

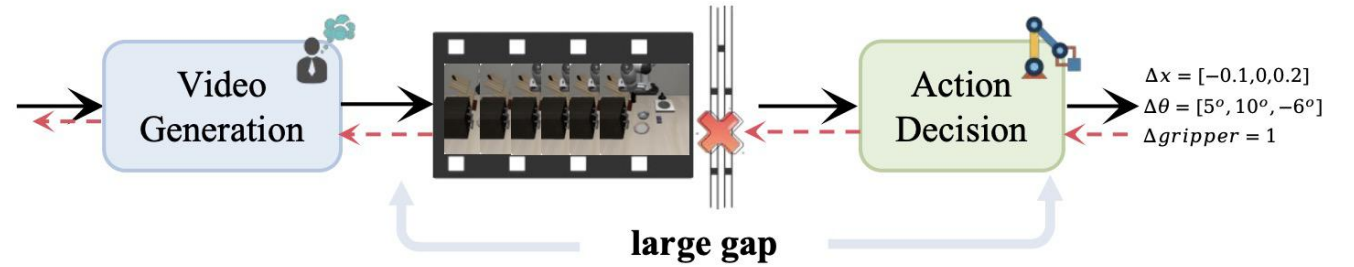
# DynBridge: Bridging Imagination and Control through Interaction Dynamics for Robot Manipulation



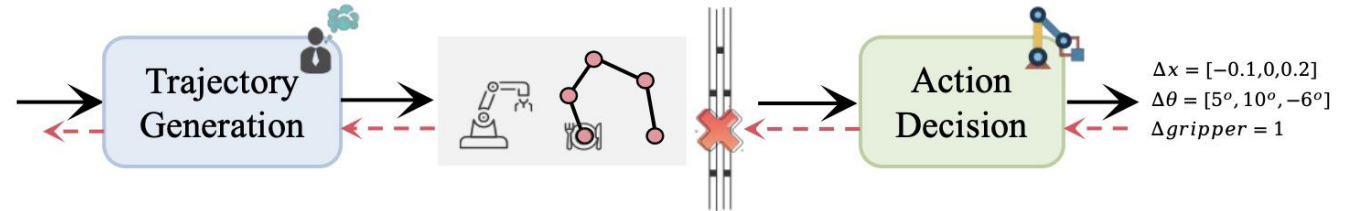
Alex Wang\*, Zhiwei Dong\*, Qicheng Bai, Chenshi Zhang, Yujie Yi,  
Guang Dai, Yong Liu, Mengmeng Wang

## Motivation:

Recent generative models allow robots to generate future visual outcomes for action guidance, yet most still address Video imagination and control independently, resulting in *visually coherent rollouts but physically inconsistent behaviors*.

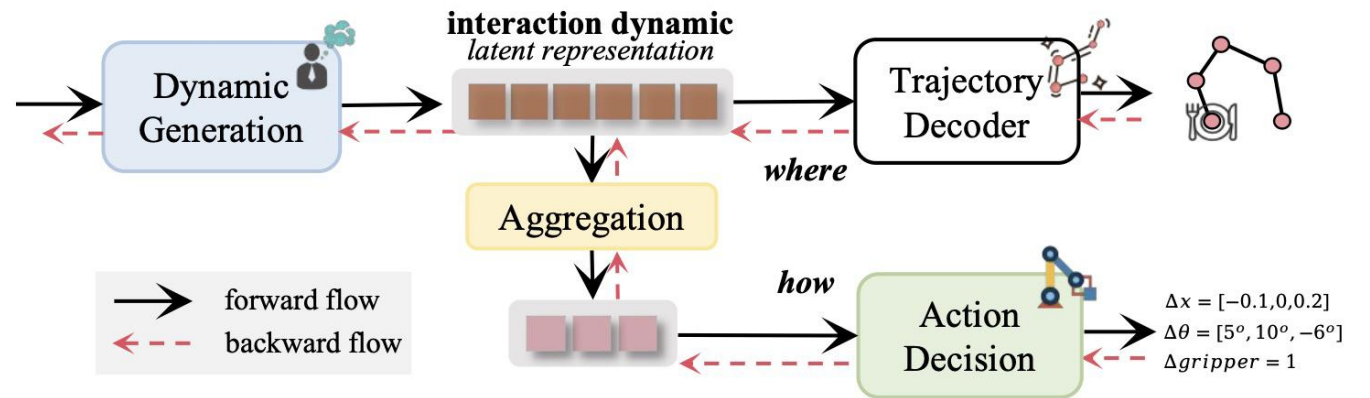


*unidirectionally* mitigates the gap via *decision-friendly generation*



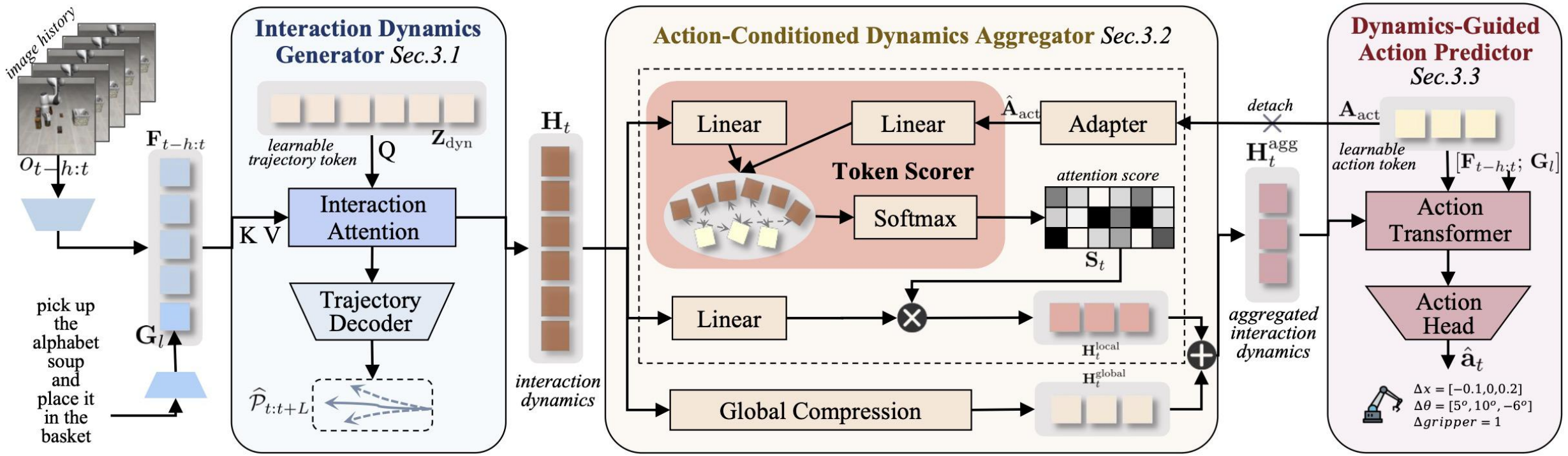
(a) Previous methods

*bidirectionally* bridges the gap via *interaction dynamics* - encoding *where* and *how*



(b) Our method

# Overview framework of DynBridge



DynBridge consists of three core modules:

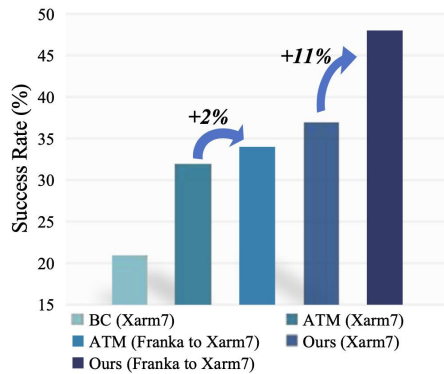
- **Interaction Dynamics Generator** – employs learnable dynamic tokens that attend across modalities to imagine future interaction dynamics.
- **Action-Conditioned Dynamics Aggregator** – aggregates dynamics for subsequent action decision, leveraging generated interaction dynamics.
- **Dynamic-Guided Action Predictor Decision** – leverages the aggregated inter- action dynamics and historical context to predict context-aware and executable actions.

# Performance comparison

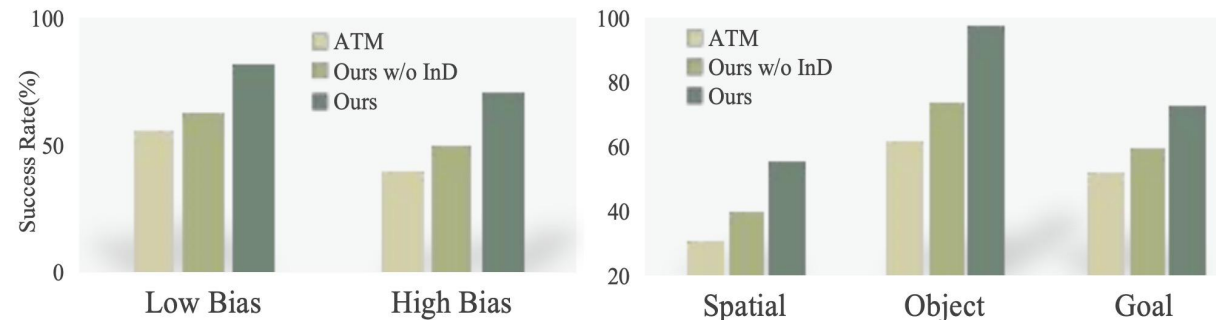
Average success rate on the LIBERO benchmark.

Methods	Ext. Data	Libero-Spatial	Libero-Object	Libero-Goal	Libero-Long	Libero-90
BC-Full-Trainset	✗	$0.71 \pm 0.03$	$0.71 \pm 0.08$	$0.76 \pm 0.01$	$0.24 \pm 0.02$	$0.38 \pm 0.01$
BC	✗	$0.39 \pm 0.08$	$0.51 \pm 0.13$	$0.42 \pm 0.04$	$0.16 \pm 0.03$	$0.29 \pm 0.01$
R3M-finetune [24]	✗	$0.49 \pm 0.04$	$0.52 \pm 0.02$	$0.05 \pm 0.01$	$0.09 \pm 0.03$	$0.09 \pm 0.00$
VPT [2]	✓	$0.38 \pm 0.04$	$0.19 \pm 0.01$	$0.01 \pm 0.01$	$0.03 \pm 0.02$	$0.04 \pm 0.01$
UniPi [11]	✓	$0.69 \pm 0.04$	$0.59 \pm 0.03$	$0.11 \pm 0.02$	$0.05 \pm 0.02$	$0.07 \pm 0.01$
ATM [36]	✓	$0.68 \pm 0.02$	$0.68 \pm 0.06$	$0.77 \pm 0.01$	$0.39 \pm 0.15$	$0.48 \pm 0.02$
GraphMimic [4]	✓	$0.88 \pm 0.02$	$0.89 \pm 0.02$	$0.87 \pm 0.03$	$0.56 \pm 0.03$	$0.67 \pm 0.01$
<b>Ours</b>	✗	<b><math>0.92 \pm 0.01</math></b>	<b><math>1.00 \pm 0.00</math></b>	<b><math>0.92 \pm 0.00</math></b>	<b><math>0.71 \pm 0.01</math></b>	<b><math>0.75 \pm 0.01</math></b>

Interaction dynamics are transferable across embodiments, capable of capturing reusable causal patterns even from failed demonstrations.



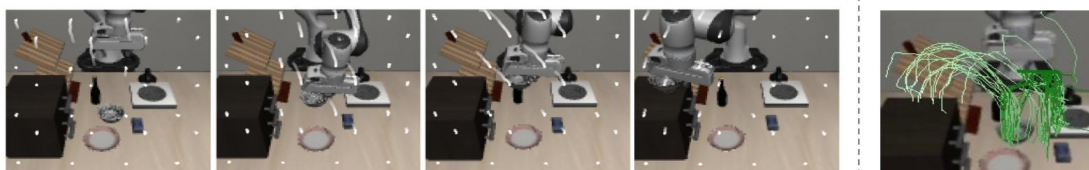
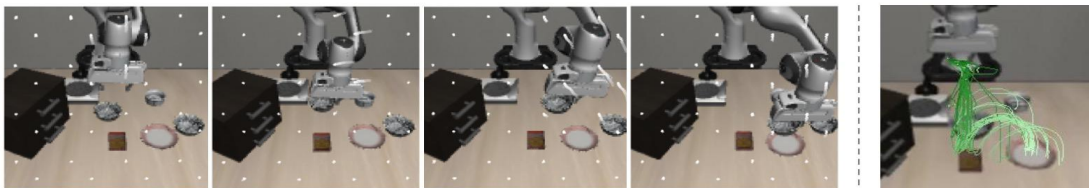
DynBridge achieves stable generalization by modeling the causal relations between robot actions and objects.



# Performance comparison

DynBridge effectively bridges the imagination – control gap.

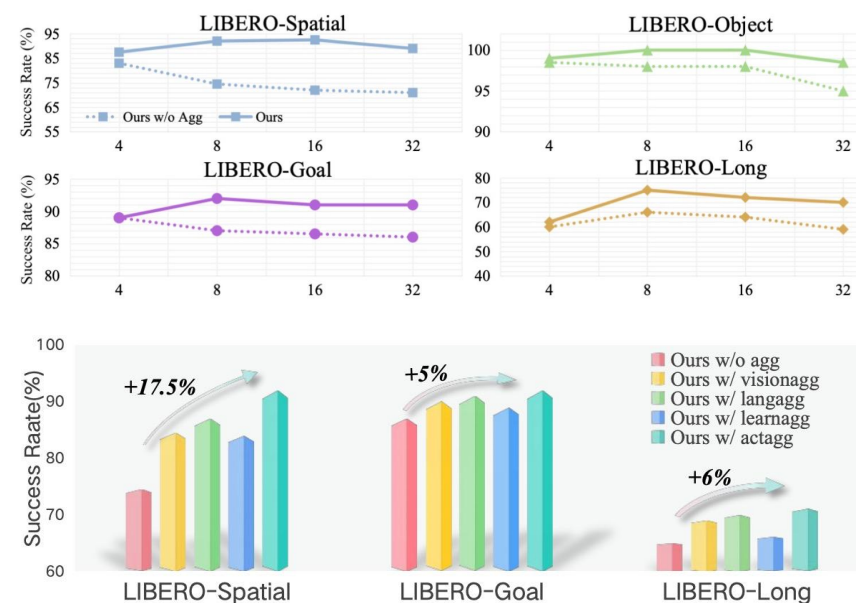
*ATM*



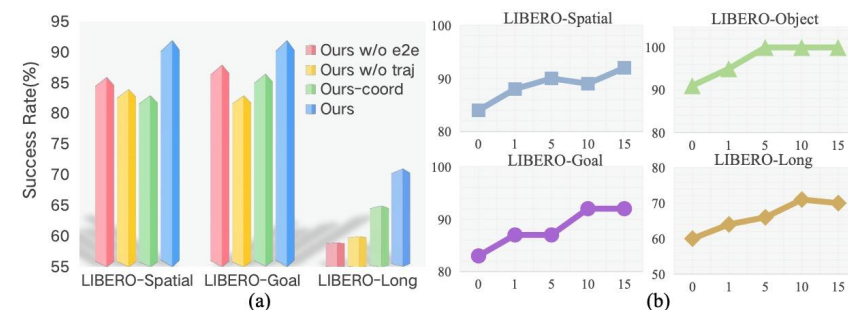
*Ours*



## Effect of the Dynamics Aggregator



## Effect of the Interaction Dynamics Generator



# Conclusions

- We introduce interaction dynamics, a latent representation that encodes where and how interactions unfold, capturing both spatial structure and physical dynamics beyond visual correlations.
- We propose DynBridge, an end-to-end framework that integrates trajectory generation and action prediction through shared interaction dynamics, effectively bridging the gap between imagination and control.
- We empirically demonstrate that DynBridge, without any additional robot-data pretraining, achieves consistent improvements and strong generalization across both simulated and real-world benchmarks.

Feel free to contact with [alexw\\_rob@163.com](mailto:alexw_rob@163.com).