WORLD LAUREATES FORUM
世界顶尖科学家论坛
2026
国际工程智能大会
2026 International Conference on AI for Engineering
World Laureates Foundation
上海世界顶尖科学家发展基金会
工程智能研究院
INSTITUTE OF AI FOR ENGINEERING

RAC-WAM: Region-wise Adversarial Consistency for Visual Robustness in World-Action Models

Lihua Que[1], Jing Wang[2], Qingshan Xu[1], Yingwei Pan[2], Sheng Zhou[3], Weidong Chen[1], Xun Yang[1], Ting Yao[2], Tao Mei[2]

[1] University of Science and Technology of China, Hefei, China, [2] HiDream.ai, Beijing, China, [3] King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

01 ABSTRACT

World-action models remain vulnerable to visual shifts. RAC-WAM regularizes action and future-video predictions with region-wise adversarial perturbations, improving robustness without extra inference cost.

Regional sensitivity

At matched RMS, background perturbations induce 1.6–2.3× the action discrepancy of robot/object perturbations.

02 METHOD

Sample a camera scope, then search robot, object and background regions separately. Allocate a shared RMS budget by sensitivity.

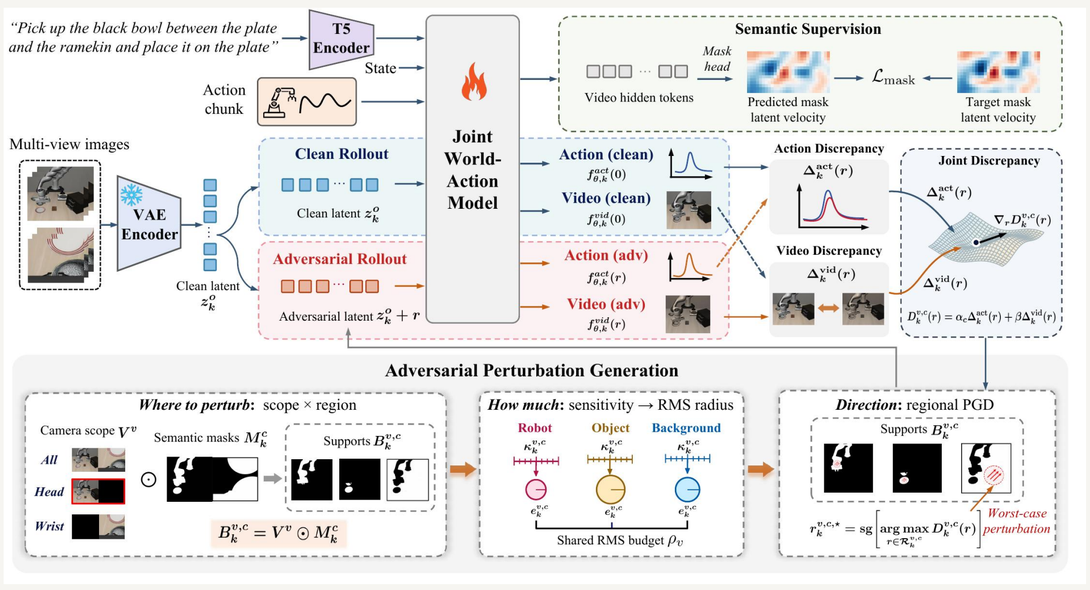


Figure 1. RAC-WAM improves visual robustness through region-wise adversarial training guided by action and video predictions.

Training only

3-step PGD and auxiliary mask supervision.
No extra demonstrations or inference modules.

03 RESULTS

LIBERO-Plus: 82.8% zero-shot success +10.6 percentage points over the baseline.

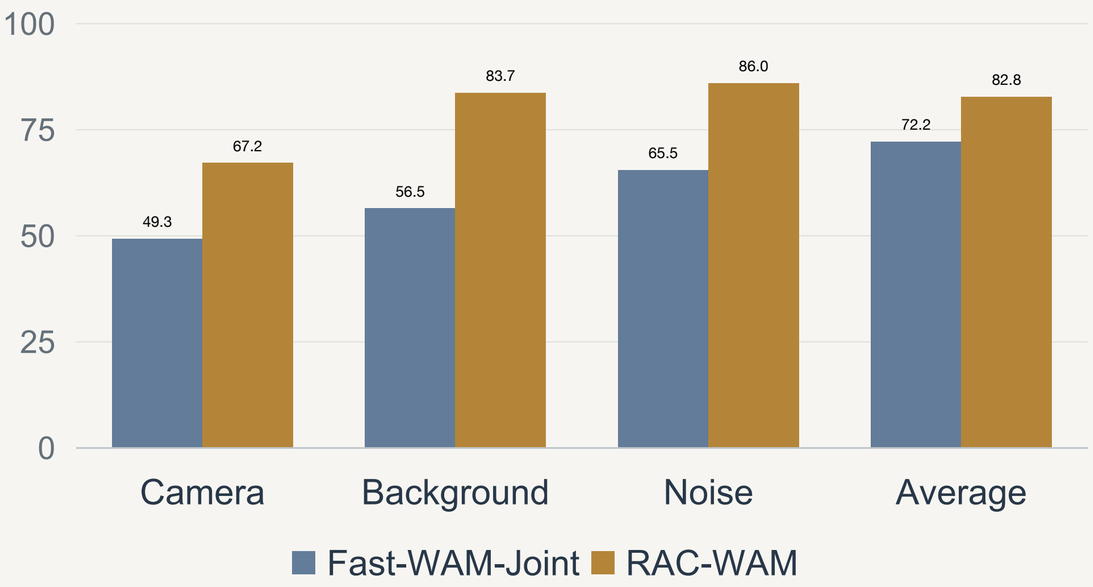
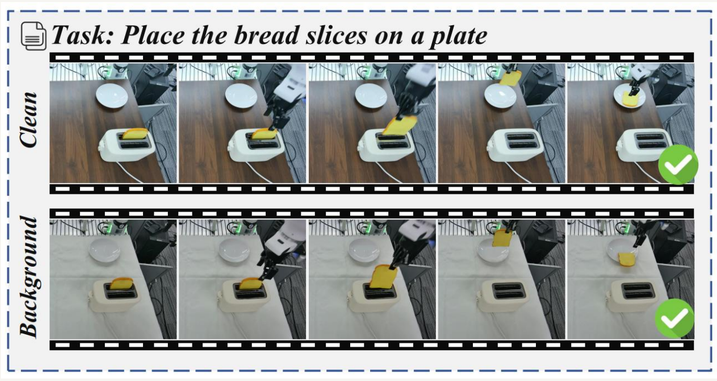


Figure 2. Success rate (%). Selected shifts and overall average. Zero-shot evaluation after LIBERO training.

Table 1. Success rate (%) vs. Fast-WAM-Joint

Benchmark	Baseline	RAC-WAM	Gain (pp)
LIBERO	97.7	98.3	+0.6
RoboTwin 2.0	94.5	97.8	+3.3
Franka robust	**17.5**	**35.4**	**+17.9**

Real-world manipulation on Franka



- 3 tasks
- 4 visual shifts
- 20 trials per task–condition.

04 DISCUSSION

Real-world distractors remain challenging.
Robust Franka success is 35.4%, leaving substantial room for improvement.

05 CONCLUSIONS

Region-aware action–video consistency improves robustness across simulation and real robots with no additional inference cost.

CONTACT

Lihua Que, qlihua@mail.ustc.edu.cn