KALM: Keypoint Abstraction using Large Models for Object-Relative Imitation Learning

Xiaolin Fang* Bo-Ruei Huang* Jiayuan Mao* Jasmine Shone Joshua B. Tenenbaum Tomás Lozano-Pérez Leslie Pack Kaelbling

Massachusetts Institute of Technology

Website

Our Goal

Robot skills that can generalize to novel object poses, camera views and object instances, with low data collection effort

Method

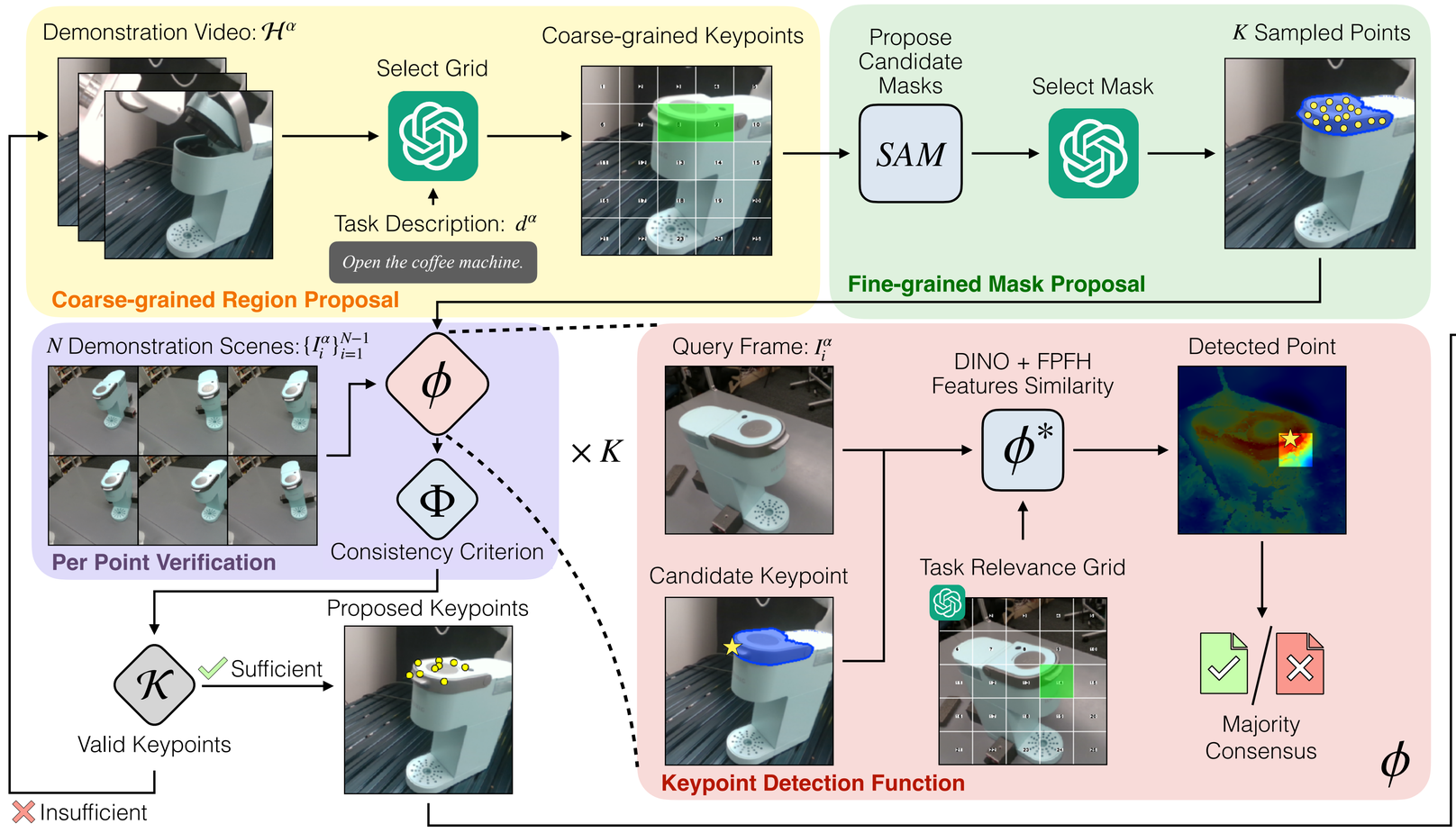
Distill keypoint abstraction by prompting and verifying keypoint proposals from VLMs using a single video and few images (~10)

Keypoint Desiderata

Task Relevance	The keypoints directly support the specific robotic task being performed
Cross-instance Consistency	The keypoints are robustly identifiable across multiple instances

KALM Overview

- Coarse-grained Region Proposal: VLM selects a grid on the image given a task description
- Fine-grained Mask Proposal: Refine the proposal with segmentation model and VLM
- Keypoint Detection Function: Detect the keypoints with DINO and FPFH features similarity
- Per Point Verification: Verify each candidate point is consistent across demonstration scenes

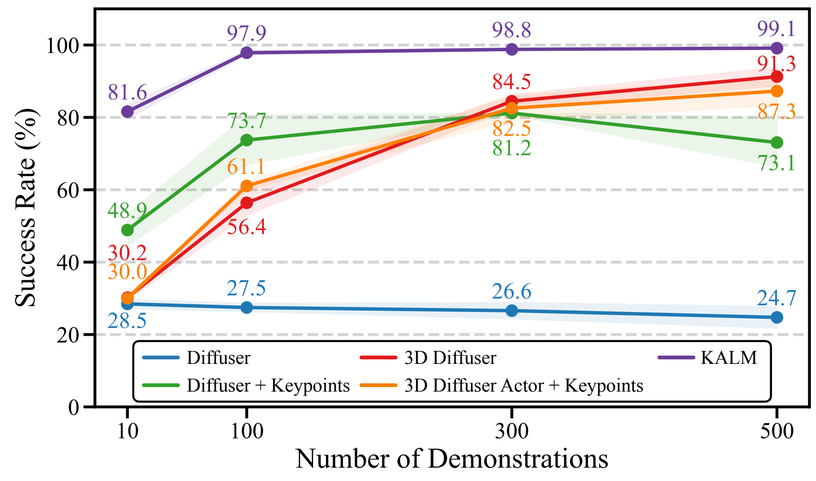


(a) Keypoint Distillation

(b) Inference Time

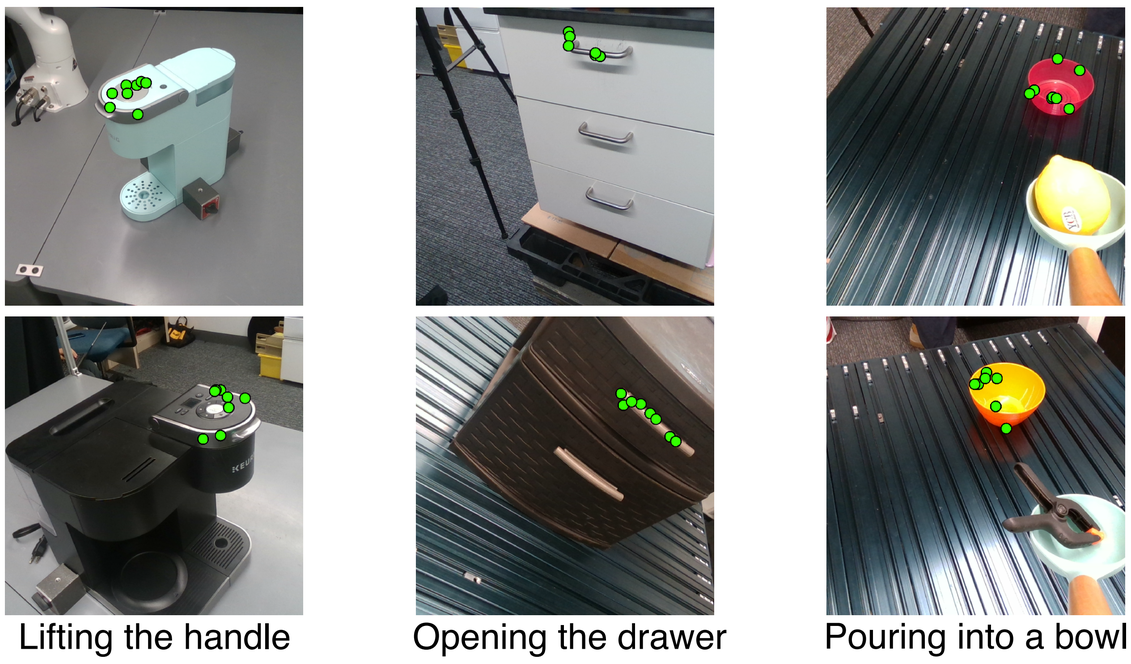
Keypoint Abstraction Improves Data Efficiency

Tasks	Without Keypoints		With Keypoints		
	RGB	RGBD	RGB	RGBD	
	Diffuser	3D Diffuser Actor	Diffuser	3D Diffuser Actor	**KALM (Ours)**
DrawerOpen	30.00 ± 8.29	30.00 ± 6.53	62.00 ± 1.41	29.33 ± 4.64	**77.00 ± 2.94**
DrawerClose	50.00 ± 1.41	53.67 ± 4.19	83.33 ± 2.62	50.67 ± 6.60	**92.33 ± 0.47**
ButtonSide	32.67 ± 2.49	37.67 ± 2.05	49.67 ± 11.09	38.67 ± 1.25	**79.67 ± 1.25**
ButtonTop	19.00 ± 3.56	21.00 ± 4.90	28.00 ± 8.04	21.00 ± 4.32	**97.33 ± 0.47**
LeverPull	10.67 ± 3.30	8.67 ± 2.36	21.33 ± 1.70	10.33 ± 4.19	**61.67 ± 6.13**



Generalization to Novel Views and Objects

Observed images and detected keypoints at testing time



Lifting the handle

Opening the drawer

Pouring into a bowl

Keypoint Verification Procedure Helps Distill Robust and Consistent Keypoints

Tasks	Without Verification		KALM (Ours)	
	View	Cross Object	View	Cross Object
Lifting Handle	1/10	0/10	9/10	6/10
Opening Drawer	2/10	0/10	6/10	7/10
Pouring into Bowl	6/10	2/10	8/10	6/10