シーン画像からのタスク実行可能なロボット動作軌跡の生成～検索拡張と拡散モデルの統合～

Norimichi Ukita
Toyota Technological Institute, Japan

PDF

Code

READ: Retrieval-Enhanced Asymmetric Diffusion for Motion Planning

Takeru Oba, Matthew Walter, and Norimichi Ukita

CVPR2024



Task: Motion Prediction from an Image

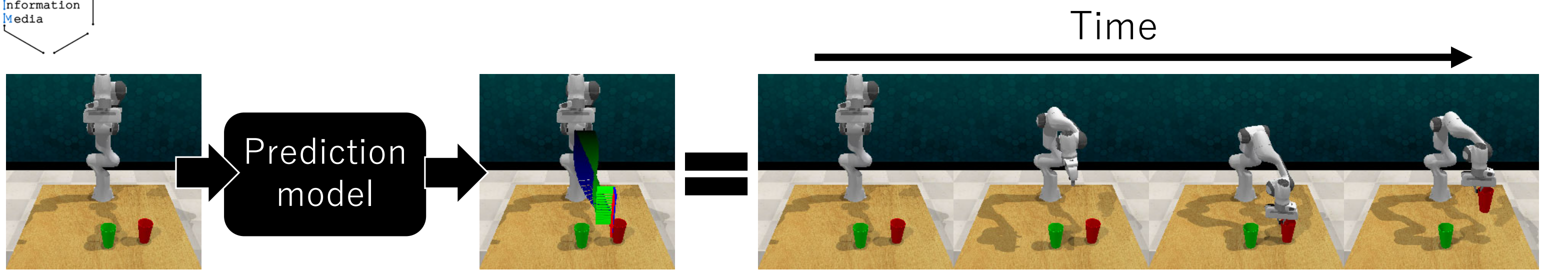


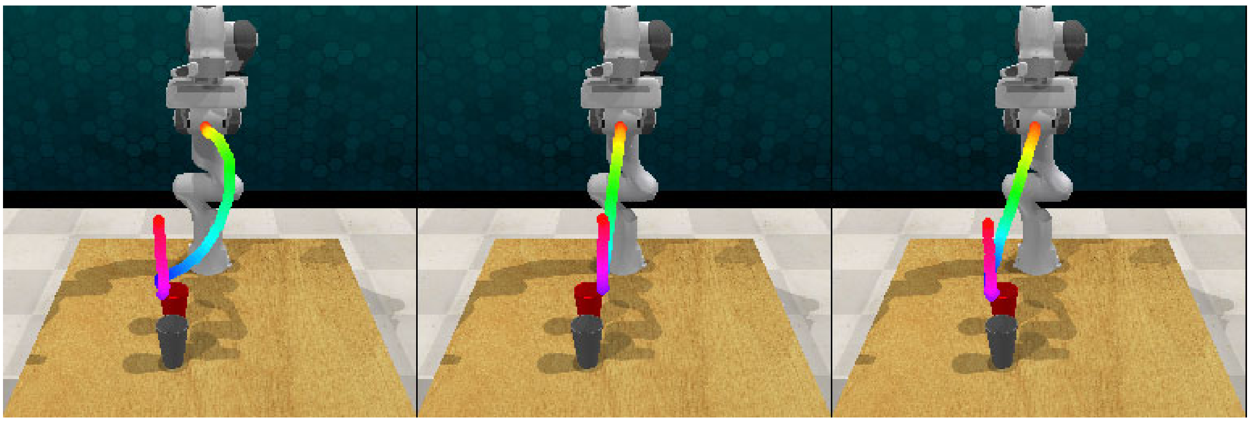
Image at t Motion from t+1 to T

Time: t Predicted Motion

1 [2.3, -1.0, 1.3, 0.6, 0.2, 0.9, 1.0]
[2.2, -0.9, 1.4, 0.6, 0.1, 1.0, 1.0]
⋮
[1.8, -0.9, 2.3, 0.5, 0.1, 1.6, 0.0]
T [1.7, -0.9, 2.3, 0.5, 0.1, 1.7, 0.0]

position angle grasp

Difficulty



- Multiple motions can accomplish the task
  - Demonstrators’ motions are stochastic.

Difficulty in Robot Motion Learning

- Stochasticity
  - Not only one but also several motions can achieve each task.
- Small number of training samples
  - Image generation >> Real robot motion planning
    - Robot motions are collected by manually controlling robots.
- Controllability
  - Complexity in articulated joint control
    - Similar motions can or cannot be achieved due to the limited range of joint motions.
- High precision
  - Small motion difference may disturb a task.

Difficulty in Robot Motion Learning

- Stochasticity
  - Probabilistic models
    - Representing multiple task-achievable motions
- Small number of training samples
  - Generative models
    - Successful in-distribution sampling from a limited number of samples
- Controllability
  - Retrieval-based motion planning
    - Motion optimization/refinement from real controllable motions
- High precision
  - High-fidelity motion refinement
    - Refinement in a high-resolution refinement space

Our Solutions for Robot Motion Learning

- Probabilistic models
  - Representing multiple task-achievable motions

➡ Diffusion models

- Generative models
  - Successful in-distribution sampling from a limited number of samples

➡ Diffusion models

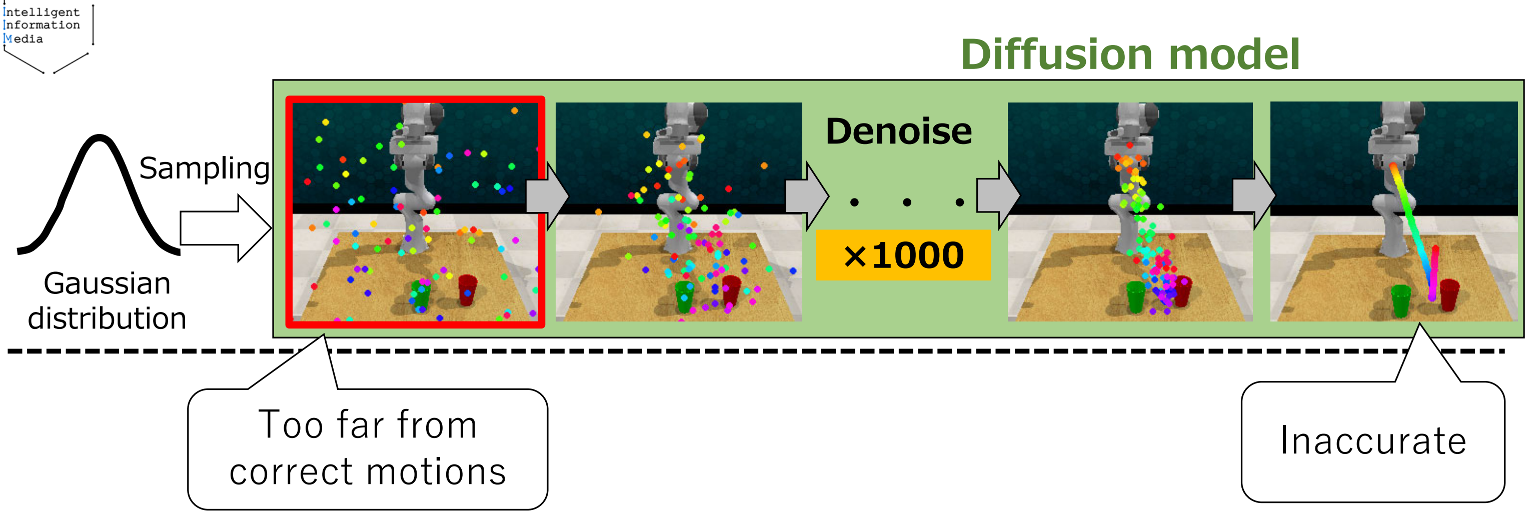
- Retrieval-based motion planning
  - Motion optimization/refinement from real controllable motions

➡ Refining real samples from intermediate diffusion steps

- High-fidelity motion refinement
  - Refinement in a high-resolution refinement space

➡ High-resolution feature space by asymmetric diffusion

Diffusion Model



Our Solutions for Robot Motion Learning

- Probabilistic models
  - Representing multiple task-achievable motions

➡ Diffusion models

- Generative models
  - Successful in-distribution sampling from a limited number of samples

➡ Diffusion models

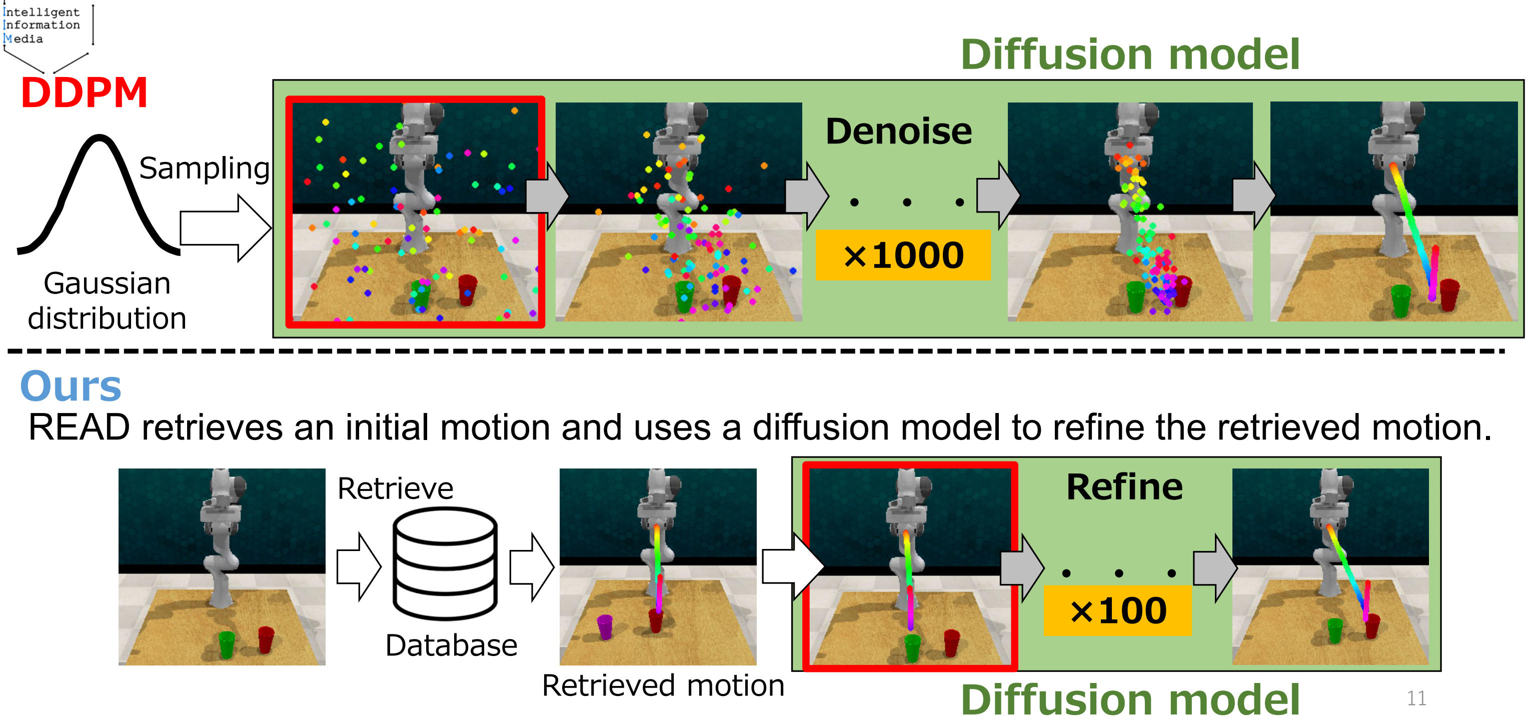
- Retrieval-based motion planning
  - Motion optimization/refinement from real controllable motions

➡ Refining real samples from intermediate diffusion steps

- High-fidelity motion refinement
  - Refinement in a high-resolution refinement space

➡ High-resolution feature space by asymmetric diffusion

Proposed Method: Refinement from **Retrieved** Motion



Our Solutions for Robot Motion Learning

- Probabilistic models
  - Representing multiple task-achievable motions

  ⇨ Diffusion models
- Generative models
  - Successful in-distribution sampling from a limited number of samples

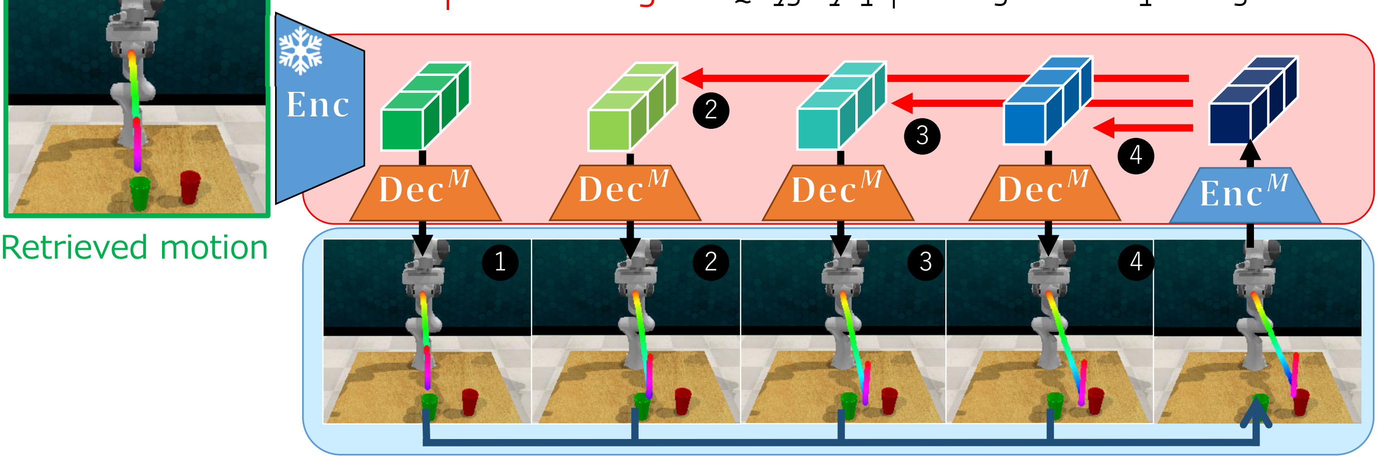
  ⇨ Diffusion models
- Retrieval-based motion planning
  - Motion optimization/refinement from real controllable motions

  ⇨ Refining real samples from intermediate diffusion steps
- High-fidelity motion refinement
  - Refinement in a high-resolution refinement space

  ⇨ High-resolution feature space by asymmetric diffusion

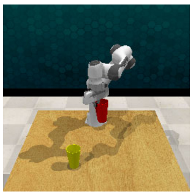
Proposed method: **Asymmetric** Diffusion

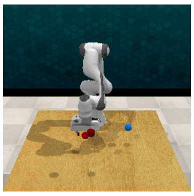
Ktw| fwi%uwthjxx?Qfyjsy%q|2inr jsxntsfq%xufhj



Wj{jwxj%uwthjxx?Twlnsfq%nlm2wjxtqzyntsxufhj

Experiments

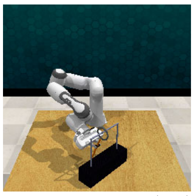

Pick up Cup (PC)

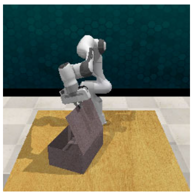

Reach Target (RT)


Push Button (PB)


Open Wine (OW)

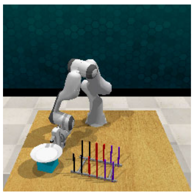

Water Plants (WP)

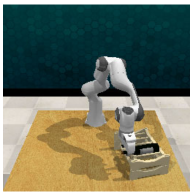

Beat the Buzz (BB)


Close Box (CB)


Put Rubbish (PR)


Put Knife (PK)


Take Plate (TP)


Stack Wine (SW)


Place Hanger (PH)

Bench mark：12 tasks
Training：**1000** for each task
Test：**100** for each task
Metrics: Success rate

Experimental Results

Task Success rates

		Avg12	PC	RT	PB	OW	CB	PR	SW	PH	WP	BB	PK	TP
no retrieval	Deterministic	72.7	64	74	88	65	85	77	99	39	95	61	35	91
	DDPM [18]	71.9	96	1	44	70	100	96	98	40	84	84	52	98
	VPSDE [44]	66.4	95	3	96	53	99	68	95	48	59	50	38	93
retrieval	VINN [34]	24.8	5	2	2	20	59	2	45	4	71	25	25	38
	DMO-EBM [32]	70.4	89	32	85	61	97	85	74	40	91	46	51	94
	VPSDE+CG [17]	19.5	26	10	5	6	72	6	30	1	51	7	5	15
	R2-Diff [33]	81.0	95	91	99	72	96	98	96	43	90	56	48	89
	READ	88.8	96	99	72	88	100	98	97	57	91	86	82	100

昨年度シンポジウムでの発表