Masked Generative Policy For Robotic Control

Lipeng Zhuang∗, Shiyu Fan∗, Florent P. Audonnet, Yingdong Ru, Edmond S.L Ho, Gerardo Aragon Camarasa, Paul Henderson

∗*Equal contribution*

ICLR2026

Motivation

Why current imitation-learning policies still struggle at execution time ?

Diffusion-based policies

Diffusion Policy, DP3

- Multi-step denoising at inference, leading to an **inference-time** bottleneck

Autoregressive / GPT-style policies

QueST

- Token-by-token decoding
- Latency increases with **sequence length**

Shared limitations

- Weak **online adaptation** to unexpected changes during execution
- No explicit **long-horizon memory**

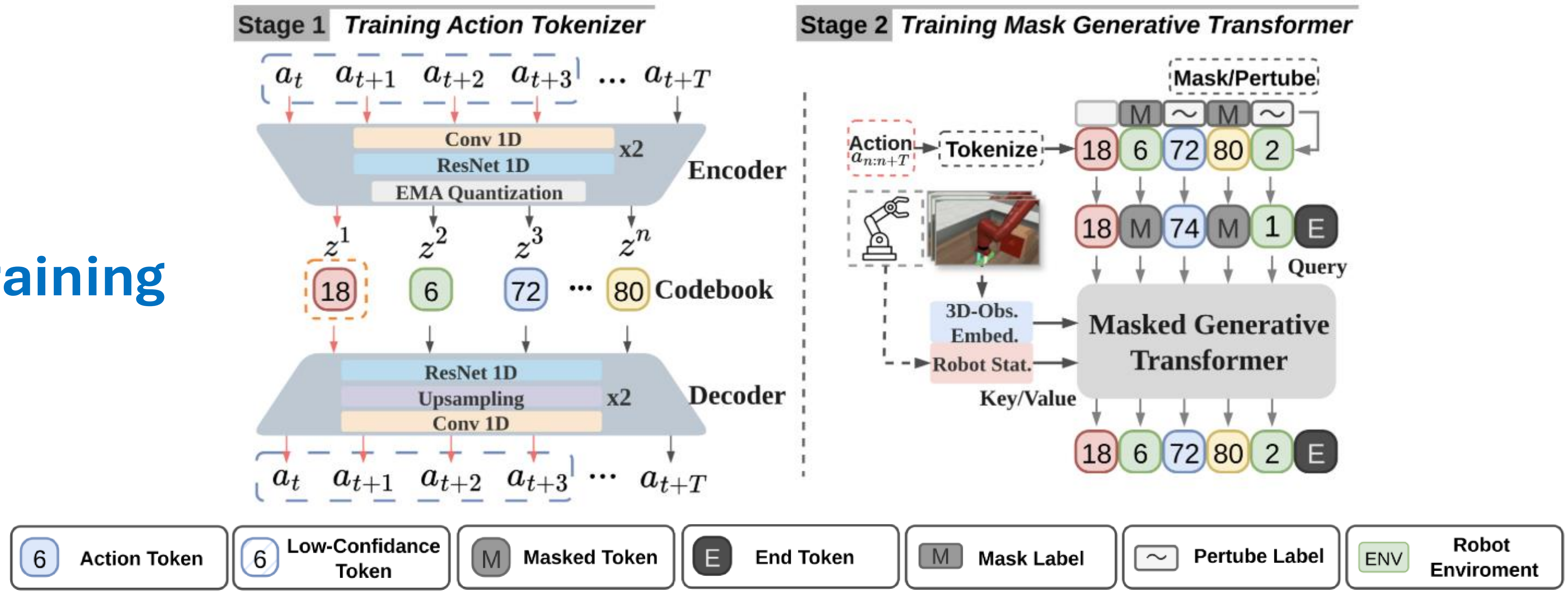
CVAS group

University of Glasgow

Method Structure

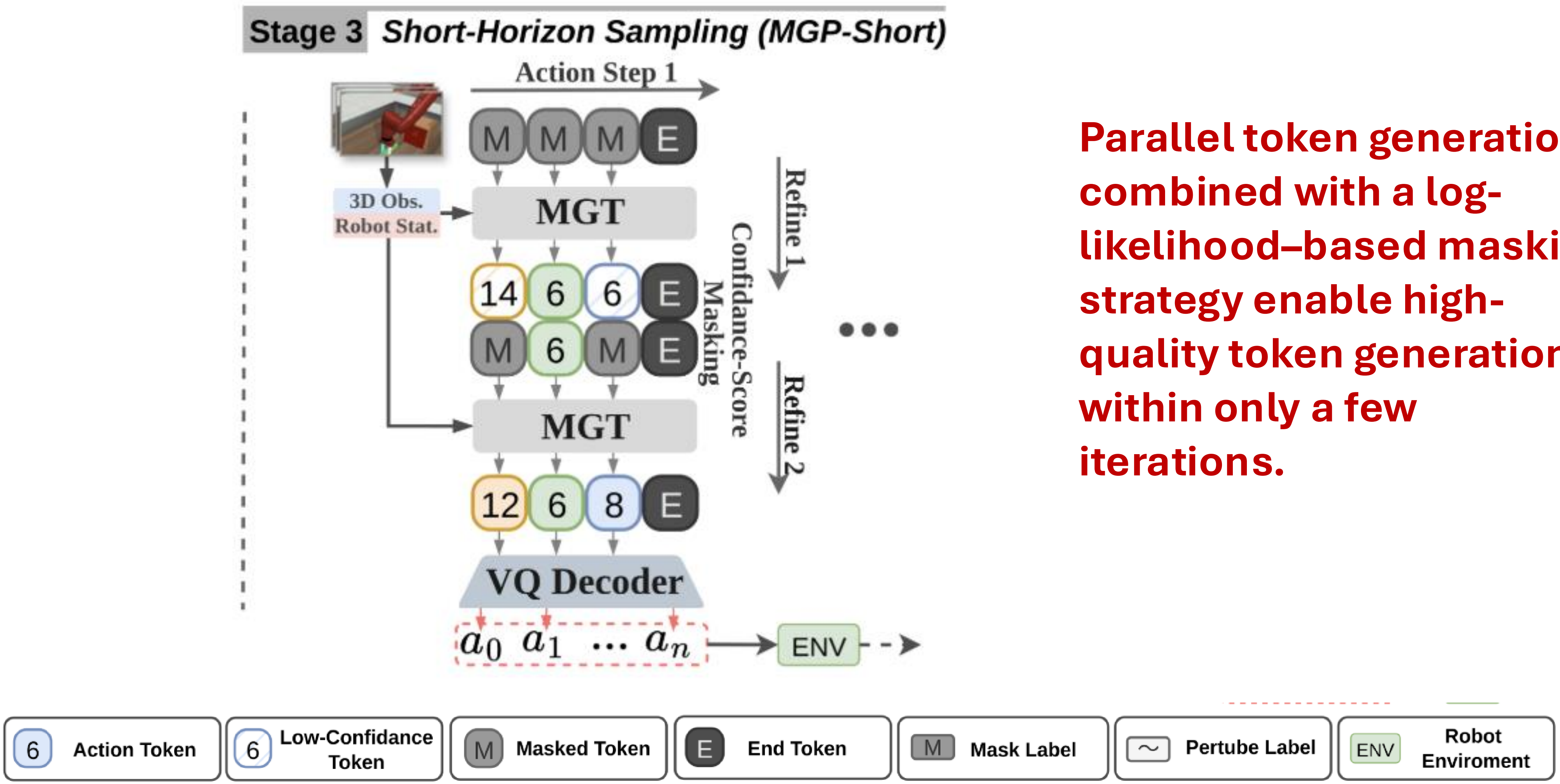
MGP is a stochastic policy that at time t samples a sequence of future actions given a conditioning including past actions, robot states, and visual observations.

Training



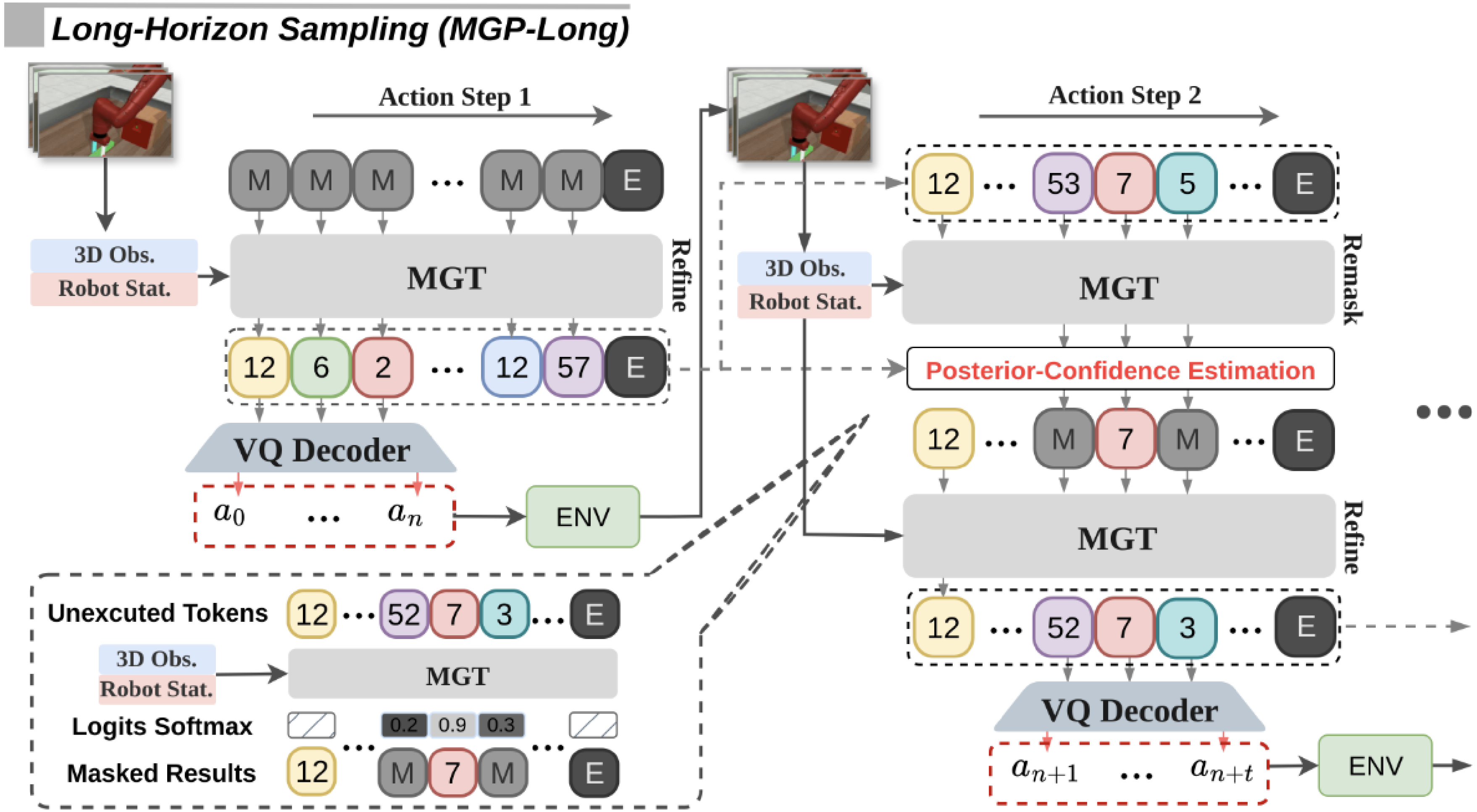
Left: Training Stage 1 - Action Tokenizer and Middle: Training Stage 2 - Masked Generative Transformer

Short Horizon Mask-And-Refine Sampling (MGP-Short)



Parallel token generation, combined with a log-likelihood–based masking strategy enable high-quality token generation within only a few iterations.

Long Horizon Mask-And-Refine Sampling (MGP-Long)



Long-horizon sampling (MGP-Long) through Adaptive Token Refinement (ATR).

Adaptive Token Refinement (ATR) strategy

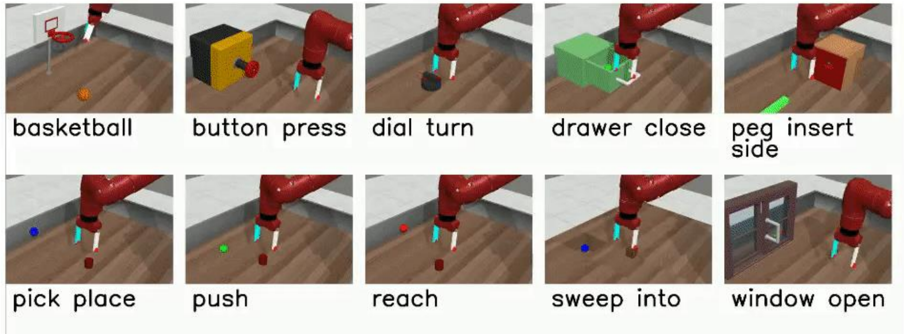
Posterior-Confidence Estimation

Variable Execution Step

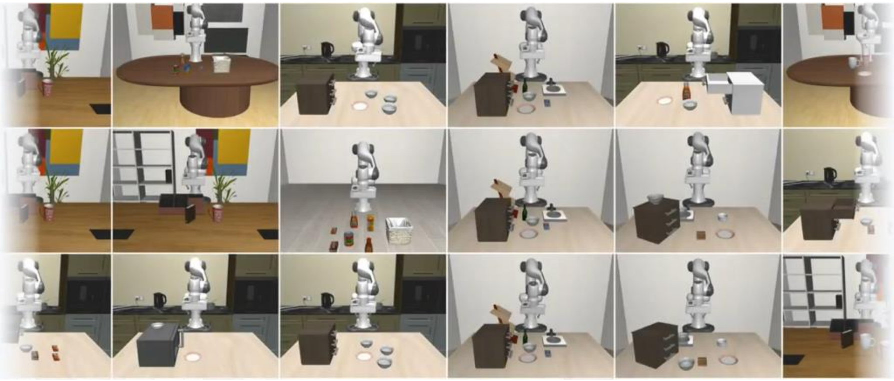
Simulation Experiments

- ❖ **Ten Baselines:** 5 diffusion based; 5 autoregressive based
- ❖ **Three Standard Benchmarks (*150 tasks in total*)**

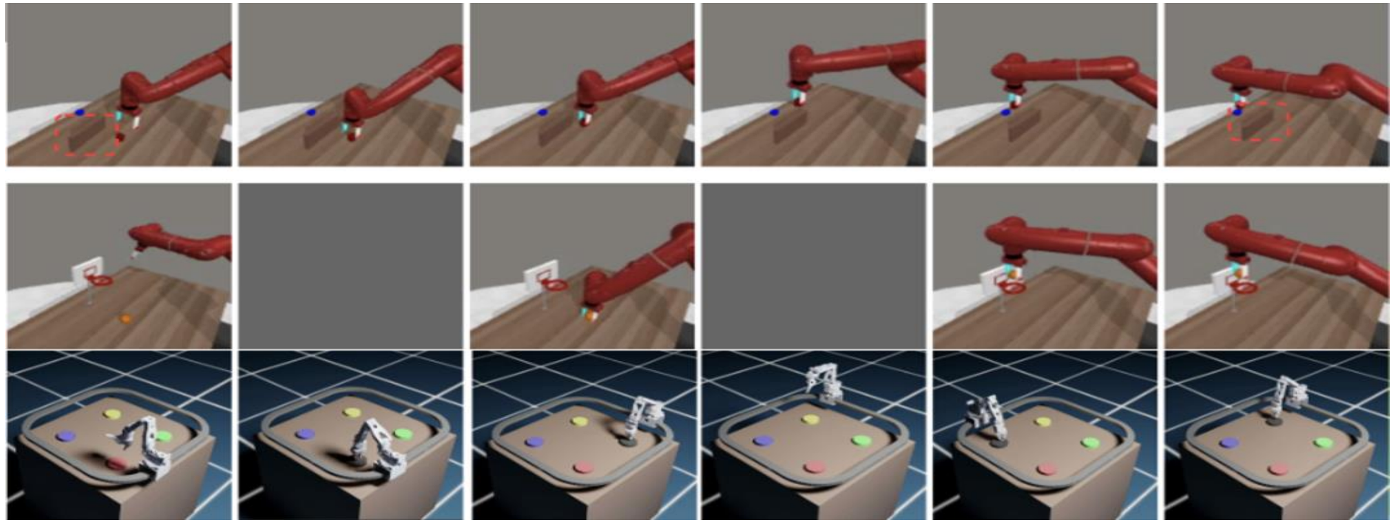
Meta-World



LIBERO-90&
LIBERO-Long



- ❖ **Three Challenging Evaluation Environments**



Dynamic environments with moving objects/obstacles

Observation missing

Two long-duration *non-Markovian* task

Summary of Standard Benchmarks Evaluation

*Evaluated on **Meta-World, LIBERO-90**, and **LIBERO-Long** benchmarks (**150+** manipulation tasks).*

Single-task

Meta-World

Avg. success

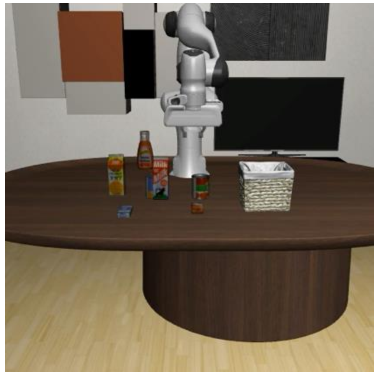
0.637 (4% increase)

49× faster than DP3

SOTA on short-horizon control

Meta-World

Libero-90



Libero-Long

Multi-task

LIBERO-90

Avg. success

0.889

3.4× faster than QueST

SOTA on multitask performance

Long-horizon

LIBERO-Long

Avg. success

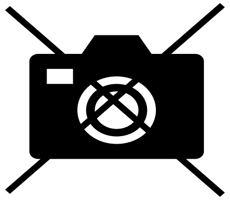
0.820 (14% increase)

10× faster than QueST

SOTA on long-horizon tasks

Robustness Across Challenging Environments

Missing Observations



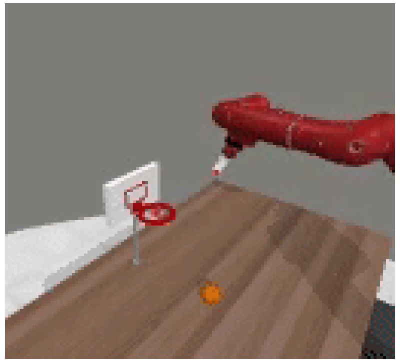
At each control cycle during inference, the current observation is withheld with some probability (even 70%)

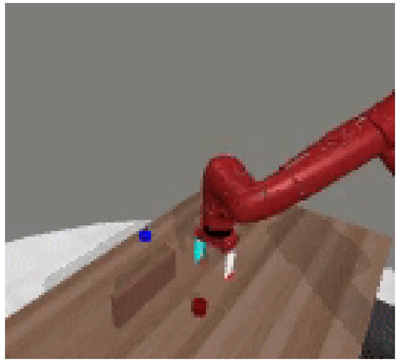
Avg. success

0.525

(**+21%** vs full-horizon,
+32% vs short-horizon)

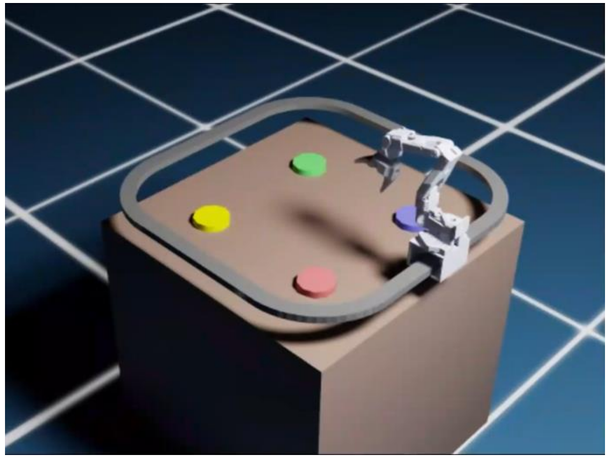
Dynamic Environments



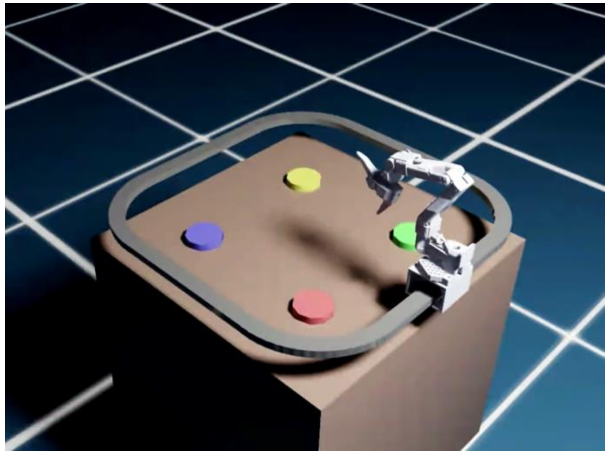


- *Best average success in five dynamic Meta-World*

Non-Markovian Tasks



Button Press Change Color

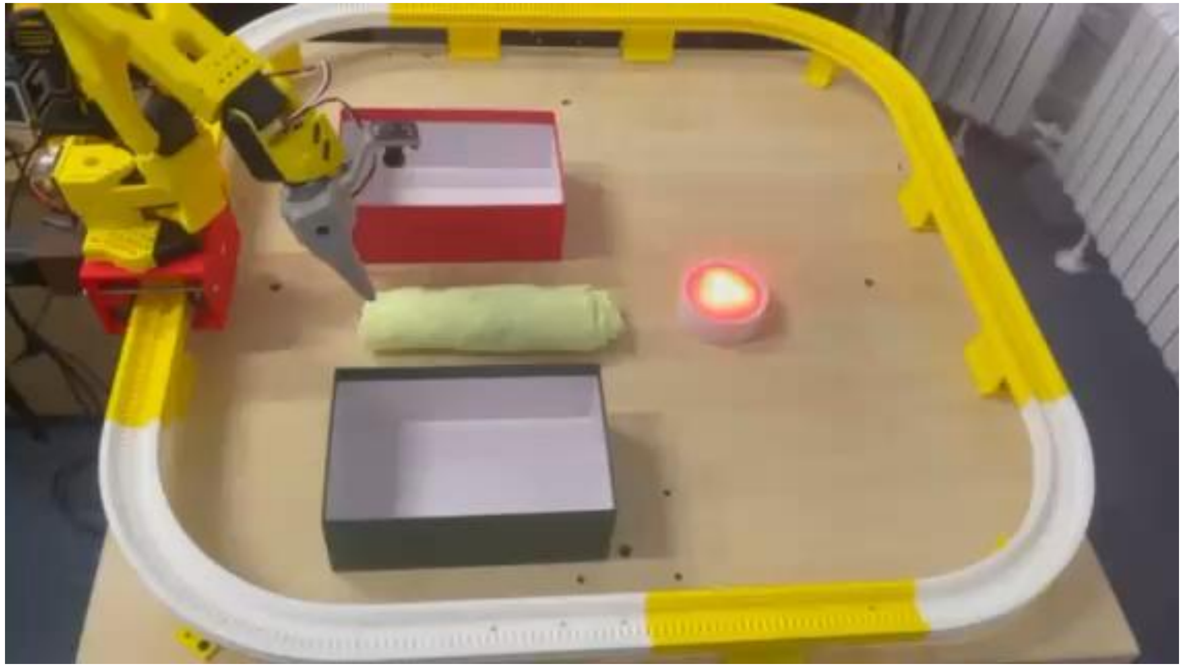


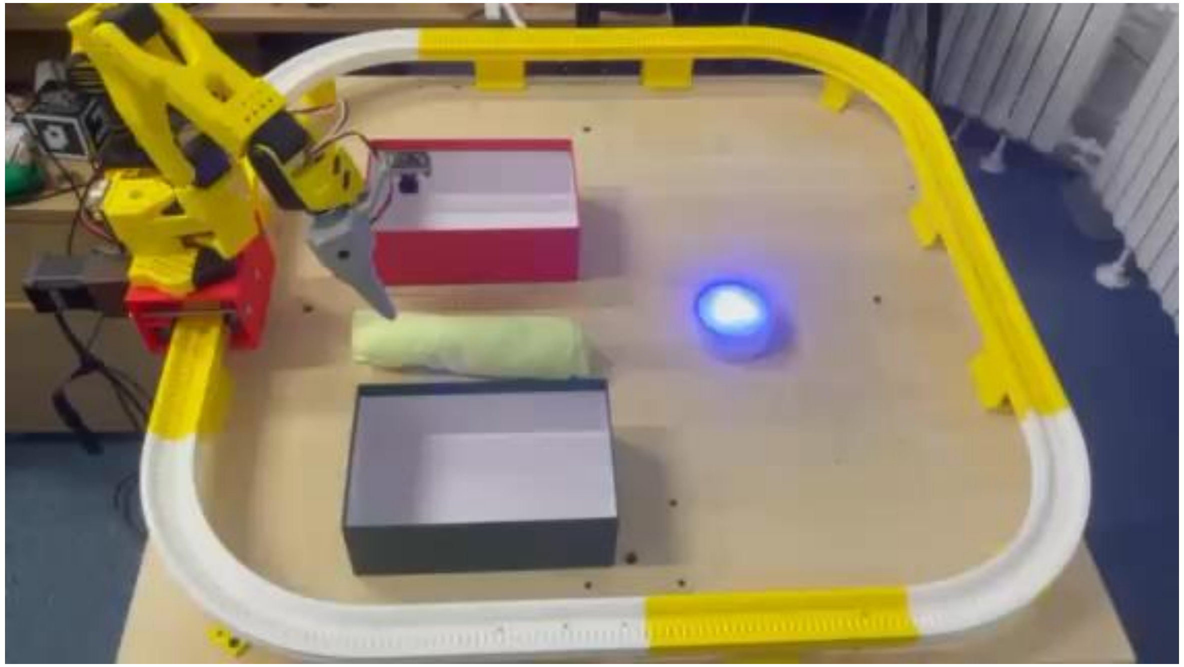
Button Press On/Off

MGP-Long is the only method that succeeds on both non-Markovian button tasks, achieving a 100% success rate.

Real World Experiments

- We collect 60 human demonstrations. approximately 100 timesteps/demonstration
- Each demonstration contains synchronized RGB images, depth maps, and colored point clouds, along with joint state and end-effector signals
- MGP-Long achieves a success rate of 96%, outperforming DP3-Full Seq (which achieves 84%).





Thanks!

For more details, please refer to our paper: *Masked Generative Policy for Robotic Control* (*https://github.com/lipeng-zhuang521/MGT*).

Do you have any questions?

lipeng.zhuang@glasgow.ac.uk s.fan.1@research.gla.ac.uk