Enabling VLA Action Self-Verification via VLM Token Probability Bucketing

Chen Zhao[1,2*], Zhuoran Wang[1,2], Haoyang Li[1,2*], Shifeng Bao[1,2*], Guanlin Li[1,2*], Youhe Feng[1,2*], Yang Li[1,2], Jie Tang[3], Jing Zhang[1,2†]

[1]Renmin University of China [2]Key Laboratory of Data Engineering and Knowledge Engineering, MOE, China [3]Tsinghua University
*Work done as interns at Zhipu AI. †Corresponding author.

A VLA that verifies its own actions.
Continuous actions → token probabilities

98.4%
LIBERO success
$\pi_{0.5}$ + TokenPB

+17.1
Seen real-world task-score points[a]

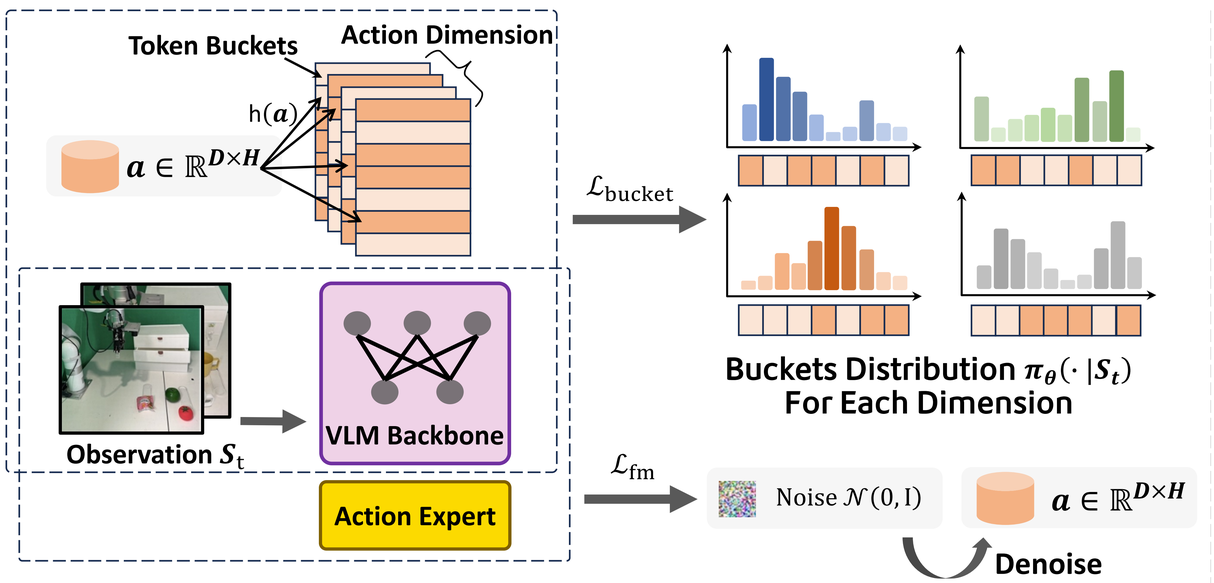
+16.4
Unseen real-world task-score points[a]

1 Why self-verification?

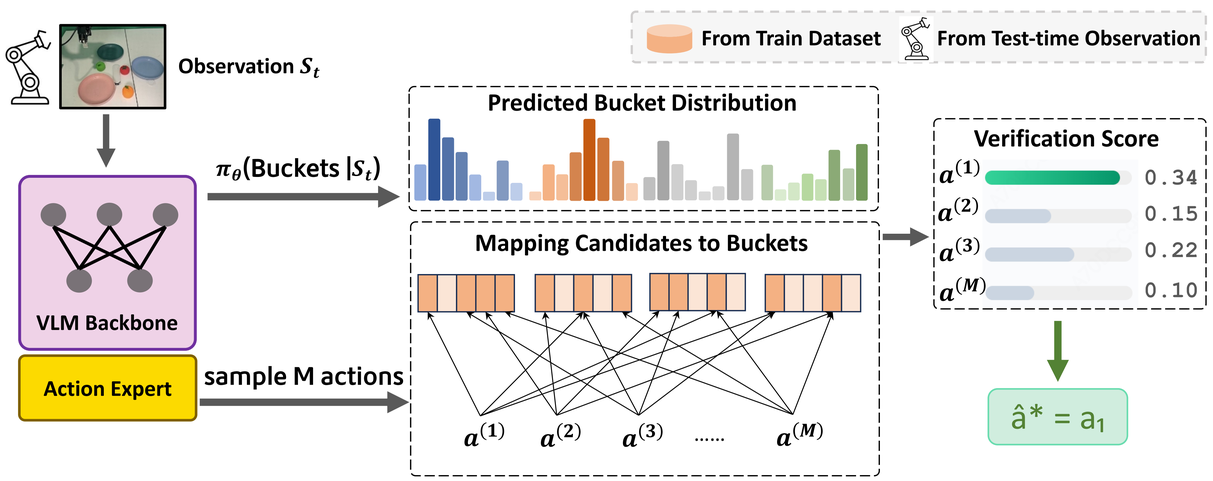
Sampling more actions helps only with a good selector. TokenPB turns continuous-action verification into token-probability estimation inside **one VLA checkpoint**.

2 TokenPB framework

Co-train action generation and bucket prediction



Sample candidates, score tokens, execute the best



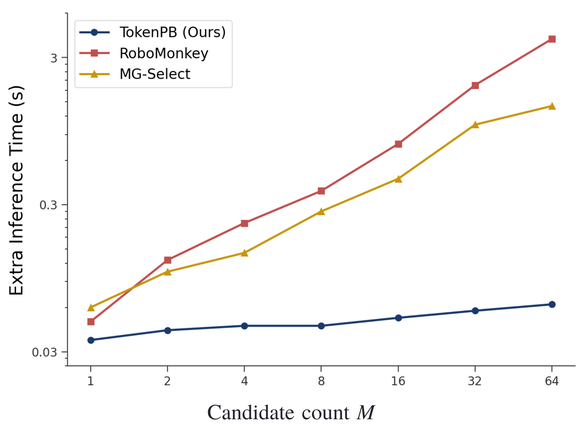
3 Main results

Same co-trained checkpoint: single sample (M = 1) vs. TokenPB (M = 16).

Benchmark	Model	Success Rate (%)					
LIBERO		**Spatial**	**Object**	**Goal**	**Long**		**Avg.**
	Qwen2.5-VL-GR00T	94.6	97.2	93.8	90.2		93.9
	Qwen2.5-VL-GR00T+TokenPB	96.6	98.8	97.0	94.6		96.8
	InternVL3.5-GR00T	95.8	96.0	94.2	92.8		94.7
	InternVL3.5-GR00T+TokenPB	97.6	**99.4**	97.2	96.8		97.8
	$\pi_{0.5}$	97.8	98.4	96.6	93.0		96.5
	$\pi_{0.5}$+TokenPB	**98.4**	**99.4**	**98.2**	**97.6**		**98.4**
CALVIN ABC→D		**Task 1**	**Task 2**	**Task 3**	**Task 4**	**Task 5**	**Avg. Len.**
	Qwen2.5-VL-GR00T	90.8	81.2	74.9	68.7	59.5	3.751
	Qwen2.5-VL-GR00T+TokenPB	93.1	87.4	81.6	75.9	71.3	4.093
	$\pi_{0.5}$	93.5	86.7	79.6	75.3	71.1	4.071
	$\pi_{0.5}$+TokenPB	**95.4**	**90.6**	**85.3**	**79.6**	**75.8**	**4.267**
RoboTwin 2.0		**Block Rank RGB**	**Open Laptop**	**Adjust Bottle**	**.....**		**Avg.**
		Easy / Hard	*Easy / Hard*	*Easy / Hard*			*Easy / Hard*
	Qwen2.5-VL-GR00T	38.0 / 11.0	39.0 / 18.0	96.0 / 16.0	…		30.1 / 6.7
	Qwen2.5-VL-GR00T+TokenPB	**52.0 / 24.0**	**55.0 / 27.0**	**97.0 / 26.0**	…		**39.3 / 10.2**

4 Efficiency and real-world validation

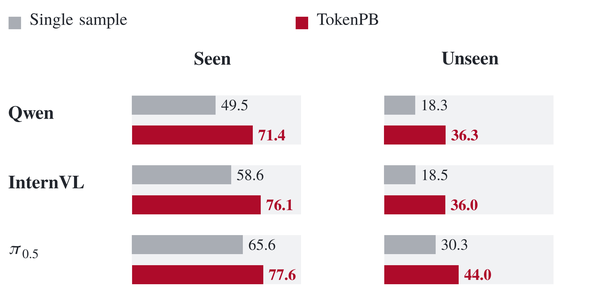
Extra inference time

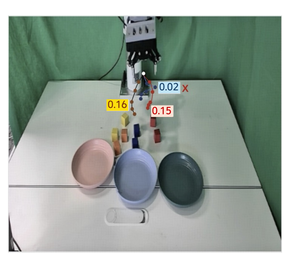


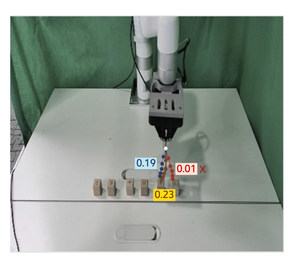
Reuse the observation-prefix KV cache. Score all candidates with batched teacher forcing.

Reported latency; candidate-token compute still grows with M. No external verifier is needed.

Real-world task score (%)







Completion score allows partial credit.
Unseen: held-out instructions / configurations.
Image labels: candidate bucketing scores.

[a]Mean task-score gain across Qwen2.5-VL, InternVL3.5, and $\pi_{0.5}$.

2021201645@ruc.edu.cn zhang-jing@ruc.edu.cn