

On-Device Diffusion Transformer Policy for Efficient Robot Manipulation

Supplementary Material

Figure A1. Real-world experiments for DP-T (first column) and MoDE (other columns). Task descriptions are shown below each image. **This figure contains an animated video. For optimal viewing, please zoom in and use a professional PDF reader.**

G. Supplementary Material

The supplement consists of the following sections:

- Section H presents the extensive experimental results on Robomimic dataset based on the DiffusionPolicy Transformer (DP-T) model.
- Section I describes the real-world experiments based on DP-T and MoDE models, including the experimental setup and results.

We provide a webpage to visualize the results of the pruned models and original models, which can be found at <https://weleen.github.io/LightDP/>.

H. Extensive Experiments based on DP-T

Models	Lift-ph	Can-ph	Square-ph	Transport-ph	Push-T	ToolHang-ph
DP-T	1.000	1.000	1.000	0.955	0.772	0.713
DP-T-D6/6-8	1.000	1.000	1.000	0.950	0.752	0.707

Models	Lift-mh	Can-mh	Square-mh	Transport-mh	Kitchen	Block Push
DP-T	1.000	1.000	0.940	0.727	0.574	1.000
DP-T-D6/6-8	1.000	1.000	0.955	0.773	0.571	1.000

Table A1. The extensive evaluation on DP-T tasks (Push-T and Robomimic), showing the success rates of the original model (DP-T) and the pruned model (DP-T-D6/6-8). The pruned model maintains performance across most tasks, with only minor drops in success rates.

In Table A1, we have provided success rate on all tasks (*i.e.*, Push-T and Robomimic) in the Diffusion Policy [6] work, which indicate that the pruned model DP-T-D6/6-8 preserves the baseline’s performance on most tasks, and the performance only drops by less than 0.02 on the tasks.

I. Real-world Experiments

Based on two models DP-T and MoDE, we deploy our *LightDP* on two robotic arms (an Inovo robot for DP-T

Models	Task 1	Models	Task 2	Task 3	Task 4
DP-T	0.80	MoDE	0.80	0.55	0.30
DP-T-D6/6-8	0.75	MoDE-10/10-12	0.75	0.50	0.30

Table A2. Real-world evaluation results based on DP-T (on a Inovo Robot) and MoDE (on a Lebai Robot). The success rates are shown for each task, with the pruned model (DP-T-D6/6-8 and MoDE-10/10-12) maintaining performance across most tasks, with only minor drops in success rates.

and a Lebai robot for MoDE), where each task is executed by 20 times. As shown in Figure A1 and Table A2, the pruned model achieves a comparable success rate on these real-world tasks. Considering that most household users are often redundant to purchase advanced device, we selected the most accessible and portable device (*i.e.*, iPhone) as the computing platform for our robotic development setup. Moreover, we also evaluate our approach based on a Jetson Orin NX (16 GB, Jetpack 5.1.1), the latency is 244.68ms (*resp.*, 37.69ms) based on DP-T (*resp.*, DP-T-D6/6-8).