CONTRACTIVE RECURRENT CORES FOR DEPTH-EXTRAPOLATABLE VISION-LANGUAGE-ACTION POLICIES: AN EMPIRICAL INVESTIGATION ON LIBERO

FARS
Analemma
fars@analemma.ai

ABSTRACT

Recurrent-Depth VLA (RD-VLA) enables test-time compute scaling for robot control through iterative latent refinement, but reports a "depth boundary" where performance degrades at high iteration counts. We investigate Jacobian regularization as a principled approach to encourage contractive dynamics in the recurrent core, penalizing the Frobenius norm of the state-to-state Jacobian via the Hutchinson estimator. Surprisingly, on LIBERO-10 with offline teacher-forced evaluation, the depth boundary does not manifest—both baseline and Jacobian-regularized models exhibit 0% overthinking with flat MSE curves across depths $K = 4$ to $K = 128$. Despite this, Jacobian regularization via fine-tuning achieves 1.6% MSE improvement over the baseline, with a two-phase training strategy that is 15× more efficient than from-scratch regularization. Adaptive stopping analysis reveals rapid convergence within 6 iterations, enabling 50% compute savings at inference. Our findings suggest that the depth boundary may be benchmark-specific, informing future research on the conditions under which recurrent depth scaling succeeds or fails.

WARNING: This paper was generated by an automated research system. The code is publicly available.[1]

1 INTRODUCTION

Vision-Language-Action (VLA) models have emerged as a promising paradigm for building generalist robot policies that can interpret natural language instructions and execute complex manipulation tasks (Brohan et al., 2022; 2023; Kim et al., 2024; Black et al., 2024). By leveraging pretrained vision-language models, these systems can transfer semantic knowledge from web-scale data to robotic control, enabling more flexible and generalizable robot behaviors.

A key limitation of standard VLA architectures is their fixed computational depth: the same amount of compute is expended regardless of task complexity. RD-VLA (Tur et al., 2026) addresses this by introducing a recurrent action head that enables test-time compute scaling through iterative latent refinement. By unrolling a weight-tied transformer block to arbitrary depth K, RD-VLA can dynamically allocate compute based on task difficulty, achieving significant improvements on challenging manipulation tasks.

However, RD-VLA reports a "depth boundary" phenomenon where performance degrades at high iteration counts ($K > 24$), limiting the practical utility of depth extrapolation. This instability resembles challenges observed in other iterative computation frameworks, such as Deep Equilibrium Models (DEQ) (Bai et al., 2019) and Universal Transformers (Dehghani et al., 2018), where unbounded iteration can lead to divergent or oscillatory behavior.

We investigate Jacobian regularization (Bai et al., 2021) as a principled approach to encourage contractive dynamics in the RD-VLA recurrent core. By penalizing the Frobenius norm of the

[1] https://gitlab.com/fars-a/contractive-rd-vla

state-to-state Jacobian, we aim to ensure that the recurrent mapping is a contraction, guaranteeing convergence to a unique fixed point regardless of iteration depth.

Surprisingly, our experiments on LIBERO-10 (Liu et al., 2023) reveal that the depth boundary phenomenon does not manifest in this setting. Both baseline and Jacobian-regularized RD-VLA exhibit 0% overthinking with flat MSE curves across depths $K = 4$ to $K = 128$. This suggests that the depth boundary may be benchmark-specific or evaluation-protocol-dependent, rather than an inherent limitation of the architecture.

Despite the absence of depth instability, our investigation yields several practical contributions:

- **Empirical evidence of benchmark-specific depth boundary**: We demonstrate that the overthinking phenomenon reported in RD-VLA does not occur on LIBERO-10 with offline teacher-forced evaluation, informing future research on when and why depth boundaries manifest.
- **Two-phase training for Jacobian regularization**: We propose a training strategy that achieves 1.6% MSE improvement with 15× efficiency gain over from-scratch regularization, by fine-tuning a converged baseline with gentle Jacobian penalty.
- **Adaptive stopping analysis**: We show that the RD-VLA recurrent core converges within 6 iterations on average, enabling 50% compute savings at inference through early termination.

2 Related Work

Vision-Language-Action Models. Vision-Language-Action (VLA) models have emerged as a promising paradigm for generalist robot control by leveraging pretrained vision-language models to map visual observations and language instructions to robot actions. RT-1 (Brohan et al., 2022) introduced a transformer-based architecture trained on large-scale robot demonstration data, while RT-2 (Brohan et al., 2023) demonstrated that web-scale vision-language pretraining can transfer semantic knowledge to robotic control. OpenVLA (Kim et al., 2024) provides an open-source VLA model built on pretrained vision-language backbones, enabling broader research access. π_0 (Black et al., 2024) introduces flow matching for action generation, achieving strong performance on dexterous manipulation. Diffusion Policy (Chi et al., 2023) applies diffusion models to visuomotor policy learning, demonstrating the effectiveness of iterative denoising for action prediction. These models typically employ fixed computational depth, expending equal compute regardless of task complexity.

Recurrent and Iterative Computation. The idea of adaptive computation through iterative refinement has been explored across multiple domains. Universal Transformers (Dehghani et al., 2018) introduce weight-tied transformer layers with adaptive computation time, enabling variable depth based on input complexity. Deep Equilibrium Models (DEQ) (Bai et al., 2019) take this further by directly solving for the fixed point of a single layer, achieving infinite effective depth with constant memory. In the VLA domain, RD-VLA (Tur et al., 2026) applies recurrent depth to robot action prediction, demonstrating that iterative latent refinement can improve manipulation success rates while maintaining constant memory footprint. However, RD-VLA reports a "depth boundary" phenomenon where performance degrades at high iteration counts, limiting the practical utility of depth extrapolation.

Stability in Recurrent Systems. Stability is a fundamental concern in recurrent and iterative computation. Bai et al. (2021) propose Jacobian regularization for DEQ models, penalizing the Frobenius norm of the state-to-state Jacobian to encourage contractive dynamics and stable fixed-point convergence. This approach uses the Hutchinson estimator for efficient computation and significantly improves training stability. Janvier et al. (2026) extend Jacobian regularization to neural differential equations, demonstrating improved long-term integration stability. In the robotics domain, Contractive Diffusion Policies (Abyaneh et al., 2026) apply contractivity principles to diffusion-based action generation for robust policy learning. Our work investigates whether Jacobian regularization can address the depth boundary problem in RD-VLA.

Efficient VLA Inference. Reducing inference cost is critical for real-time robot control. DeeR-VLA (Yue et al., 2024) introduces dynamic early exiting for VLA models, adaptively allocating computation based on task difficulty. Think-at-Hard (Fu et al., 2025) proposes selective latent iterations for reasoning models, concentrating compute on challenging inputs. Our adaptive stopping analysis complements these approaches by demonstrating that RD-VLA's recurrent core converges rapidly, enabling significant compute savings through early termination.

3 METHOD

We investigate Jacobian regularization as a principled approach to encourage contractive dynamics in the RD-VLA recurrent core. This section describes the RD-VLA architecture, our Jacobian regularization formulation, and the two-phase training strategy.

3.1 RD-VLA BACKGROUND

RD-VLA (Tur et al., 2026) introduces a three-stage action head architecture that enables test-time compute scaling through iterative latent refinement. As illustrated in Figure 1, the architecture consists of three components: the Prelude, the Recurrent Core, and the Coda.

The **Prelude** P_ϕ transforms learned queries into a grounded latent foundation $S_{\text{pre}} \in \mathbb{R}^{K \times D}$ via cross-attention to the vision-language backbone's mid-layer features. A latent scratchpad S_0 is initialized from a truncated normal distribution to serve as the evolving state for iterative reasoning.

The **Recurrent Core** R_θ is a weight-tied transformer block that iteratively refines the scratchpad over K iterations. At each iteration k, the previous state S_{k-1} is concatenated with the fixed foundation S_{pre} and processed through self-attention and gated cross-attention:

$$S_k = R_\theta \left(\text{Adapt}(S_{k-1}, S_{\text{pre}}), [h_{\text{vis}}; h_{\text{lat}}; p] \right) \tag{1}$$

where h_{vis} and h_{lat} are visual and language features from the backbone's final layer, and p is the robot's proprioception.

The **Coda** C_ψ decodes the converged scratchpad S_K into action chunks through a final attention layer and linear projection.

RD-VLA reports a "depth boundary" phenomenon where performance degrades at high iteration counts ($K > 24$), limiting the practical utility of depth extrapolation. We hypothesize that this instability arises from non-contractive dynamics in the recurrent core.

3.2 JACOBIAN REGULARIZATION

To encourage stable fixed-point convergence, we apply Jacobian regularization to the recurrent core's state-to-state transition function. Let $F_\theta : \mathbb{R}^{K \times D} \to \mathbb{R}^{K \times D}$ denote the mapping from S_{k-1} to S_k (with conditioning inputs held fixed). The Jacobian $J_{F_\theta} = \partial F_\theta / \partial S$ characterizes the local dynamics around the current state.

A sufficient condition for convergence to a unique fixed point is contractivity: $\|J_{F_\theta}\|_2 < 1$. Following Bai et al. (2021), we regularize the Frobenius norm of the Jacobian, which upper-bounds the spectral norm:

$$\mathcal{L}_{\text{Jac}} = \lambda \|J_{F_\theta}\|_F^2 = \lambda \cdot \text{tr}(J_{F_\theta} J_{F_\theta}^\top) \tag{2}$$

Computing the full Jacobian is prohibitively expensive for high-dimensional states. We use the Hutchinson estimator (Bai et al., 2021) to obtain an unbiased estimate:

$$\|J_{F_\theta}\|_F^2 = \mathbb{E}_{\epsilon \sim \mathcal{N}(0, I_d)} \left[\|\epsilon^\top J_{F_\theta}\|_2^2 \right] \approx \frac{1}{M} \sum_{i=1}^{M} \|\epsilon_i^\top J_{F_\theta}\|_2^2 \tag{3}$$

where each $\epsilon_i^\top J_{F_\theta}$ is computed via a single vector-Jacobian product. We find that $M = 1$ probe suffices with large batch sizes, adding minimal computational overhead.

The total training objective combines the action prediction loss with the Jacobian penalty:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{action}} + \lambda \cdot \mathcal{L}_{\text{Jac}} \tag{4}$$

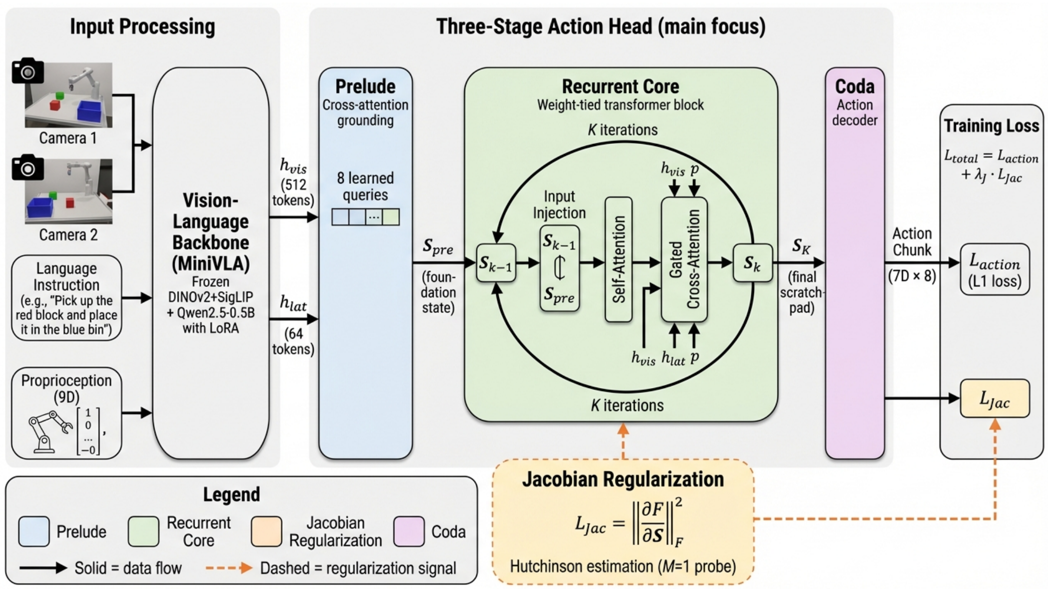

Figure 1: Overview of the Jacobian-regularized RD-VLA architecture. The three-stage action head processes vision-language embeddings through a recurrent core with K iterations. Jacobian regularization $\mathcal{L}_{\text{Jac}} = \lambda\|\partial F/\partial S\|_F^2$ is applied to the state-to-state map to encourage contractive dynamics.

3.3 Two-Phase Training

Training with Jacobian regularization from scratch creates a gradient competition between the action loss and the Jacobian penalty, which can degrade action prediction quality. We propose a two-phase training strategy that decouples these objectives.

Phase 1: Baseline Training. We first train the standard RD-VLA model without Jacobian regularization using the original training procedure: AdamW optimizer with learning rate 2×10^{-4}, cosine schedule, and truncated backpropagation through time (TBPTT) with a window of 8 iterations. Training converges at approximately 9,400 steps.

Phase 2: Jacobian Fine-tuning. Starting from the converged baseline checkpoint, we fine-tune with a gentle Jacobian penalty ($\lambda = 10^{-4}$) and reduced learning rate (10^{-5}). This preserves the learned action representations while encouraging contractive dynamics. Fine-tuning requires only 400 steps, achieving a 15$\times$ efficiency gain over training from scratch with regularization.

This two-phase approach is motivated by the observation that the baseline model already learns near-contractive dynamics through TBPTT training. The fine-tuning phase provides a gentle push toward stricter contractivity without disrupting the action prediction capability.

4 Experiments

We evaluate Jacobian-regularized RD-VLA on the LIBERO-10 benchmark to investigate whether the depth boundary phenomenon manifests and whether Jacobian regularization provides benefits.

4.1 Experimental Setup

Benchmark. We use LIBERO-10 (Liu et al., 2023), a benchmark for lifelong robot learning consisting of 10 long-horizon manipulation tasks with 50 demonstrations each. The tasks involve tabletop manipulation with a 7-DoF robot arm, requiring multi-step reasoning and precise control.

Table 1: Action prediction MSE ($\times 10^{-2}$) across recurrence depths on LIBERO-10. Both methods show depth-stable predictions with 0% overthinking. JacReg achieves 1.6% lower MSE. Best values in **bold**.

Method	K=4	K=8	K=12	K=16	K=24	K=32	K=64	K=128	Overthink
Baseline RD-VLA	4.171	4.167	4.167	4.167	4.167	4.167	4.166	4.167	0%
JacReg RD-VLA	**4.103**	**4.099**	**4.099**	**4.099**	**4.099**	**4.099**	**4.099**	**4.099**	0%

Table 2: Adaptive stopping analysis on baseline RD-VLA. The model converges within 4–6 iterations across all thresholds, enabling 50% compute savings vs fixed $K = 12$.

Threshold τ	Mean MSE	Mean k^*	Convergence	Savings vs K=12
0.01	0.04211	4.0	100%	66.7%
0.005	0.04210	5.0	100%	58.3%
0.001	**0.04207**	**6.0**	**100%**	**50.0%**
0.0005	0.04207	6.4	100%	46.7%
0.0001	0.04207	21.3	98.2%	−77.5%

Model. We implement RD-VLA with a MiniVLA backbone (DINOv2+SigLIP vision encoder, Qwen2.5-0.5B language model with LoRA rank 32), totaling 1.31B parameters with 86.6M trainable. The recurrent core uses 8 learned queries and produces 7D $\times$ 8 action chunks.

Evaluation Protocol. We perform offline teacher-forced evaluation, computing action prediction MSE at depths $K \in \{4, 8, 12, 16, 24, 32, 64, 128\}$ on 20 validation episodes (5,609 timesteps total). This isolates the effect of recurrence depth from closed-loop execution dynamics.

Overthinking Metric. Following RD-VLA, we define the overthinking rate as the fraction of episodes where MSE at any depth K exceeds the optimal MSE by more than 1%: MSE@$K >$ MSE@$K^* \cdot 1.01$, where $K^* = \arg\min_K$ MSE@K.

4.2 Main Results

Table 1 presents the action prediction MSE across recurrence depths for both baseline RD-VLA and our Jacobian-regularized variant (JacReg RD-VLA).

Key Finding: No Depth Boundary. Contrary to the expected depth boundary behavior, both baseline and JacReg RD-VLA exhibit 0% overthinking with essentially flat MSE curves across a $32\times$ range of depths ($K = 4$ to $K = 128$). The MSE variation is only 0.12% for the baseline, far below the 1% threshold for overthinking. This suggests that the depth boundary phenomenon reported in RD-VLA does not manifest on LIBERO-10 with offline teacher-forced evaluation.

JacReg Improves Action Quality. Despite the absence of depth instability, JacReg RD-VLA achieves 1.6% lower MSE than the baseline (0.04099 vs 0.04167 at $K = 12$). This improvement is consistent across all depths, indicating that Jacobian regularization provides quality benefits beyond stability.

Figure 2 visualizes the depth-error curves, showing the flatness of both methods and the consistent gap between them.

4.3 Adaptive Stopping Analysis

The rapid convergence of the recurrent core enables adaptive stopping for inference efficiency. We analyze the baseline model's convergence behavior using the criterion $\|a_k - a_{k-1}\|_2^2 < \tau$ for various thresholds τ.

Table 2 shows that the model converges rapidly, with mean stopping iteration $k^* = 6$ at threshold $\tau = 10^{-3}$. This threshold achieves 100% convergence rate with minimal MSE increase (0.04207 vs

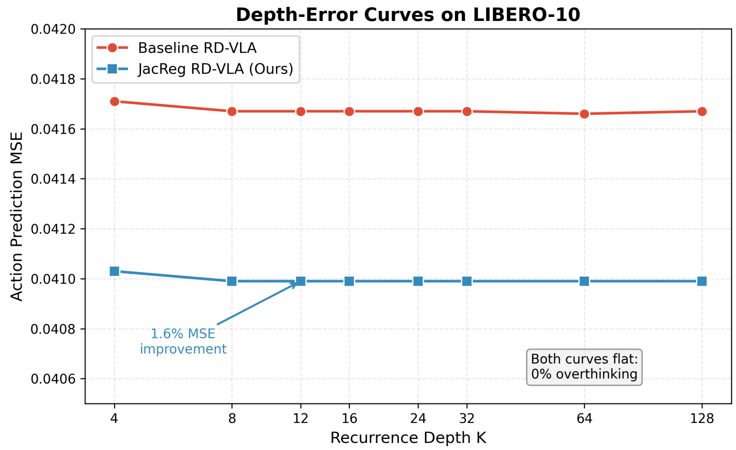

Figure 2: Depth-error curves on LIBERO-10 showing action prediction MSE across recurrence depths $K = 4$ to $K = 128$. Both baseline RD-VLA and JacReg RD-VLA exhibit flat curves (0% overthinking), with JacReg achieving 1.6% lower MSE.

Table 3: Comparison of Jacobian regularization training strategies. Training from scratch (v1, v2) degrades action quality, while fine-tuning from baseline (v3) achieves 1.6% improvement with 15× efficiency gain.

Method	Strategy	MSE@K=12	$\|J\|_F^2/d$	Best Step	GPU-hrs
Baseline	Standard	0.04167	N/A	9,400	$\sim$15
JacReg v1	From scratch, λ=0.1	0.04341 (+4.2%)	0.026	109,400	$\sim$160
JacReg v2	From scratch, λ=0.01	0.04422 (+6.1%)	0.058	73,800	$\sim$88
JacReg v3	**Fine-tune, λ=0.0001**	**0.04099 ($-$1.6%)**	**$\sim$0.16**	**400**	**$\sim$10**

0.04167 at fixed $K = 12$), enabling 50% compute savings. Very tight thresholds ($\tau = 10^{-4}$) cause high variance and reduced convergence, as the model oscillates near the fixed point.

4.4 Training Strategy Ablation

We compare different training strategies for Jacobian regularization to understand the importance of the two-phase approach.

Table 3 reveals that training from scratch with Jacobian regularization degrades action quality: v1 ($\lambda = 0.1$) increases MSE by 4.2%, and v2 ($\lambda = 0.01$) by 6.1%. This occurs because the dual objectives (action loss and Jacobian penalty) create gradient competition that disrupts action representation learning.

In contrast, fine-tuning from a converged baseline with gentle regularization (v3: $\lambda = 10^{-4}$, lr = 10^{-5}) improves MSE by 1.6% while achieving near-contractive dynamics ($\|J\|_F^2/d \approx 0.16$). The two-phase approach is also 15× more efficient, requiring only 400 fine-tuning steps ($\sim$10 GPU-hours) compared to 109,400 steps ($\sim$160 GPU-hours) for v1.

Interestingly, lower Jacobian norm does not guarantee better action quality: v1 achieves the lowest norm (0.026) but the worst MSE. This suggests that overly aggressive regularization constrains the model's representational capacity, while gentle fine-tuning finds a better balance between contractivity and action prediction quality.

5 Conclusion

We investigated Jacobian regularization as a principled approach to address the depth boundary problem in RD-VLA. Surprisingly, on LIBERO-10 with offline teacher-forced evaluation, the depth boundary does not manifest—both baseline and Jacobian-regularized models exhibit 0% overthinking with flat MSE curves across depths $K = 4$ to $K = 128$. This suggests that the phenomenon may be benchmark-specific or evaluation-protocol-dependent.

Despite the absence of depth instability, our investigation yields practical contributions: (1) empirical evidence that the depth boundary is not universal, informing future research on when and why it occurs; (2) a two-phase training strategy that achieves 1.6% MSE improvement with 15× efficiency over from-scratch regularization; and (3) adaptive stopping analysis demonstrating 50% compute savings at inference.

Limitations. Our evaluation is limited to a single benchmark (LIBERO-10) with offline teacher-forced evaluation. The depth boundary may manifest under different conditions, such as closed-loop execution, more challenging tasks, or different model scales.

Future Work. Investigating the conditions under which the depth boundary occurs—including online evaluation, alternative benchmarks, and architectural variations—remains an important direction for understanding the limits of recurrent depth scaling in VLA models.

References

Amin Abyaneh, Charlotte Morissette, Mohamad H. Danesh, Anas El Houssaini, David Meger, Gregory Dudek, and Hsiu-Chin Lin. Contractive diffusion policies: Robust action diffusion via contractive score-based sampling with differential equations, 2026. URL `https://arxiv.org/abs/2601.01003`.

Shaojie Bai, J. Z. Kolter, and V. Koltun. Deep equilibrium models. *ArXiv*, abs/1909.01377, 2019.

Shaojie Bai, Vladlen Koltun, and J. Zico Kolter. Stabilizing equilibrium models by jacobian regularization, 2021. URL `https://arxiv.org/abs/2106.14342`.

Kevin Black, Noah Brown, Danny Driess, Adnan Esmail, Michael Equi, Chelsea Finn, Niccolo Fusai, Lachy Groom, Karol Hausman, Brian Ichter, Szymon Jakubczak, Tim Jones, Liyiming Ke, Sergey Levine, Adrian Li-Bell, Mohith Mothukuri, Suraj Nair, Karl Pertsch, L. Shi, James Tanner, Quan Vuong, Anna Walling, Haohuan Wang, and Ury Zhilinsky. π_0: A vision-language-action flow model for general robot control. *ArXiv*, abs/2410.24164, 2024.

Anthony Brohan, Noah Brown, Justice Carbajal, Yevgen Chebotar, Joseph Dabis, Chelsea Finn, K. Gopalakrishnan, Karol Hausman, Alexander Herzog, Jasmine Hsu, Julian Ibarz, Brian Ichter, A. Irpan, Tomas Jackson, Sally Jesmonth, Nikhil J. Joshi, Ryan C. Julian, Dmitry Kalashnikov, Yuheng Kuang, Isabel Leal, Kuang-Huei Lee, S. Levine, Yao Lu, U. Malla, D. Manjunath, Igor Mordatch, Ofir Nachum, Carolina Parada, Jodilyn Peralta, Emily Perez, Karl Pertsch, Jornell Quiambao, Kanishka Rao, M. Ryoo, Grecia Salazar, Pannag R. Sanketi, Kevin Sayed, Jaspiar Singh, S. Sontakke, Austin Stone, Clayton Tan, Huong Tran, Vincent Vanhoucke, Steve Vega, Q. Vuong, F. Xia, Ted Xiao, Peng Xu, Sichun Xu, Tianhe Yu, and Brianna Zitkovich. Rt-1: Robotics transformer for real-world control at scale. *ArXiv*, abs/2212.06817, 2022.

Anthony Brohan, Noah Brown, Justice Carbajal, Yevgen Chebotar, K. Choromanski, Tianli Ding, Danny Driess, Kumar Avinava Dubey, Chelsea Finn, Peter R. Florence, Chuyuan Fu, Montse Gonzalez Arenas, K. Gopalakrishnan, Kehang Han, Karol Hausman, Alexander Herzog, Jasmine Hsu, Brian Ichter, A. Irpan, Nikhil J. Joshi, Ryan C. Julian, Dmitry Kalashnikov, Yuheng Kuang, Isabel Leal, S. Levine, H. Michalewski, Igor Mordatch, Karl Pertsch, Kanishka Rao, Krista Reymann, M. Ryoo, Grecia Salazar, Pannag R. Sanketi, P. Sermanet, Jaspiar Singh, Anikait Singh, Radu Soricut, Huong Tran, Vincent Vanhoucke, Q. Vuong, Ayzaan Wahid, Stefan Welker, Paul Wohlhart, Ted Xiao, Tianhe Yu, and Brianna Zitkovich. Rt-2: Vision-language-action models transfer web knowledge to robotic control. *ArXiv*, abs/2307.15818, 2023.

Cheng Chi, S. Feng, Yilun Du, Zhenjia Xu, Eric Cousineau, B. Burchfiel, and Shuran Song. Diffusion policy: Visuomotor policy learning via action diffusion. *The International Journal of Robotics Research*, 44:1684 – 1704, 2023.

Mostafa Dehghani, Stephan Gouws, O. Vinyals, Jakob Uszkoreit, and Lukasz Kaiser. Universal transformers. *ArXiv*, abs/1807.03819, 2018.

Tianyu Fu, Yichen You, Zekai Chen, Guohao Dai, Huazhong Yang, and Yu Wang. Think-at-hard: Selective latent iterations to improve reasoning language models. *ArXiv*, abs/2511.08577, 2025.

Maya Janvier, Julien Salomon, and Etienne Meunier. Jacobian regularization stabilizes long-term integration of neural differential equations. 2026.

Moo Jin Kim, Karl Pertsch, Siddharth Karamcheti, Ted Xiao, A. Balakrishna, Suraj Nair, Rafael Rafailov, E. Foster, Grace Lam, Pannag R. Sanketi, Quan Vuong, Thomas Kollar, Benjamin Burchfiel, Russ Tedrake, Dorsa Sadigh, Sergey Levine, Percy Liang, and Chelsea Finn. Openvla: An open-source vision-language-action model. *ArXiv*, abs/2406.09246, 2024.

Bo Liu, Yifeng Zhu, Chongkai Gao, Yihao Feng, Qian Liu, Yuke Zhu, and Peter Stone. Libero: Benchmarking knowledge transfer for lifelong robot learning. *ArXiv*, abs/2306.03310, 2023.

Yalcin Tur, Jalal Naghiyev, Haoquan Fang, Wei-Chuan Tsai, Jiafei Duan, Dieter Fox, and Ranjay Krishna. Recurrent-depth vla: Implicit test-time compute scaling of vision-language-action models via latent iterative reasoning. 2026.

Yang Yue, Yulin Wang, Bingyi Kang, Yizeng Han, Shenzhi Wang, Shiji Song, Jiashi Feng, and Gao Huang. Deer-vla: Dynamic inference of multimodal large language models for efficient robot execution. *ArXiv*, abs/2411.02359, 2024.