UniPi: Learning Universal Policies via Text-Guided Video Generation

Yilun Du*, Sherry Yang*, Bo Dai, Hanjun Dai, Ofir Nachum, Josh Tenenbaum, Dale Schuurmans, Pieter Abbeel

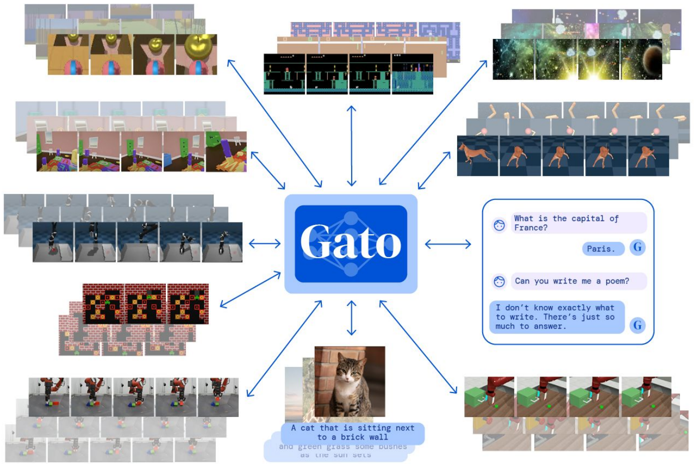
MIT Google Brain University of Alberta Berkeley University of California

Goal: Generalist Agent

Why: Generalization across environments / tasks

Goal: Generalist Agent

Why: Generalization across environments / tasks
Examples: Gato, Multi-Game DT, Scaled QL, RT-1



Gato

Multi-Game DT

Goal: Generalist Agent

Challenge

- Environment diversity: Different state action spaces

Goal: Generalist Agent

Challenge

- Environment diversity: Different state action spaces
- Reward diversity: Different reward functions

Goal: Generalist Agent

Challenge

- Environment diversity: Different state action spaces
- Reward diversity: Different reward functions

Previous Solution

- Tokenization. Might loose knowledge from pretrained models

Goal: Generalist Agent

Challenge

- Environment diversity: Different state action spaces
- Reward diversity: Different reward functions

Previous Solution

- Tokenization. Might loose knowledge from pretrained models
- Text-as-task. Unified notion of task / reward for all envs

Goal: Generalist Agent

Challenge

- Environment diversity: Different state action spaces
- Reward diversity: Different reward functions

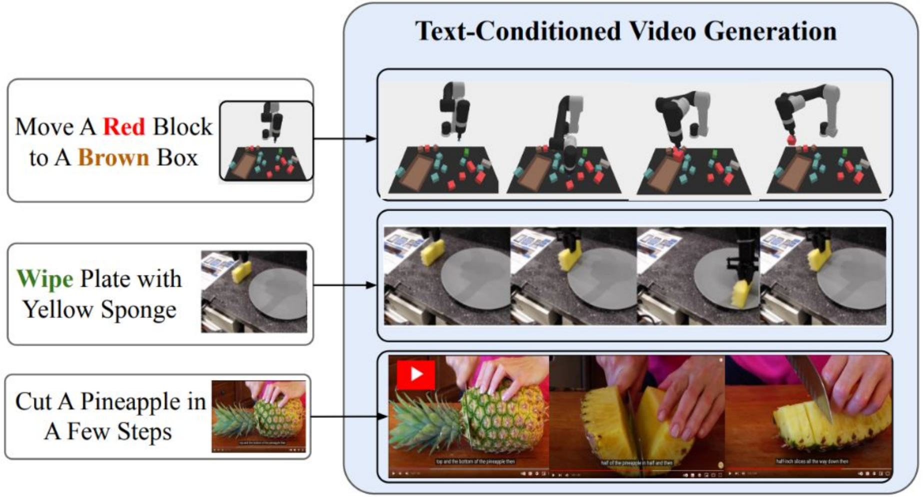
Previous Solution

- Tokenization. Might loose knowledge from pretrained models
- Text-as-task. Unified notion of task / reward for all envs

Our Solution

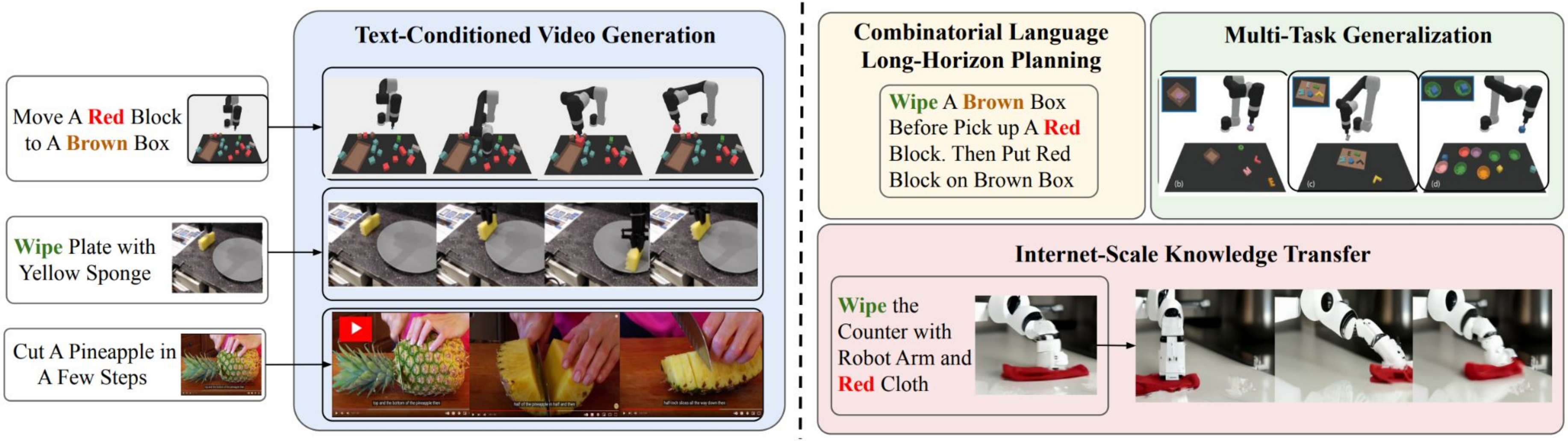
- Video-as-policy: Unified state-action spaces for all envs
- Text-as-task

UniPi: Universal Policy via Text-Conditioned Video Generation



Pretrain general purpose policies on wide sources of data (simulated, real robots and YouTube).

UniPi: Universal Policy via Text-Conditioned Video Generation

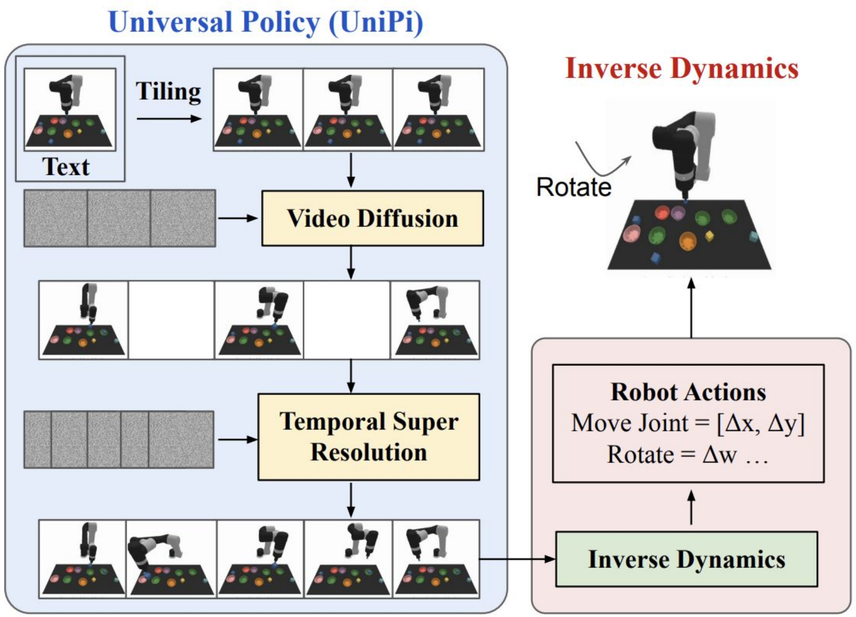


Pretrain general purpose policies on wide sources of data (simulated, real robots and YouTube). Generalize to multi-task settings requiring combinatorical language generalization, long-horizon planning, or internet-scale knowledge.

UniPi Implementation

Conditional Video Synthesis

- Conditioned on the first frame



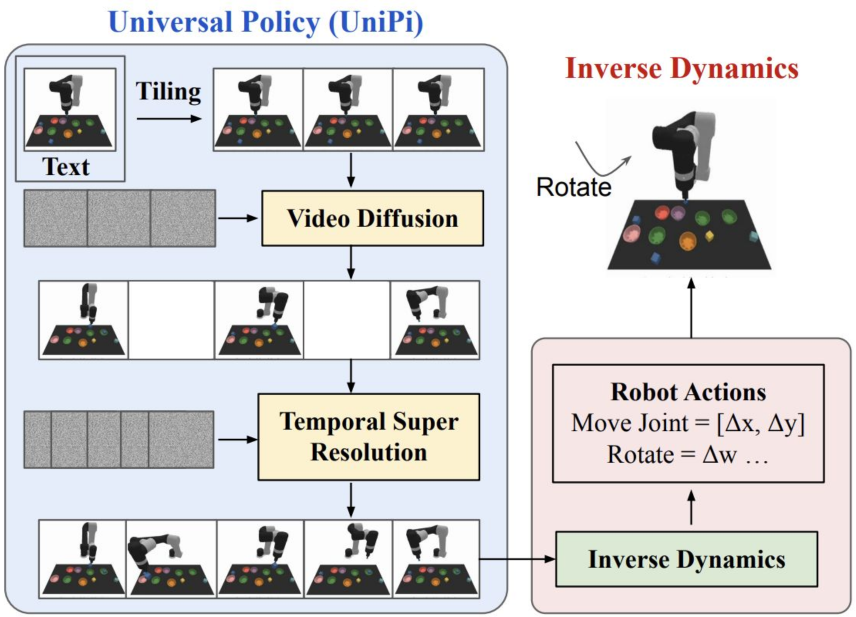
UniPi Implementation

Conditional Video Synthesis

- Conditioned on the first frame

Trajectory Consistency through Tiling

- Frames are replicated across time



UniPi Implementation

Conditional Video Synthesis

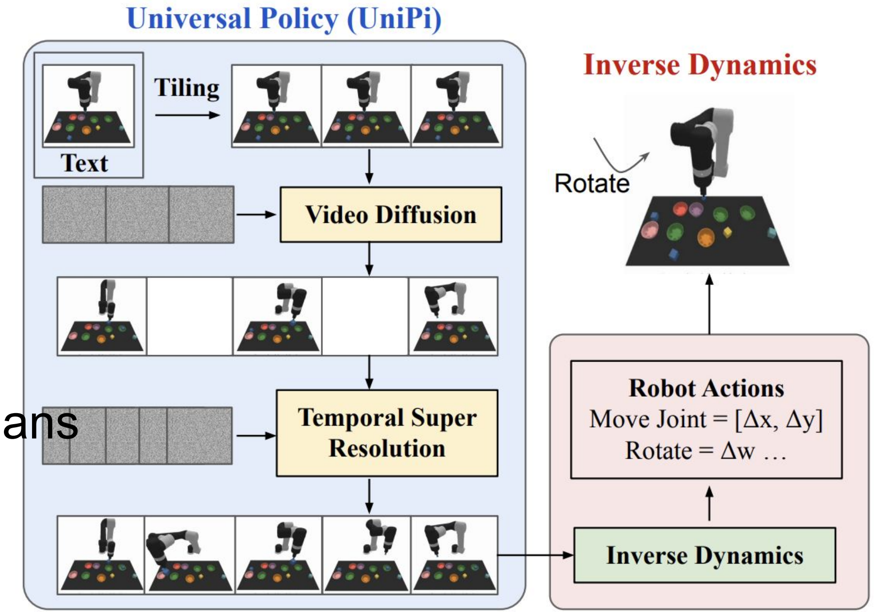
- Conditioned on the first frame

Trajectory Consistency through Tiling

- Frames are replicated across time

Hierarchical Planning

- Temporal super-resolution to refine plans



UniPi Implementation

Conditional Video Synthesis

- Conditioned on the first frame

Trajectory Consistency through Tiling

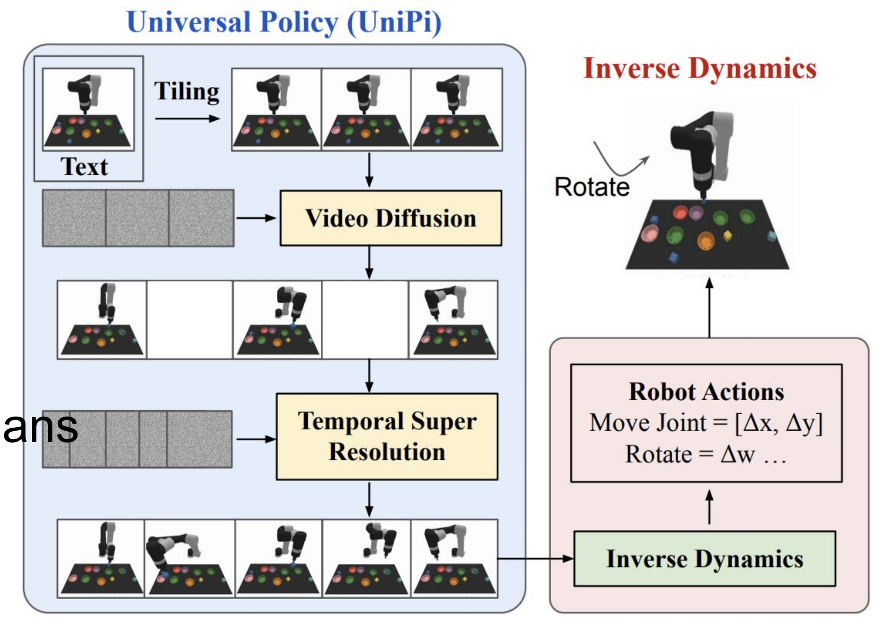
- Frames are replicated across time

Hierarchical Planning

- Temporal super-resolution to refine plans

Task Specific Action Adaptation

- Inverse dynamics model to recover control actions from videos



UniPi Implementation

Conditional Video Synthesis

- Conditioned on the first frame

Trajectory Consistency through Tiling

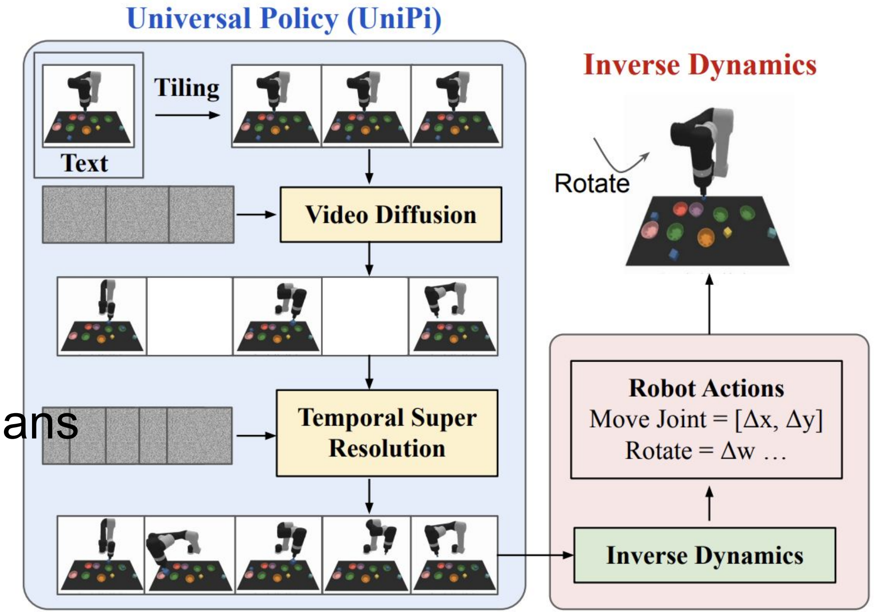
- Frames are replicated across time

Hierarchical Planning

- Temporal super-resolution to refine plans

Task Specific Action Adaptation

- Inverse dynamics model to recover control actions from videos

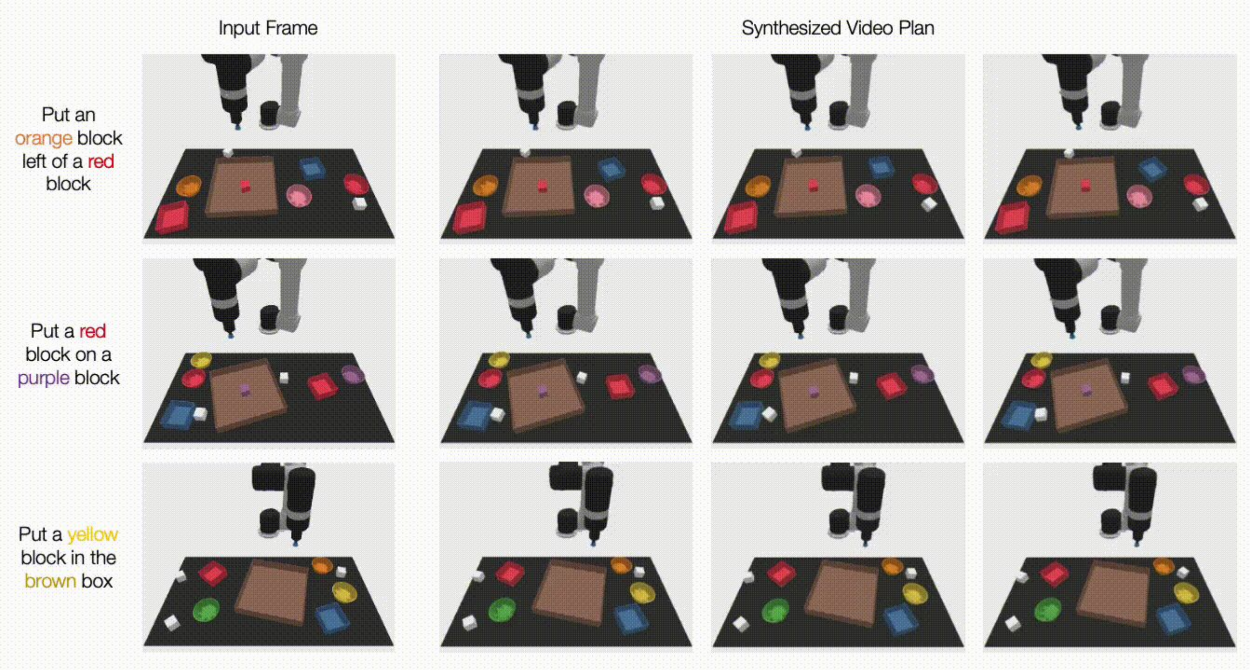


10B Imagen Video model

Tiny Regression model

UniPi Capabilities

Combinatorial Policy Synthesis



UniPi can synthesize a diverse set of different behaviors which satisfy unseen language subgoals.

UniPi Evaluation

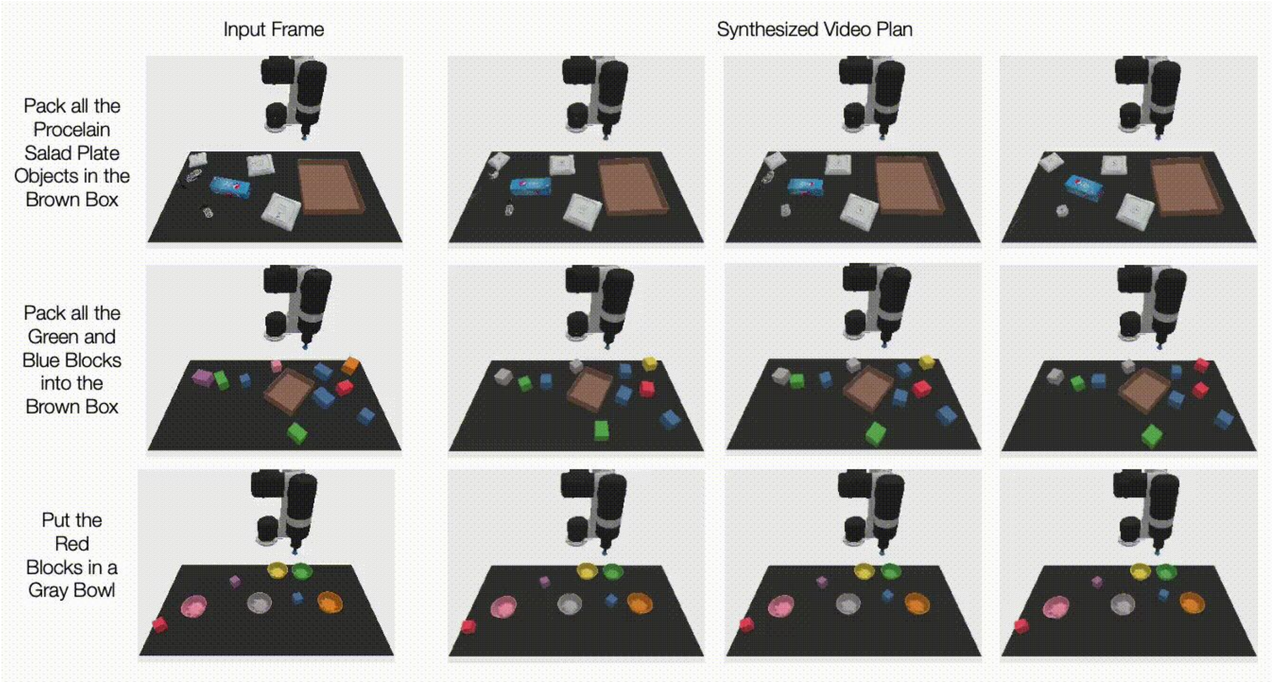
Combinatorial Generalization

Model	Seen		Novel	
	Place	**Relation**	**Place**	**Relation**
State + Transformer BC (Brohan et al., 2022)	19.4 ± 3.7	8.2 ± 2.0	11.9 ± 4.9	3.7 ± 2.1
Image + Transformer BC (Brohan et al., 2022)	9.4 ± 2.2	11.9 ± 1.8	9.7 ± 4.5	7.3 ± 2.6
Image + TT (Janner et al., 2021)	17.4 ± 2.9	12.8 ± 1.8	13.2 ± 4.1	9.1 ± 2.5
Diffuser (Janner et al., 2022)	9.0 ± 1.2	11.2 ± 1.0	12.5 ± 2.4	9.6 ± 1.7
UniPi (Ours)	$\mathbf{59.1} \pm 2.5$	$\mathbf{53.2} \pm 2.0$	$\mathbf{60.1} \pm 3.9$	$\mathbf{46.1} \pm 3.0$

Table 1. **Task Completion Accuracy on Combinatorial Environment.** UniPi generalizes well to both seen and novel combinations of language prompts in Place (e.g., place X in Y) and Relation (e.g., place X to the left of Y) tasks.

UniPi Capabilities

Multi-Environment Transfer



UniPi can synthesize a diverse set of different behaviors which satisfy unseen language tasks.

UniPi Evaluation

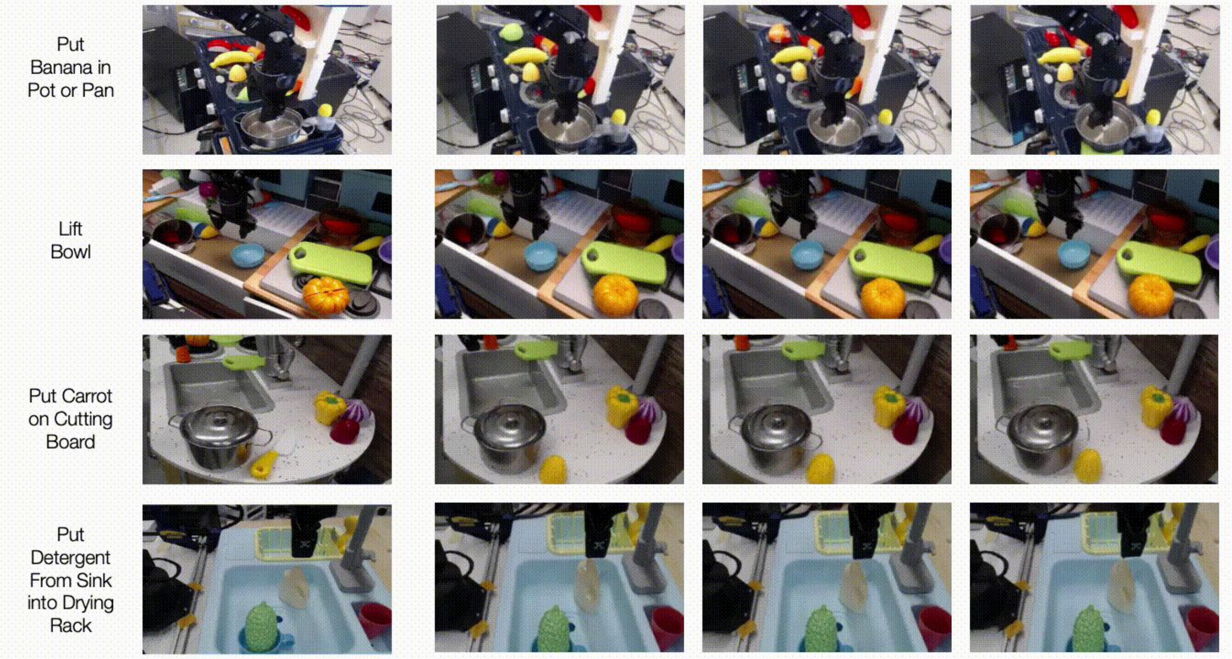
Multi-Task Generalization

Model	Place Bowl	Pack Object	Pack Pair
State + Transformer BC	9.8 ± 2.6	21.7 ± 3.5	1.3 ± 0.9
Image + Transformer BC	5.3 ± 1.9	5.7 ± 2.1	7.8 ± 2.6
Image + TT	4.9 ± 2.1	19.8 ± 0.4	2.3 ± 1.6
Diffuser	14.8 ± 2.9	15.9 ± 2.7	10.5 ± 2.4
UniPi (Ours)	$\mathbf{51.6} \pm 3.6$	$\mathbf{75.5} \pm 3.1$	$\mathbf{45.7} \pm 3.7$

Table 3. **Task Completion Accuracy on Multitask Environment.** UniPi generalizes well to new environments when trained on a set of different multi-task environments.

UniPi Capabilities

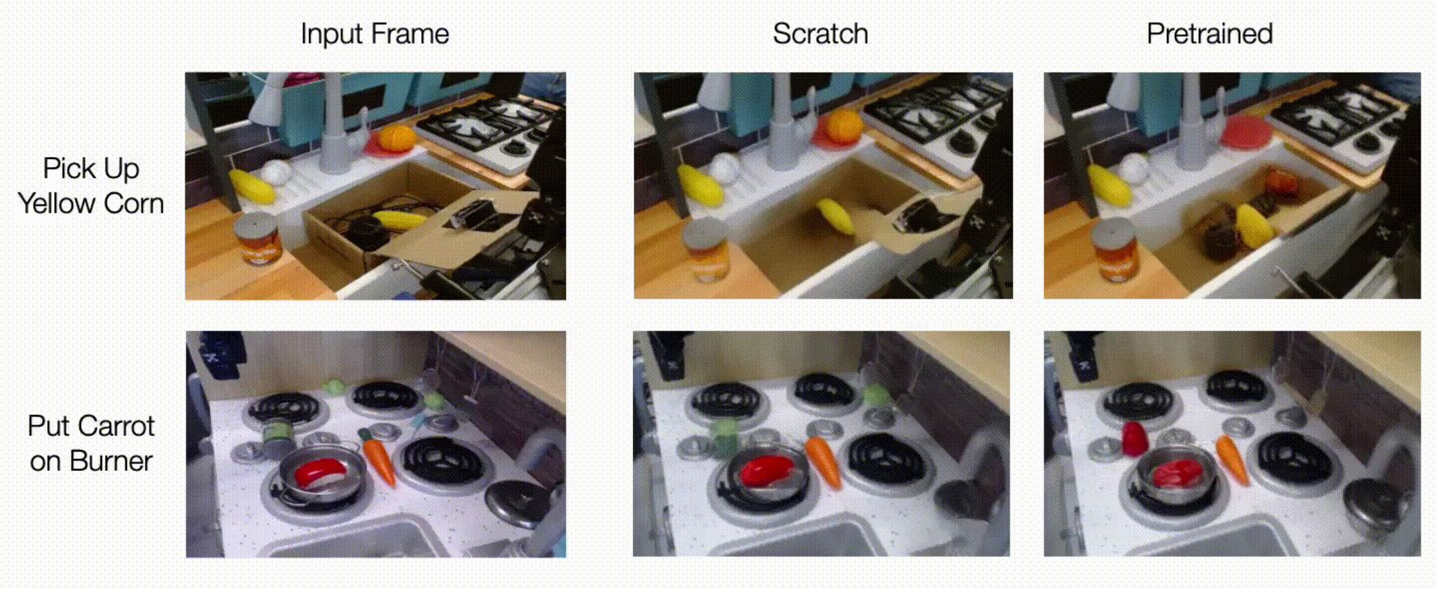
Real World Transfer



Given language instructions on unseen real images, UniPi is able to synthesize a diverse set of different behaviors which satisfy language instructions.

UniPi Capabilities

Pretraining on internet-scale data is important



UniPi Evaluation

Real-World Generalization: Pretraining on internet data is important

Model (24x40)	CLIP Score ↑	FID ↓	FVD ↓
No Pretrain	24.43 ± 0.04	17.75 ± 0.56	288.02 ± 10.45
Pretrain	$\mathbf{24.54} \pm 0.03$	$\mathbf{14.54} \pm 0.57$	$\mathbf{264.66} \pm 13.64$

Table 4. **Video Generation Quality of UniPi on Real Environment.** The use of existing data on the internet improves video plan predictions under all metrics considered.

UniPi Evaluation

Ablation: all components of UniPi are important

Frame Condition	Frame Consistency	Temporal Heirarchy	Place	Relation
No	No	No	13.2 ± 3.2	12.4 ± 2.4
Yes	No	No	52.4 ± 2.9	34.7 ± 2.6
Yes	Yes	No	53.2 ± 3.0	39.4 ± 2.8
Yes	Yes	Yes	$\mathbf{59.1} \pm 2.5$	$\mathbf{53.2} \pm 2.0$

Table 2. **Task Completion Accuracy Ablations.** Each component of UniPi improves its performance.

Questions?