World–Value–Action Model: Implicit Planning for Vision–Language–Action Systems

Runze Li[1*], Hongyin Zhang[1*], Junxi Jin[1], Qixin Zeng[1], Zifeng Zhuang[1], Yiqi Tang[1], Shangke Lyu[2†], Donglin Wang[1†]

[1] Westlake University [2] Nanjing University

NeurIPS 2026

Headline Numbers

75.6%

Real-world average success

Three tasks · 15 trials per task

+40.0 pp vs GE-ACT

98.1% LIBERO average

+2.6 pp on Long

Why planning?

Acting now requires reasoning about what happens next.

Direct action prediction offers limited lookahead. Errors accumulate across long manipulation tasks.

Why world + value?

FEASIBLE FUTURES

Predict what could happen.

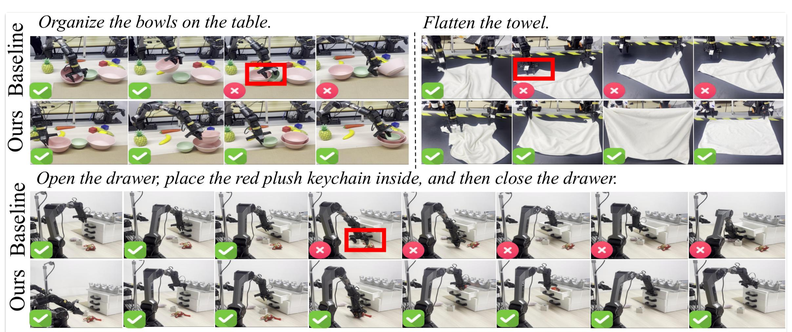
+

VALUABLE FUTURES

Choose useful outcomes.

Latent generation focuses exploration. Value guidance improves selection.

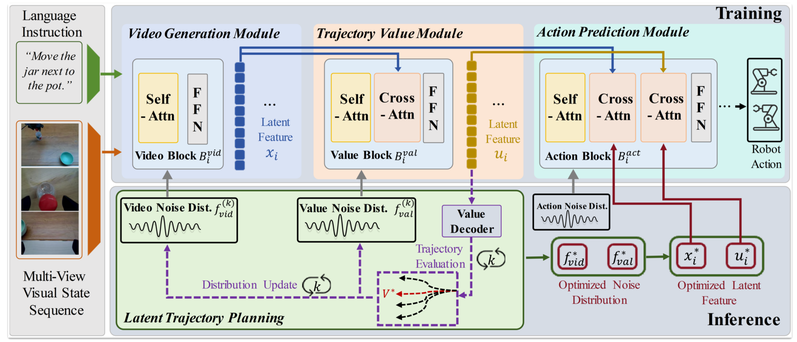
Multi-step robot execution



GE-ACT (Baseline) vs WAV (Ours)

Plan in latent space

1. Generate candidate future video features.
2. Evaluate trajectory values and select elite samples.
3. Refine future priors, repeat, then decode actions.



World, value, and action modules; source Figure 1.

$$\phi^{(m)} = \max_n \frac{\mathbb{E}[v^{(m,n)}]}{\mathrm{Std}[v^{(m,n)}] + \epsilon}$$

Score futures by the most reliable value estimate.

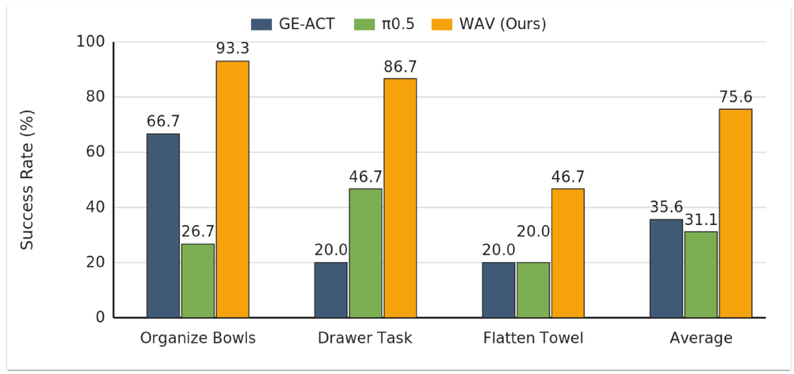
K = 3 iterations balance success and inference cost.

Does planning matter?

INFERENCE STRATEGY	AVG(%).	LONG(%)
No planning	96.4	91.8
Matched Best-of-N	97.2	93.6
WAV	**98.1**	**94.4**

Appendix Table 2. Same frozen modules, forward-pass counts matched for Best-of-N and WAV.

Real-world task completion



Three tasks, 15 trials each; full completion only.

+40.0 pp Real-world average vs GE-ACT

LIBERO simulation

METHOD	AVG.	LONG
GE-ACT	96.5	94.4
π0.5	96.8	92.4
WAV (2.2B)	**98.1**	**94.4**

Table 1 success (%). WAV leads on average and matches the best Long score.

Takeaway & trade-off

Predict futures. Evaluate value. Refine priors. Then act.

Higher task success comes with additional inference time and GPU memory. The paper uses **K = 3** as an efficiency compromise.