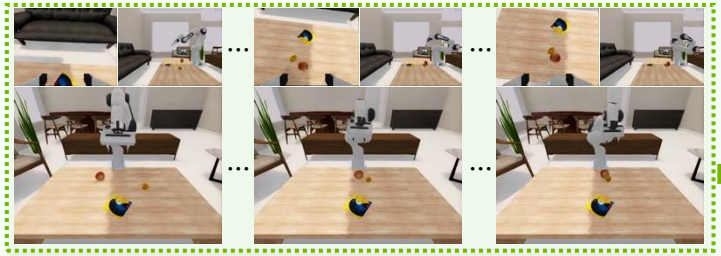
**Objects**

Third-Person Cameras
*Microsoft Azure Kinect DK*
*25 FPS @ 1280x720*
*front / side views*

Wrist Camera
*Intel RealSense D435i*
*25 FPS @ 1280x720*

**Scenes and Sensors**

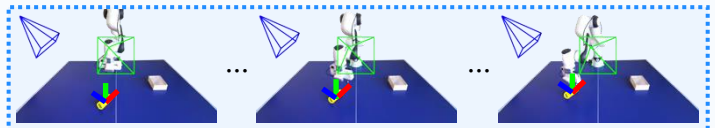
Environment Setup



**Isaac Sim Simulator**

Object Position | Object Rotation | Object Velocity

Real-time 2D Object Tracking



Real-time 6D Object Pose Estimation

**Real-world "Simulator"**

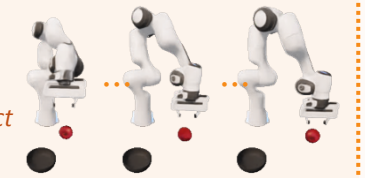
Object Position | Object Rotation | Object Velocity

Object State Acquisition

**(S1)** Approach Object

Actions
- *Move above object*

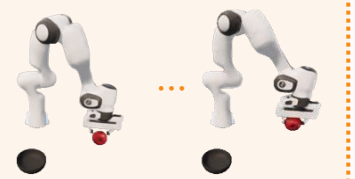
Transition
- *Gripper above object*



**(S2)** Grasp & Lift

Actions
- *Close gripper*
- *Lift the object*

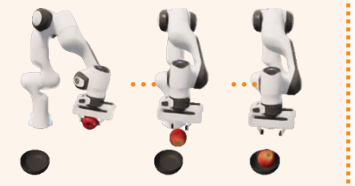
Transition
- *Object grasped*



**(S3)** Approach Target & Place

Actions
- *Move to the target*
- *Release object*

Transition
- *Object placed*



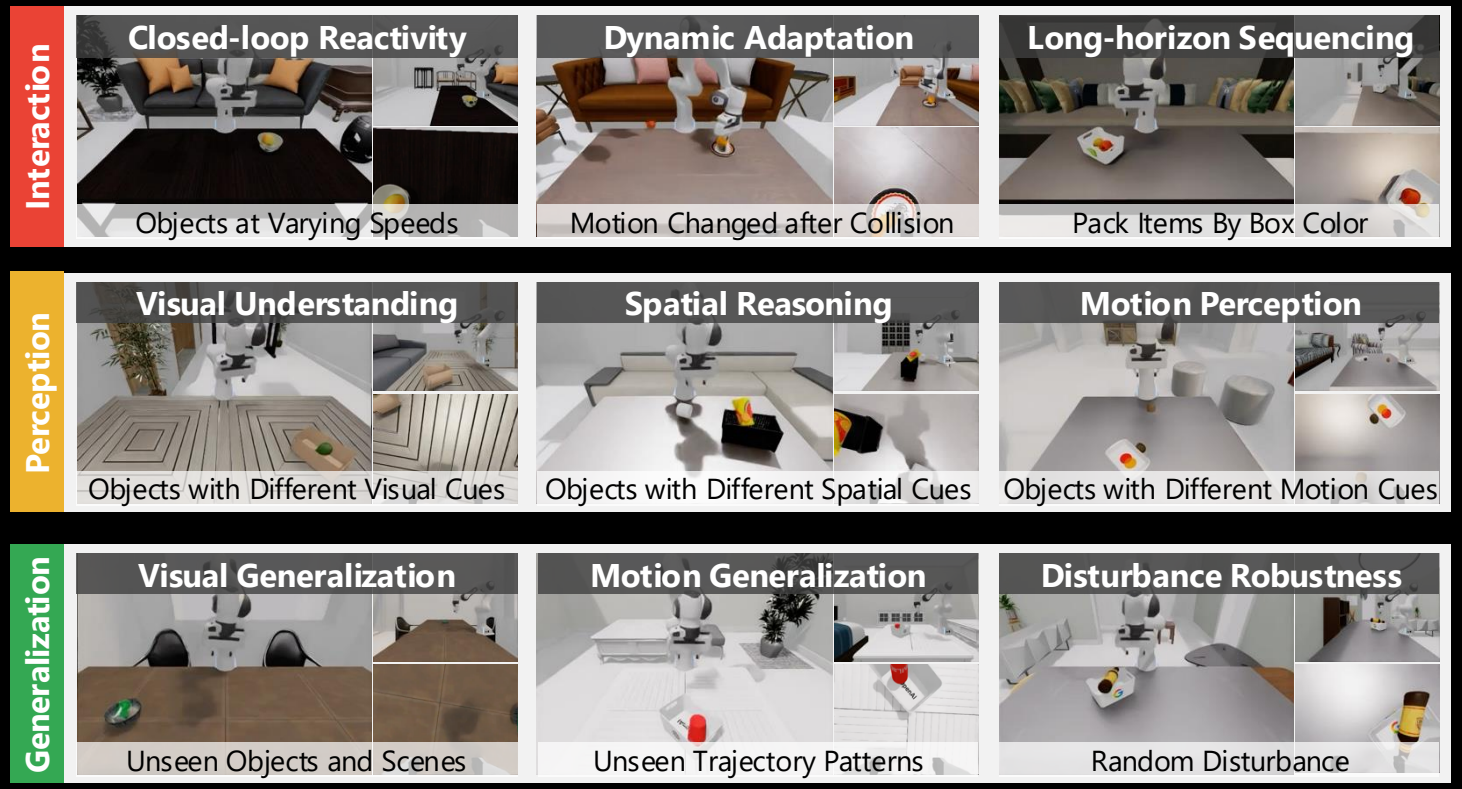
**(S4)** Reset

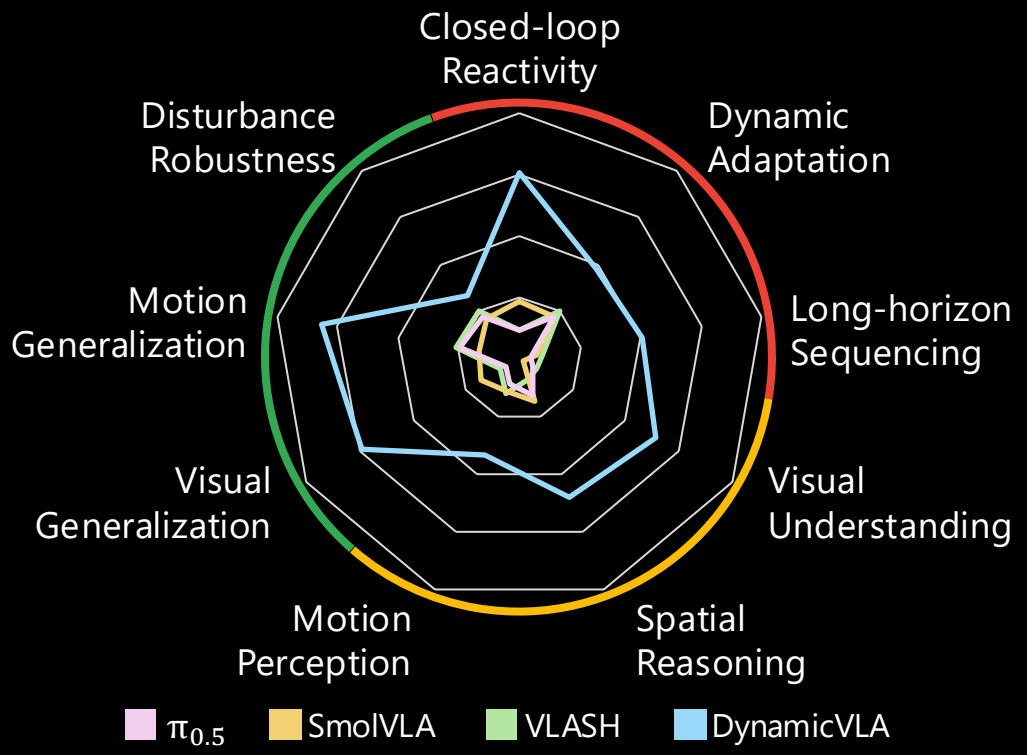
Actions
- *Return to home pose*

*Object drop detected (return to S1)*

State-machine Controller

# Dynamic Object Manipulation Benchmark

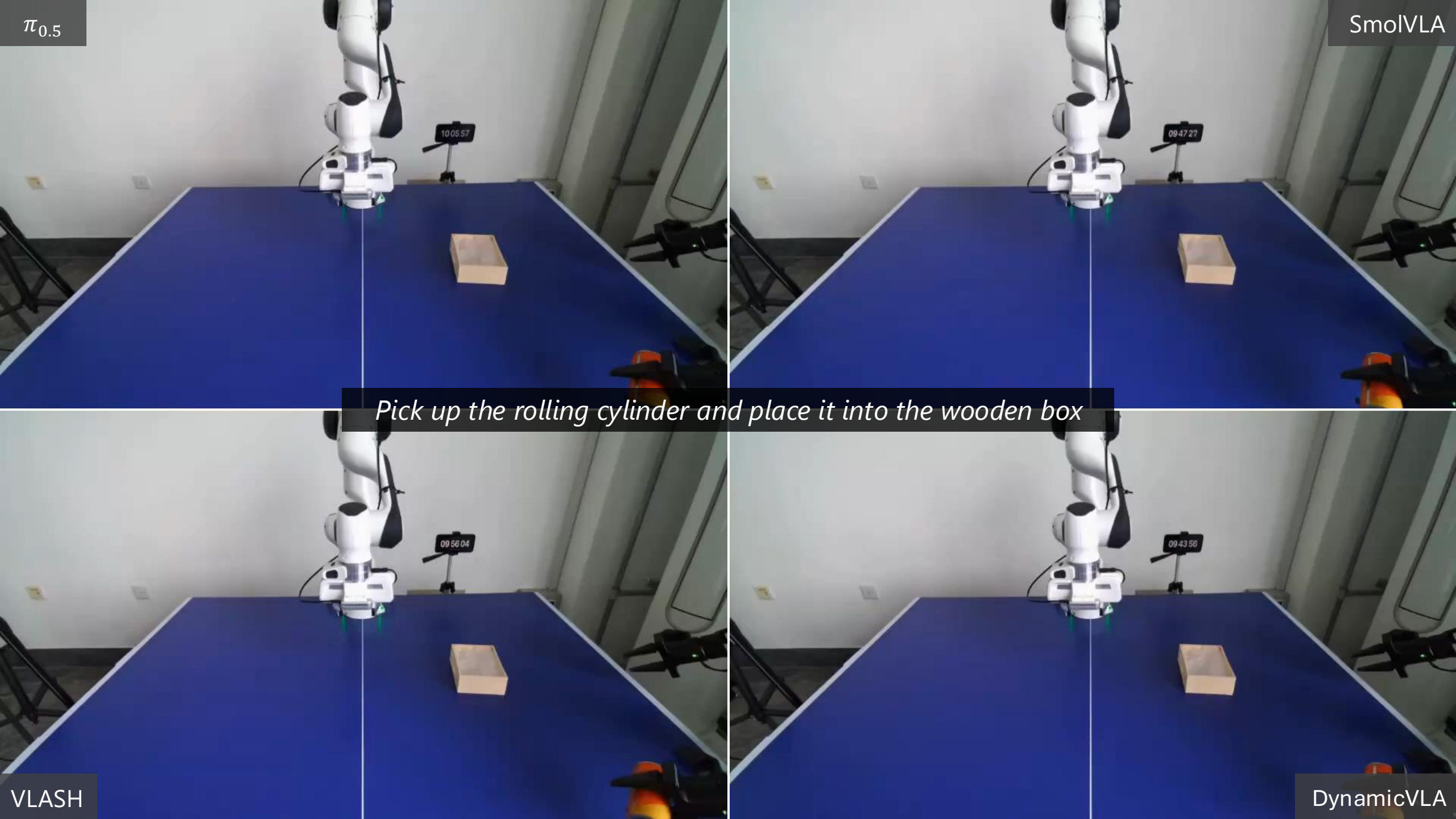


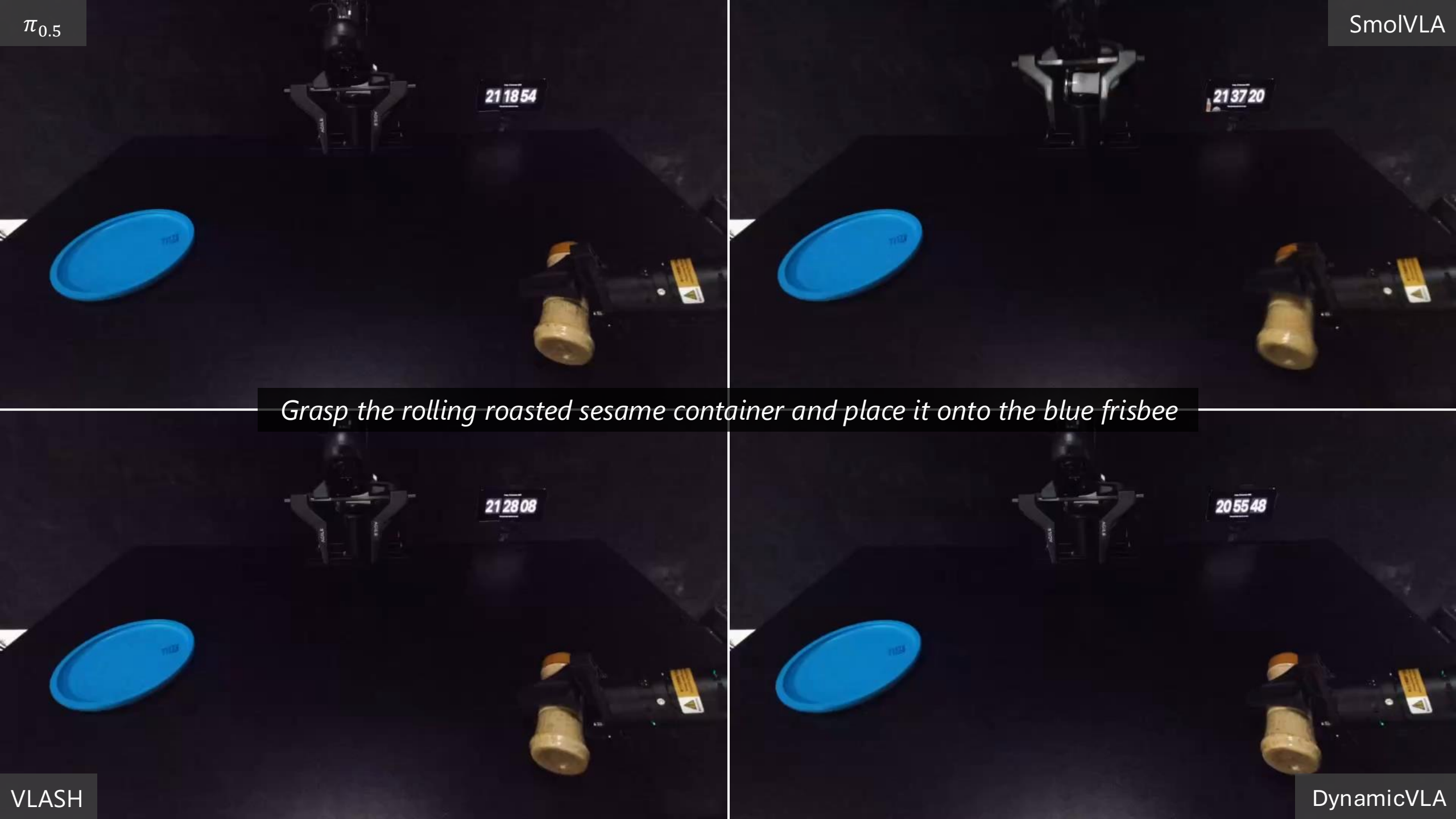




(All actions shown above are generated by DynamicVLA)

# Interaction

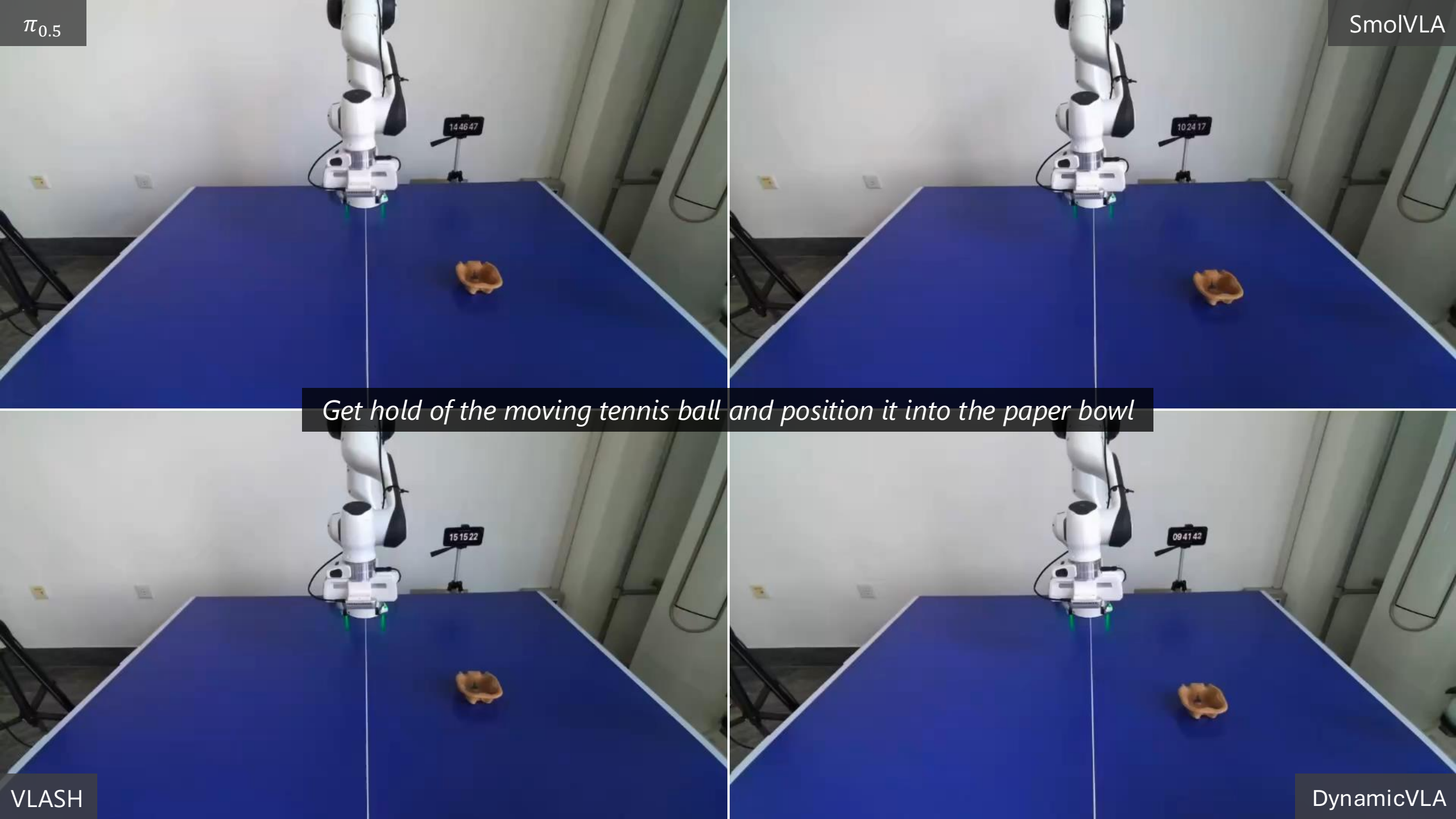
Evaluates closed-loop interaction under evolving object motion

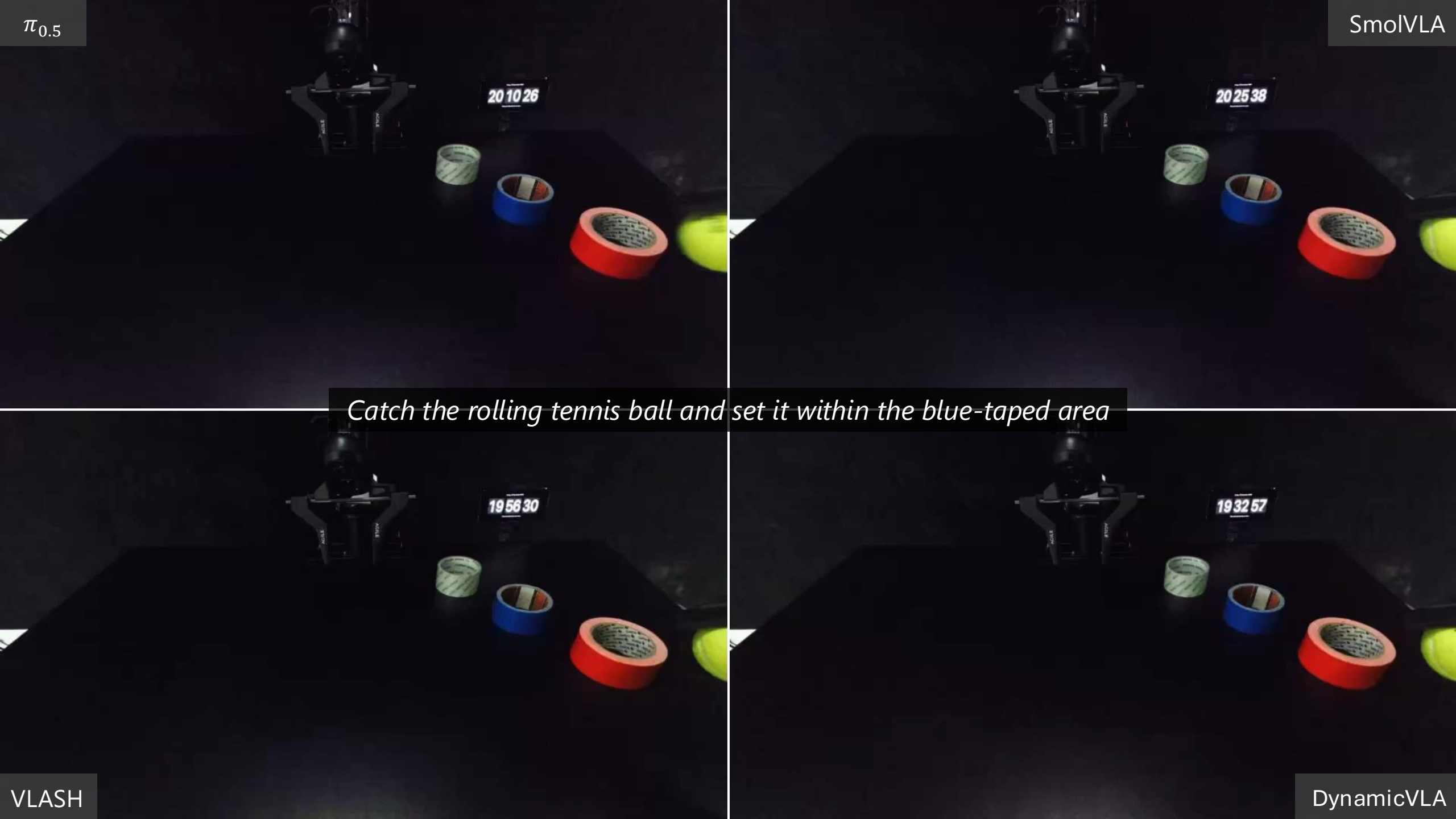

$\pi_{0.5}$
SmolVLA
Pick up the rolling cylinder and place it into the wooden box
VLASH
DynamicVLA


$\pi_{0.5}$
SmolVLA
Grasp the rolling roasted sesame container and place it onto the blue frisbee
VLASH
DynamicVLA

# Perception

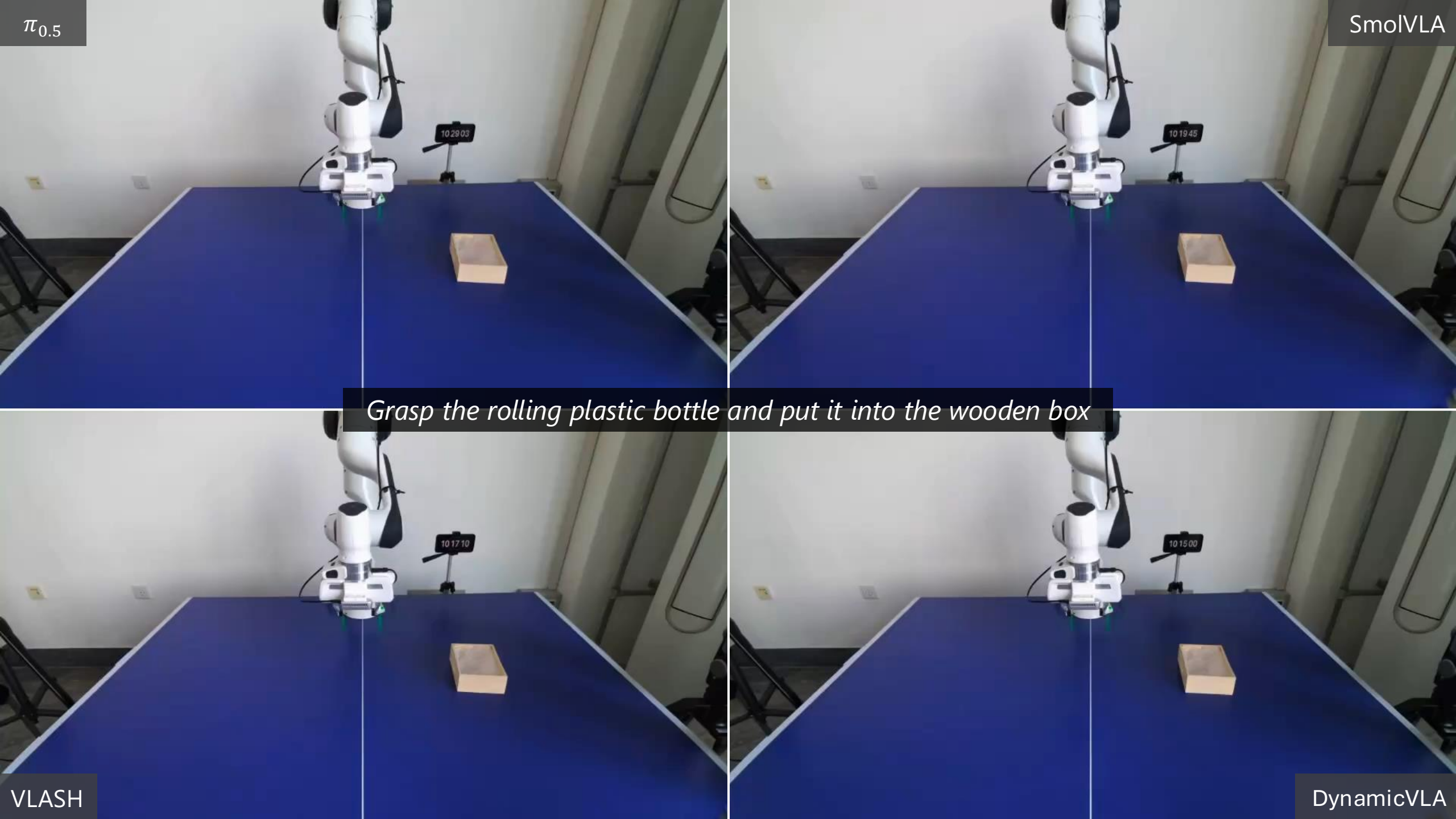
Evaluates perceptual grounding under motion

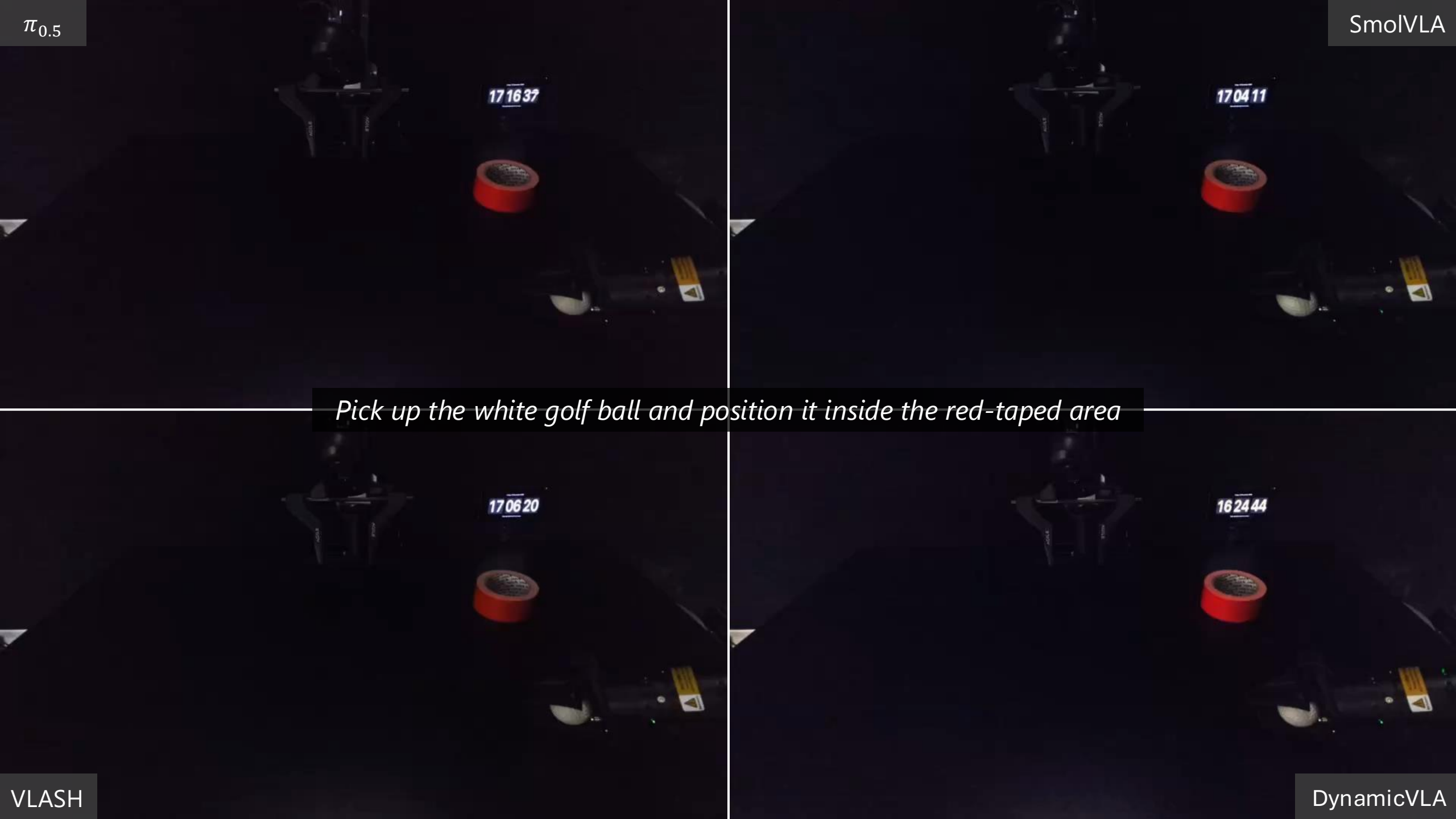

$\pi_{0.5}$
SmolVLA
Get hold of the moving tennis ball and position it into the paper bowl
VLASH
DynamicVLA


$\pi_{0.5}$
SmolVLA
20 10 26
20 25 38
Catch the rolling tennis ball and set it within the blue-taped area
19 56 30
19 32 57
VLASH
DynamicVLA

# Generalization

Evaluates generalization to unseen objects and motion patterns


$\pi_{0.5}$
SmolVLA
Grasp the rolling plastic bottle and put it into the wooden box
VLASH
DynamicVLA


$\pi_{0.5}$
SmolVLA
17 16 37
17 04 11
Pick up the white golf ball and position it inside the red-taped area
17 06 20
16 24 44
VLASH
DynamicVLA

# Thank You!

Project Page

https://haozhexie.com/project/dynamic-vla