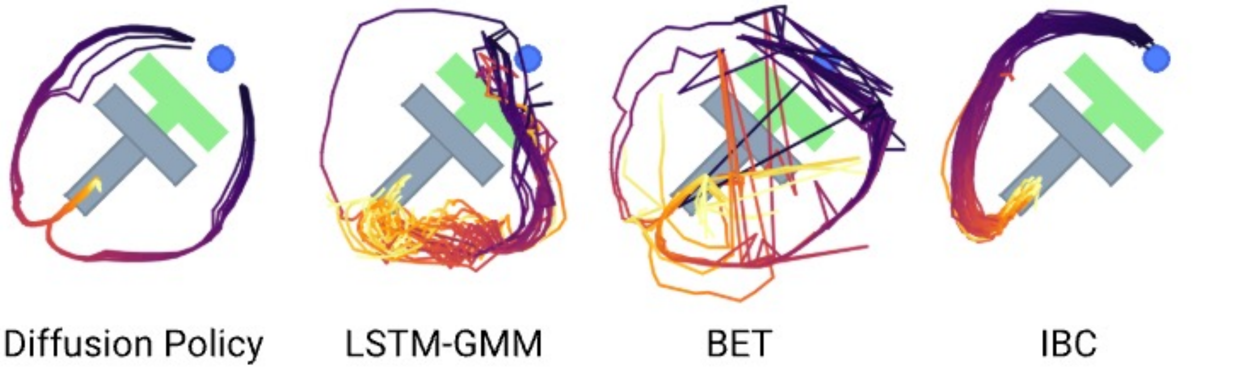
FreqPolicy: Efficient Flow-based Visuomotor Policy via Frequency Consistency

Yifei Su[*], Ning Liu[*†], Dong Chen, Zhen Zhao, Kun Wu, Meng Li, Zhiyuan Xu, Zhengping Che[†‡], Jian Tang[‡]

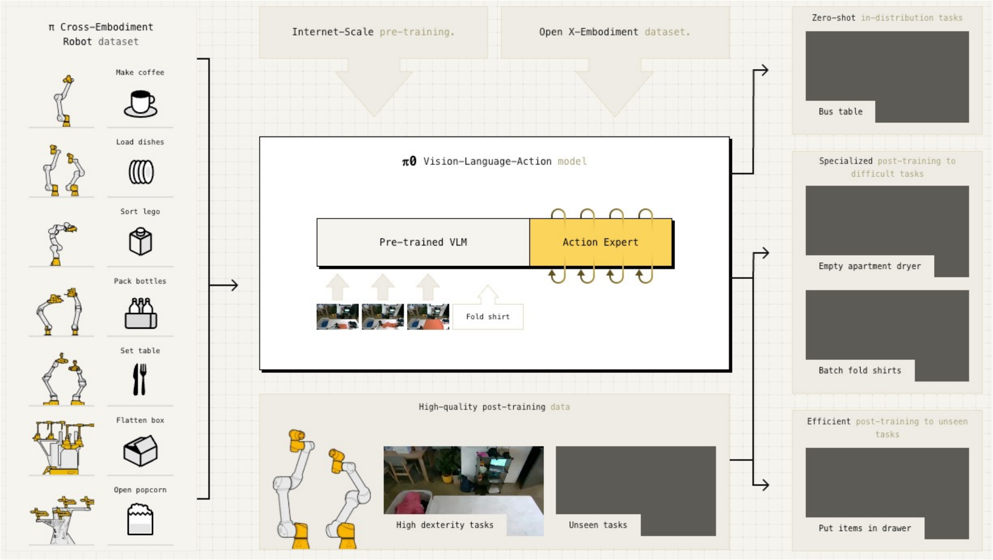
NEURAL INFORMATION PROCESSING SYSTEMS



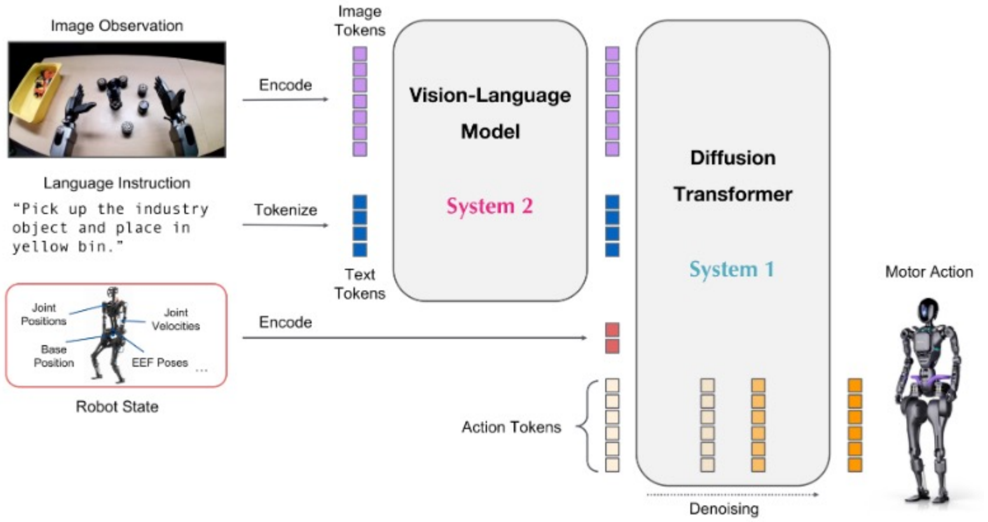
Among various visuomotor policies, generative model-based ones—such as diffusion policies [1] or flow matching policies, have gained excellent performance and widespread popularity.

Chi, Cheng, et al. "Diffusion policy: Visuomotor policy learning via action diffusion." *The International Journal of Robotics Research* 44.10-11 (2025): 1684-1704.

NEURAL INFORMATION PROCESSING SYSTEMS



10 NFE for PI0, PI0.5

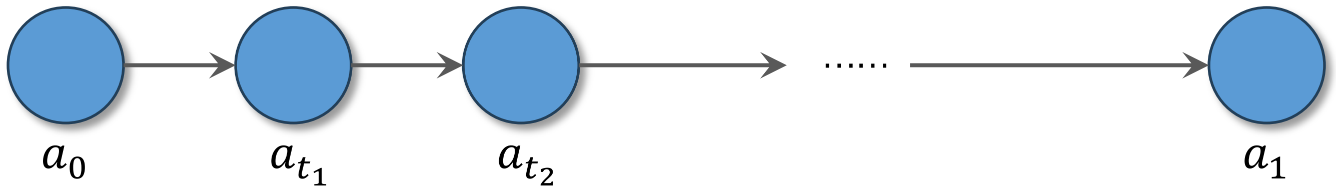


4 NFE for GR00T N1, N1.5

X-HUMANOID Institute of Automation Chinese Academy of Sciences

NEURAL INFORMATION PROCESSING SYSTEMS

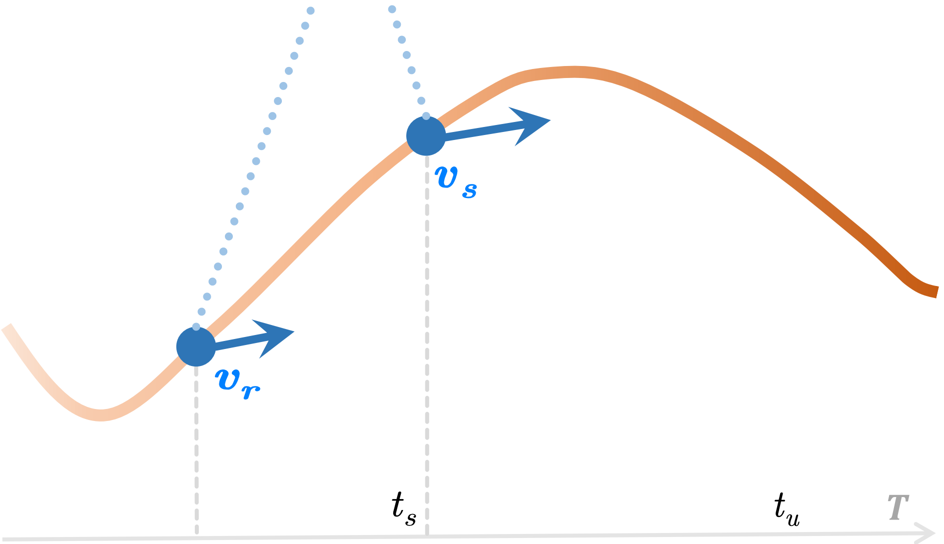
Multi-step inference policies



One-step inference policies

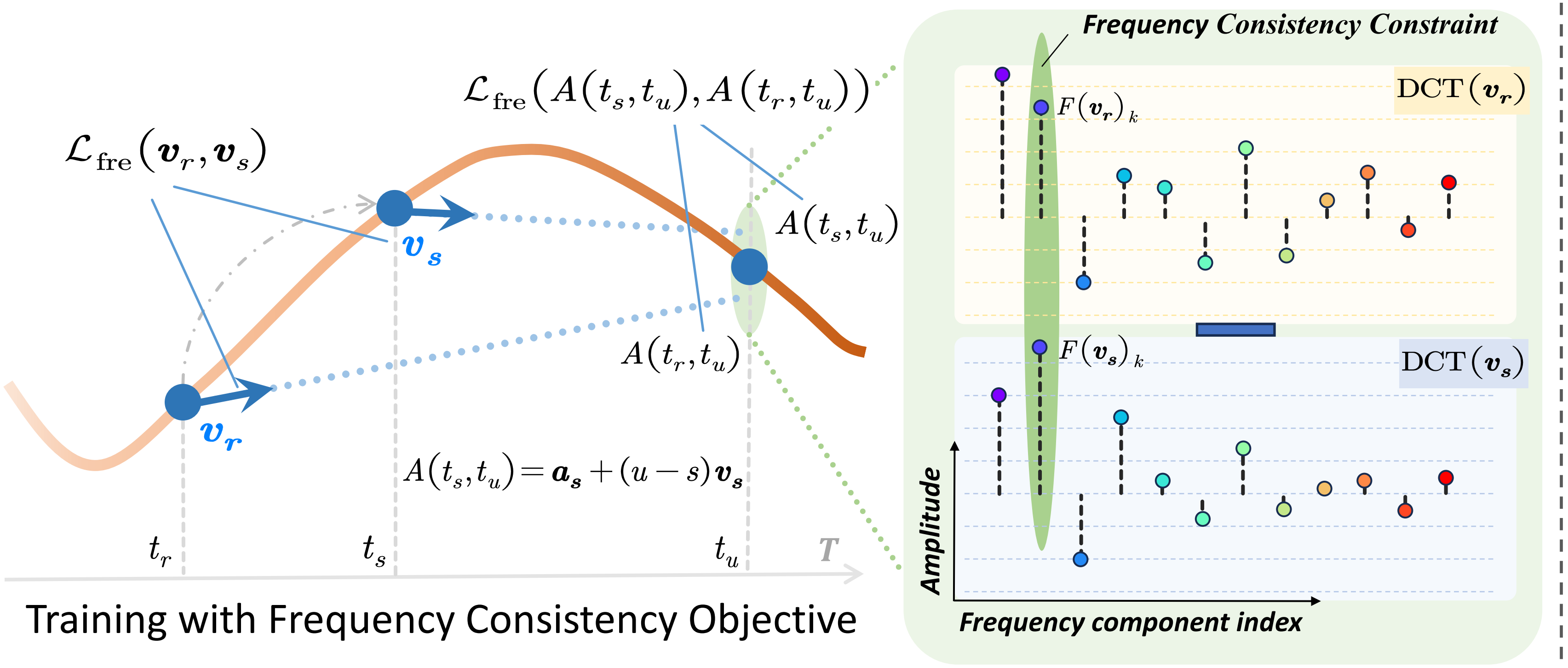
NEURAL INFORMATION PROCESSING SYSTEMS

$$u(t, a_t) = \int_0^1 u(z, a_z)\, dz = a_1 - a_0$$



Training with Basic Flow Matching Objective

NEURAL INFORMATION
PROCESSING SYSTEMS



Training with Frequency Consistency Objective

X-HUMANOID
Institute of Automation Chinese Academy of Sciences

To leverage the temporal of the action chunks, we strengthen the temporal constraint with the frequency regularization using the type-II Discrete Cosine Transform (DCT): :

$$F(v_t)_k = \sum_{n=0}^{H-1} v_t(n) \cdot \cos\left[\frac{\pi}{N}\left(n+\frac{1}{2}\right)k\right], \quad \text{for } k=0,\dots,H-1$$

$$\mathrm{Sim}(v_r, v_s) = \| F(v_r) - F(v_s) \|_2$$

The frequency components in each action chunks vary over the task execution. We propose an adaptive weighting scheme that focus on frequency components with larger discrepancies.

$$w_k = \frac{\exp \| F(v_r)_k - F(v_s)_k \|_2)}{\sum_{j=0}^{H-1} \exp \| F(v_r)_j - F(v_s)_j \|_2)}$$

$$\mathrm{Sim}(v_r, v_s) = \sum_{k=0}^{H-1} w_k \cdot \| F(v_r)_k - F(v_s)_k \|_2$$

NEURAL INFORMATION PROCESSING SYSTEMS

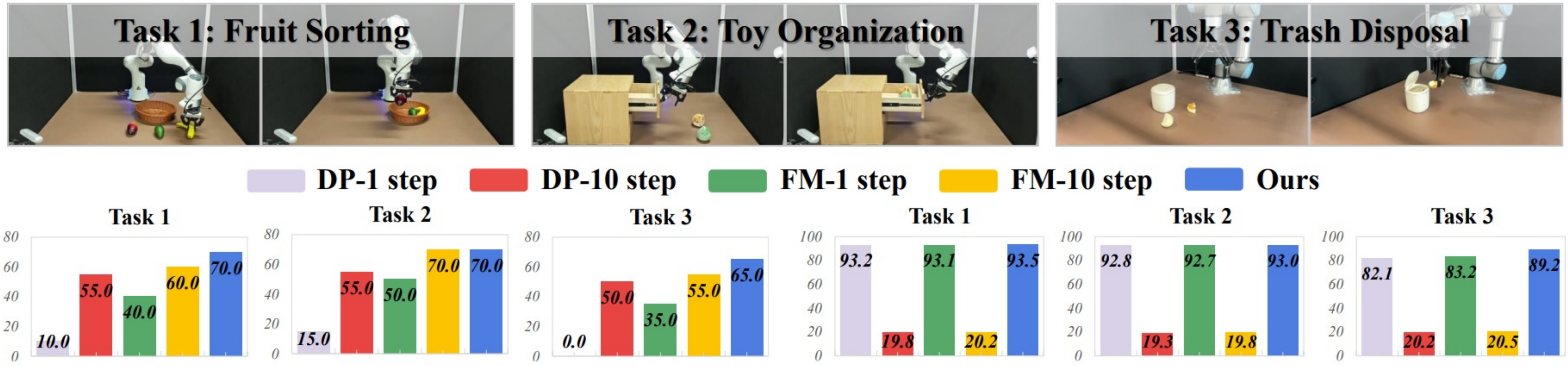
Method	NFE	Lift	Can	Square	Transport	Toolhang
DDPM [11]	15	1.00	$0.98 \pm .01$	$0.91 \pm .01$	$0.80 \pm .04$	$0.52 \pm .05$
DDiM [11]	15	1.00	$0.99 \pm .01$	$0.92 \pm .03$	$0.79 \pm .04$	$0.55 \pm .05$
RectifiedFlow* [40]	15	1.00	$0.96 \pm .02$	$0.90 \pm .02$	$0.84 \pm .04$	$0.90 \pm .02$
ActionFlow [19]	10	1.00	$0.99 \pm .01$	$0.87 \pm .10$	-	$0.81 \pm .09$
AdaFlow [24]	-	1.00	1.00	0.98	0.92	0.88
ConsistencyPolicy* [53]	3	1.00	$0.95 \pm .02$	$0.96 \pm .01$	$0.88 \pm .02$	$0.77 \pm .03$
DDiM [11]	1	0.04	$0.00 \pm .00$	$0.00 \pm .00$	$0.00 \pm .00$	$0.00 \pm .00$
ConsistencyPolicy* [53]	1	1.00	$0.98 \pm .01$	$0.92 \pm .02$	$0.78 \pm .03$	$0.70 \pm .03$
Consistent-FM* [68]	1	1.00	$0.94 \pm .02$	$0.90 \pm .01$	$0.84 \pm .02$	$0.80 \pm .02$
IMLE Policy [57]	1	1.00	0.98	0.82	0.90	0.81
Ours	1	1.00	$\mathbf{0.98} \pm \mathbf{.02}$	$\mathbf{0.92} \pm \mathbf{.02}$	$\mathbf{0.90} \pm \mathbf{.02}$	$\mathbf{0.85} \pm \mathbf{.03}$

FreqPolicy outperforms previous one-step methods such as CP and IMLE Policy. Furthermore, it even surpasses several classical multi-step policies.

Method	NFE	Spatial (%)	Object (%)	Goal (%)	Long (%)	Average (%)	Speed (Hz)
OpenVLA-DP [32]	50	92.0	75.0	93.4	11.8	68.1	0.32
OpenVLA-FlowMatching*	1	95.0	97.6	96.0	85.2	93.5	5.92
OpenVLA-FlowMatching*	10	96.0	97.2	**97.8**	83.6	93.7	1.26
OpenVLA-FreqPolicy (**Ours**)	1	**97.0**	**98.6**	96.0	**87.6**	**94.8**	**6.05**

Our FreqPolicy hits 94.8% average success, outperforms most baselines, excels vs. OpenVLA-DP (87.6% vs.11.8% in Long), and is faster than OpenVLA-Flow Matching (1 vs.10 NFE).

NEURAL INFORMATION
PROCESSING SYSTEMS



(a) Success rate (%) on three real-world tasks.

(b) Inference speed (Hz) on three real-world tasks.

X-HUMANOID

Institute of Automation Chinese Academy of Sciences

Thanks !