ROBOT LEARNING | DIFFUSION POLICIES | TEST-TIME ACTION SELECTION

Look Twice, Act Once

Condition-Sensitive Test-Time Selection for Robot Action Policies

Suraj Reddy Massachusetts Institute of Technology surajrdy@mit.edu

Training-free selector for frozen stochastic action-chunk policies | corrupted-reference contrast | no policy updates

Massachusetts Institute of Technology

CROSS
Condition-sensitive reference-observation sampling

Select the action chunk that is likely under the real observation, but unlikely when task-relevant evidence is corrupted.

+15 pp	57.3%	0
image clean success	keypoint aggregate	policy updates

MOTIVATION AND CONTRIBUTION

Diffusion policies can generate several plausible action chunks for the same scene. The deployment question is not only *which sample is likely*, but which sample is actually **grounded in the current observation**.

Core test. A good action should become less plausible when task-relevant visual/state evidence is corrupted.

CROSS contribution

1. Sample candidates under the real observation.
2. Sample references from corrupted observations using the **same frozen policy**.
3. Select chunks that are likely for the real scene and atypical for corrupted evidence.

no retraining | no reward model | sample access only | frozen policy

CROSS SELECTOR

At each control step t, draw real-observation candidates and corrupted-reference chunks:

$$A^1, \dots, A^K \sim \pi_\theta(\cdot \mid o_t), \qquad \tilde{A}^1, \dots, \tilde{A}^R \sim \pi_\theta(\cdot \mid \tilde{o}_t).$$

Score each real candidate by a local density contrast:

$$s_{\text{CROSS}}(A) = \log \hat{p}_{\text{real}}(A) - \lambda \log \hat{p}_{\text{ref}}(A)$$

with RBF kernel density over flattened action chunks:

$$\hat{p}_{\mathcal{S}}(A) = \frac{1}{|\mathcal{S}|} \sum_{B \in \mathcal{S}} \exp\left(-\frac{\|\text{vec}(A) - \text{vec}(B)\|_2^2}{2\sigma^2} \right).$$

Execute the first action from the selected chunk:

$$A_t^\star = \arg\max_{A \in \mathcal{A}_{\text{real}}} s(A), \qquad a_t^\star = A_t^\star[0].$$

Real-density term
keeps the selected chunk near the candidate cloud produced by the true observation.

Reference penalty
downweights chunks that remain typical after task evidence is removed.

ALGORITHM AND EXPERIMENTS

Selector loop

1. Generate K real candidates and R corrupted references.
2. Estimate real and reference densities in action-chunk space.
3. Penalize candidates that are also dense under corrupted references.
4. Execute the first action of the best-scoring chunk.

Default budgets

Evaluation	Budget	λ
PushT rollouts	$K=4, R=2$	1.0
Franka/GELLO offline	$K=4, R=4$	1.0

Baselines

- Vanilla sampling and Random-K.
- Medoid / density-only selection.
- BID-style temporal coherence and CROSS+BID.

Evaluation
Closed-loop PushT rollouts plus offline Franka/GELLO action-selection MSE diagnostics.

CAVEATS AND DEPLOYMENT NOTES

What the data supports

- CROSS improves clean PushT image success over vanilla sampling.
- On keypoint PushT, the gain is aggregate rather than uniformly dominant in every perturbation.
- Offline Franka/GELLO diagnostics support the corrupted-reference term beyond PushT.

Limitations

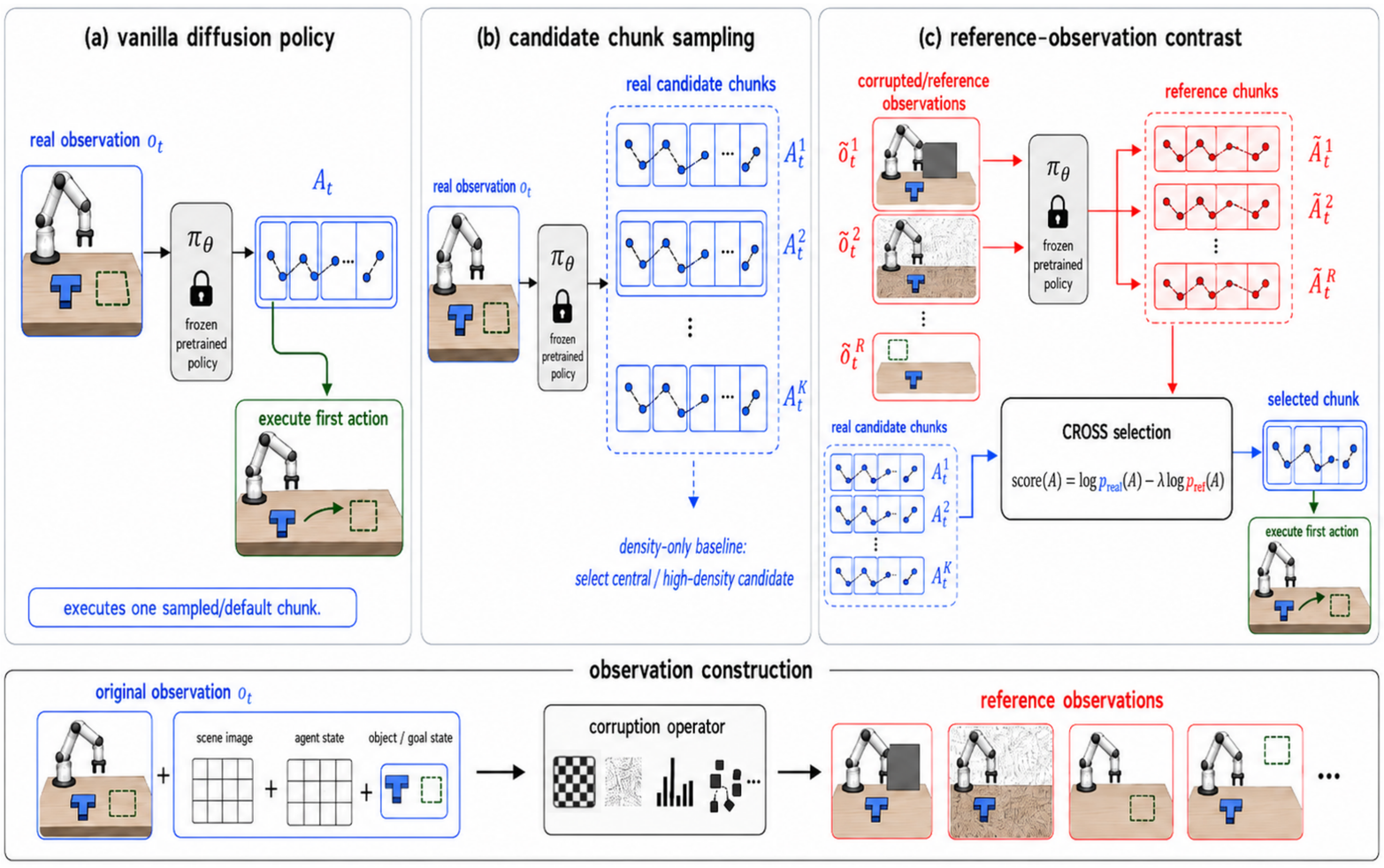
- Closed-loop evidence is still centered on PushT.
- CROSS depends on informative corruptions and candidate diversity.
- Extra samples roughly double PushT step latency in the reported implementation.

RESULTS AT A GLANCE

+15 pp clean image success over vanilla	**57.3%** keypoint aggregate — best aggregate selector
0.821 image max reward — highest clean score	**4 / 5** offline diagnostics — lowest MSE settings

Bottom line. Observation grounding can be tested at inference time by contrasting action density under real and corrupted observations.

PAPER OVERVIEW: VANILLA SAMPLING VS. CROSS



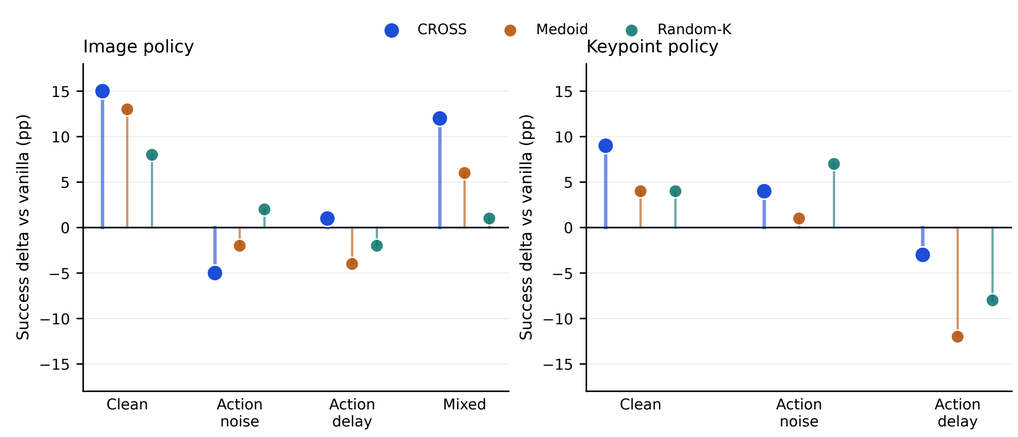
Vanilla diffusion-policy sampling executes a single sampled chunk. CROSS contrasts candidate chunks from the real observation against reference chunks sampled after corrupting task-relevant evidence, and executes the candidate that is likely for the real scene but atypical under the corrupted one.

CLOSED-LOOP PUSHT RESULTS AND ROBUSTNESS

Rollout results (frozen checkpoints)

Selector	Success (%)	Max reward
IMAGE, CLEAN (100 EPISODES)		
Vanilla	21.0	0.782
Random-K	29.0	0.782
Medoid	34.0	0.785
BID	30.0	0.746
CROSS	**36.0**	**0.821**
CROSS+BID	29.0	0.762
KEYPOINT, AGGREGATE (300 EPISODES)		
Vanilla	54.0	0.944
Random-K	55.0	0.947
Medoid	51.7	0.953
BID	52.3	0.947
CROSS	**57.3**	**0.954**
CROSS+BID	54.7	0.942

Paired test (image, clean). CROSS improves over vanilla by $+15.0$ pp on matched seeds, 95% bootstrap CI $[+4.0, +26.0]$, exact paired $p = 0.0167$.

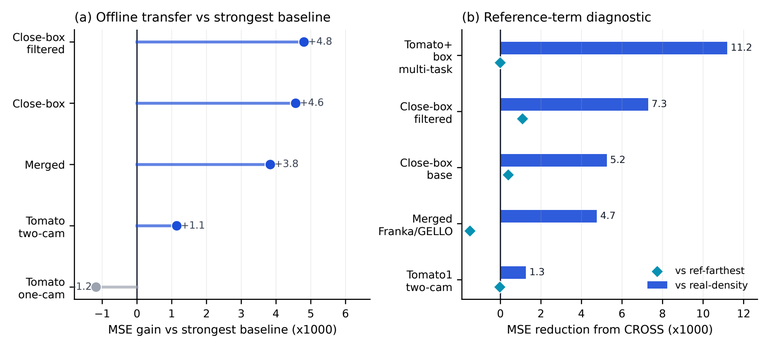


Success deltas relative to vanilla sampling under matched perturbations (100 episodes per cell). CROSS improves clean and mixed image rollouts and clean/action-noise keypoint rollouts.

Where it helps
CROSS targets weak observation grounding. It helps most when candidate chunks are plausible but insufficiently condition-specific, and less directly when actuation noise or action delay dominates.

36.0% image clean CROSS success	**21.0%** vanilla image clean — baseline success	**57.3%** keypoint aggregate — best aggregate selector	**0.821** highest clean max reward

OFFLINE TRANSFER BEYOND PUSHT



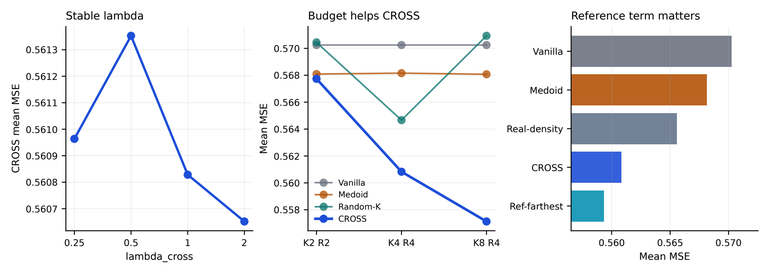
Offline Franka/GELLO action-selection diagnostics: lower selected-action MSE is better. (a) CROSS vs. the strongest non-CROSS selector per diagnostic. (b) CROSS improves over real-density-only selection in all five settings. These are diagnostics, not closed-loop success results.

Selected-action MSE (lower is better)

Diagnostic	Vanilla	CROSS
Merged Franka/GELLO	0.5702	**0.5608**
Close-box base	0.6455	**0.6351**
Close-box filtered	0.5585	**0.5496**
Tomato two-cam	0.6096	**0.6066**
Tomato one-cam	0.6787	0.6770

4 / 5 lowest MSE among all selectors	**5 / 5** beats real-density-only — reference term adds signal

ABLATIONS AND INFERENCE COST



Ablations over sampling budget, reference weight, and corruption mode on the merged Franka/GELLO diagnostic.

Sensitivity
Increasing the budget from $K{=}2, R{=}2$ to $K{=}8, R{=}4$ improves MSE from 0.5677 to 0.5571, a modest gain for the added sampling. A sweep over $\lambda \in \{0.25, \dots, 2.0\}$ is comparatively flat. Corruption choice matters most:

Corruption mode	MSE
Zero-all	**0.5497**
Dropout-dims	0.5590
Mixed (default)	0.5608
Noise-all	0.5663

Inference cost

Policy	Vanilla	CROSS
Image	0.114 s/step	0.227 s/step
Keypoint	0.012 s/step	0.024 s/step

Practical interpretation. CROSS is a better selector, not a faster sampler; use it when extra test-time sampling is acceptable.

CROSS: Condition-Sensitive Test-Time Selection for Robot Action Policies
Frozen stochastic policy + corrupted-reference sampling + density contrast. Training-free and reward-model-free.
Acknowledgments: Thanks to Nathan Zhao for providing compute resources and the Masason Foundation for research support.

Contact: surajrdy@mit.edu
Web: surajmreddy.com