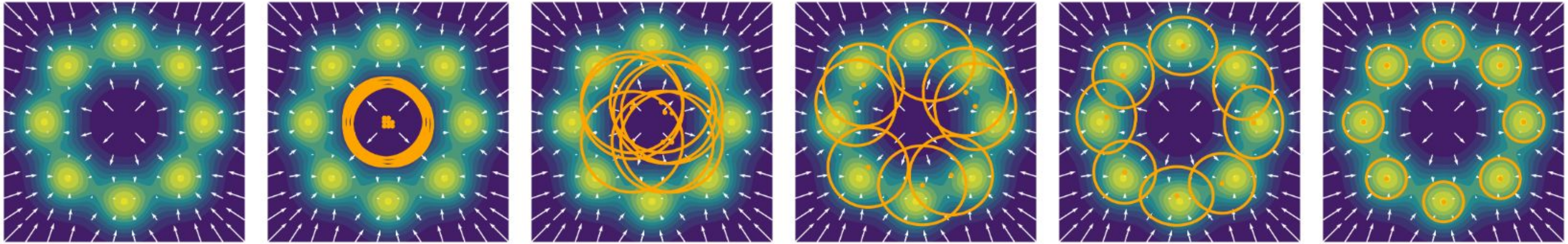
Variational Diffusion Distillation

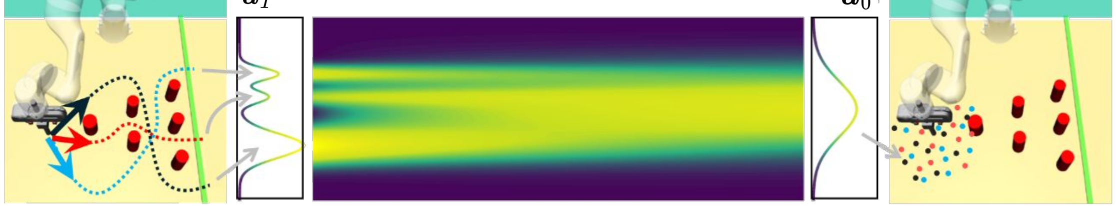
Variational Distillation of Diffusion Policies into Mixture of Experts

Hongyi Zhou, Denis Blessing, Ge Li, Onur Celik, Xiaogang Jia, Gerhard Neumann, Rudolf Lioutikov



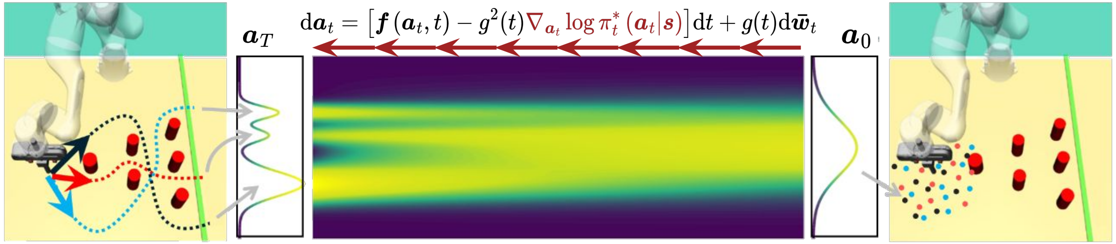
Variational Diffusion Distillation

$$\mathrm{d}\boldsymbol{a}_t = \left[\boldsymbol{f}\left(\boldsymbol{a}_t, t\right) - g^2(t)\nabla_{\boldsymbol{a}_t}\log\pi_t^*\left(\boldsymbol{a}_t|\boldsymbol{s}\right)\right]\mathrm{d}t + g(t)\mathrm{d}\bar{\boldsymbol{w}}_t$$



Diffusion

Variational Diffusion Distillation

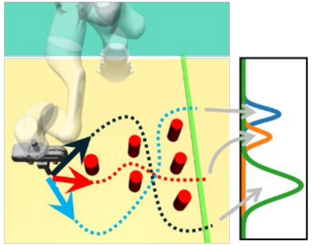


Diffusion

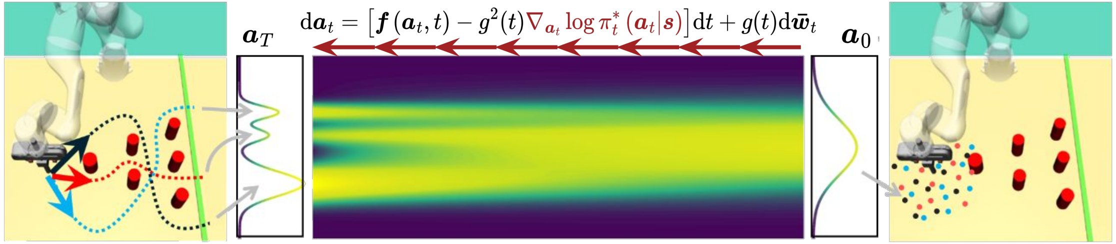
stable training | **slow inference**

intractable likelihood

Variational Diffusion Distillation



MoE

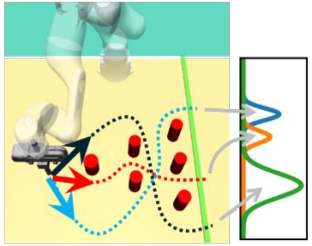


Diffusion

stable training | slow inference

intractable likelihood

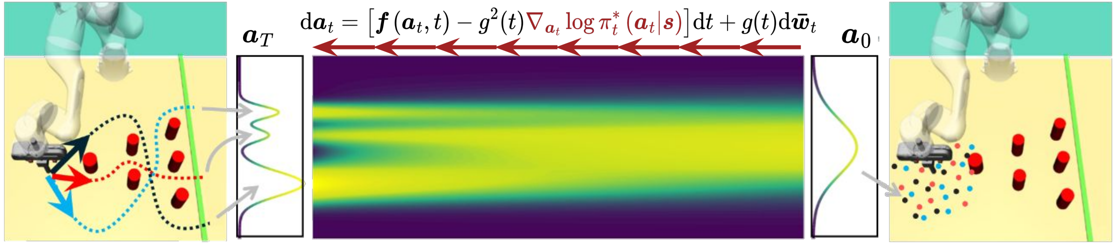
Variational Diffusion Distillation



MoE

tractable likelihood

unstable training | fast inference

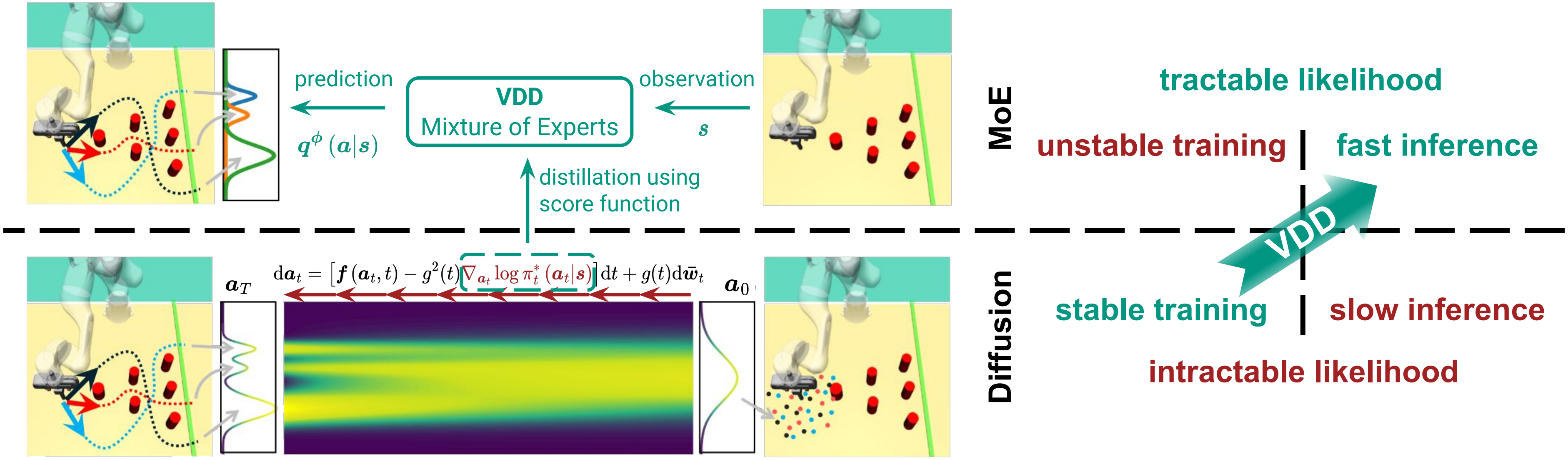


Diffusion

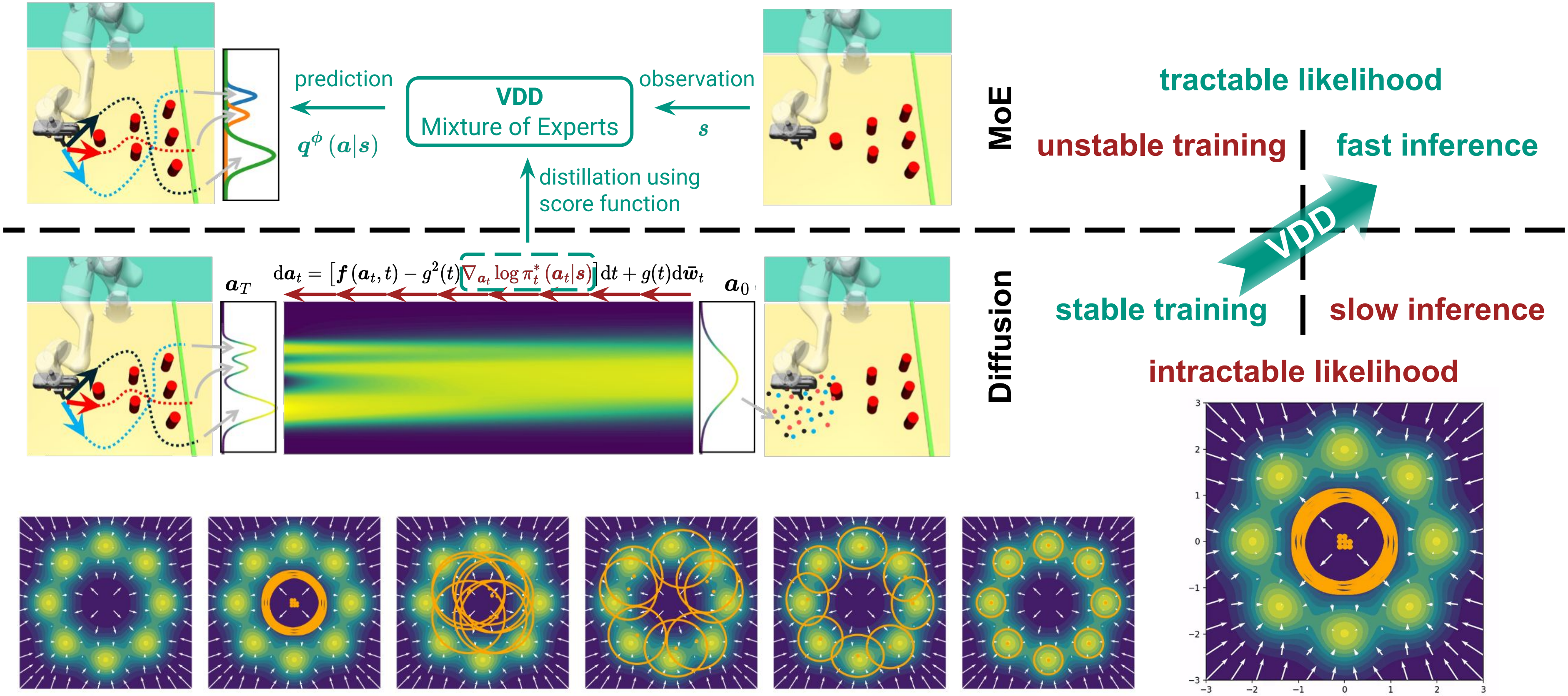
stable training | slow inference

intractable likelihood

Variational Diffusion Distillation



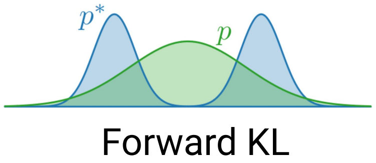
Variational Diffusion Distillation

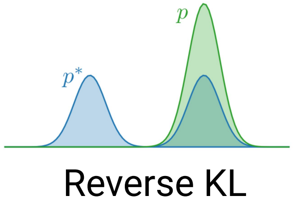


Policy Distillation via Variational Inference

Variational Inference minimize the expected reverse KL

$$\min_{\phi} J(\phi) = \min_{\phi} \mathbb{E}_{\mu(\mathbf{s})} D_{\mathrm{KL}}(q^{\phi}(\mathbf{a}|\mathbf{s}) \| \pi(\mathbf{a}|\mathbf{s}))$$

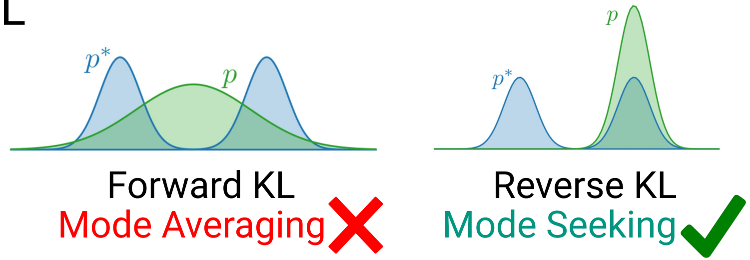




Policy Distillation via Variational Inference

Variational Inference minimize the expected reverse KL

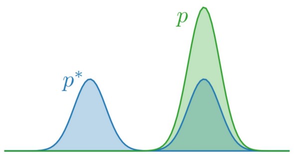
$$\min_{\phi} J(\phi) = \min_{\phi} \mathbb{E}_{\mu(\mathbf{s})} D_{\mathrm{KL}}(q^{\phi}(\mathbf{a}|\mathbf{s}) \| \pi(\mathbf{a}|\mathbf{s}))$$



Policy Distillation via Variational Inference

Variational Inference minimize the expected reverse KL

$$\min_{\phi} J(\phi) = \min_{\phi} \mathbb{E}_{\mu(\mathbf{s})} D_{\mathrm{KL}}(\underbrace{q^{\phi}(\mathbf{a}|\mathbf{s})}_{\text{Distilled Model}} \| \pi(\mathbf{a}|\mathbf{s}))$$



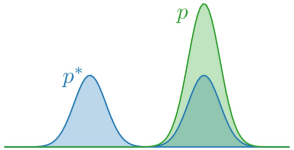
Reverse KL
Mode Seeking

Policy Distillation via Variational Inference

Variational Inference minimize the expected reverse KL

$$\min_{\phi} J(\phi) = \min_{\phi} \mathbb{E}_{\mu(\mathbf{s})} D_{\mathrm{KL}}(\underbrace{q^{\phi}(\mathbf{a}|\mathbf{s})}_{\text{Distilled Model}} \| \underbrace{\pi(\mathbf{a}|\mathbf{s})}_{\text{Diffusion Likelihood, Unknown}})$$

👎



Reverse KL

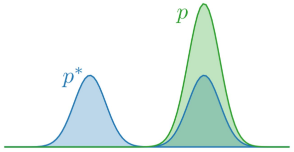
Mode Seeking ✓

Policy Distillation via Variational Inference

Variational Inference minimize the expected reverse KL

$$\min_{\phi} J(\phi) = \min_{\phi} \mathbb{E}_{\mu(\mathbf{s})} D_{\mathrm{KL}}(\underbrace{q^{\phi}(\mathbf{a}|\mathbf{s})}_{\text{Distilled Model}} \| \underbrace{\pi(\mathbf{a}|\mathbf{s})}_{\text{Diffusion Likelihood, Unknown}})$$

$$\nabla_{\phi} J(\phi) \approx \frac{M}{N} \sum_{\mathbf{s}_i \sim \mu} \mathbb{E}_{p(\boldsymbol{\epsilon})} \left[\nabla_{\phi} \log q^{\phi}(\boldsymbol{h}^{\phi}(\boldsymbol{\epsilon}, \mathbf{s}_i)|\mathbf{s}_i) - \nabla_{\phi} \log \pi(\boldsymbol{h}^{\phi}(\boldsymbol{\epsilon}, \mathbf{s}_i)|\mathbf{s}_i) \right]$$



Reverse KL

Mode Seeking

Policy Distillation via Variational Inference

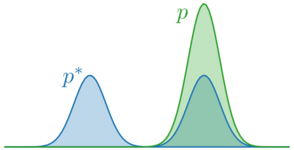
Variational Inference minimize the expected reverse KL

$$\min_{\phi} J(\phi) = \min_{\phi} \mathbb{E}_{\mu(\mathbf{s})} D_{\mathrm{KL}}(\underbrace{q^{\phi}(\mathbf{a}|\mathbf{s})}_{\text{Distilled Model}} \| \underbrace{\pi(\mathbf{a}|\mathbf{s})}_{\text{Diffusion Likelihood, Unknown}})$$

$$\nabla_{\phi} J(\phi) \approx \frac{M}{N} \sum_{\mathbf{s}_i \sim \mu} \mathbb{E}_{p(\boldsymbol{\epsilon})} \left[\nabla_{\phi} \log q^{\phi}(\boldsymbol{h}^{\phi}(\boldsymbol{\epsilon}, \mathbf{s}_i)|\mathbf{s}_i) - \nabla_{\phi} \log \pi(\boldsymbol{h}^{\phi}(\boldsymbol{\epsilon}, \mathbf{s}_i)|\mathbf{s}_i)\right]$$

Reparameterization Trick

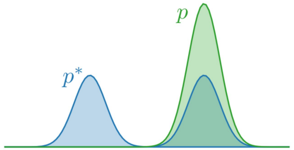
$$\underbrace{(\nabla_{\mathbf{a}} \log \pi(\mathbf{a}|\mathbf{s}_i))}_{\text{Diffusion Score, Known!}} \nabla_{\phi} \boldsymbol{h}^{\phi}(\boldsymbol{\epsilon}, \mathbf{s}_i)$$



Reverse KL
Mode Seeking

Policy Distillation via Variational Inference

Variational Inference minimize the expected reverse KL



Reverse KL
Mode Seeking ✓

$$\min_{\phi} J(\phi) = \min_{\phi} \mathbb{E}_{\mu(\mathbf{s})} D_{\mathrm{KL}}\Big(\underbrace{q^{\phi}(\mathbf{a}|\mathbf{s})}_{\text{Distilled Model}} \,\Big\|\, \underbrace{\pi(\mathbf{a}|\mathbf{s})}_{\text{Diffusion Likelihood, Unknown}}\Big)$$

$$\nabla_{\phi} J(\phi) \approx \frac{M}{N} \sum_{\mathbf{s}_i \sim \mu} \mathbb{E}_{p(\boldsymbol{\epsilon})} \left[\nabla_{\phi} \log q^{\phi}(\boldsymbol{h}^{\phi}(\boldsymbol{\epsilon}, \mathbf{s}_i)|\mathbf{s}_i) - \nabla_{\phi} \log \pi(\boldsymbol{h}^{\phi}(\boldsymbol{\epsilon}, \mathbf{s}_i)|\mathbf{s}_i) \right]$$

Reparameterization Trick

$$\underbrace{\left(\nabla_{\mathbf{a}} \log \pi(\mathbf{a}|\mathbf{s}_i)\right)}_{\text{Diffusion Score, Known!}} \nabla_{\phi} \boldsymbol{h}^{\phi}(\boldsymbol{\epsilon}, \mathbf{s}_i)$$

Similar objectives were also used in recent works[3, 4].

[1] Wang Z, Lu C, Wang Y, Bao F, Li C, Su H, Zhu J. Prolificdreamer: High-fidelity and diverse text-to-3d generation with variational score distillation. NeurIPS 2023.
[2] Chen H, Lu C, Wang Z, Su H, Zhu J. Score regularized policy optimization through diffusion behavior. ICLR 2024.

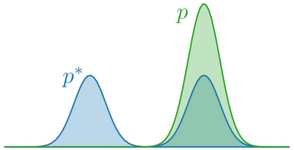
Policy Distillation via Variational Inference

NEURAL INFORMATION PROCESSING SYSTEMS

Karlsruhe Institute of Technology

Variational Inference minimize the expected reverse KL

$$\min_{\phi} J(\phi) = \min_{\phi} \mathbb{E}_{\mu(\mathbf{s})} D_{\mathrm{KL}}(\underbrace{q^{\phi}(\mathbf{a}|\mathbf{s})}_{\text{Distilled Model}} \| \underbrace{\pi(\mathbf{a}|\mathbf{s})}_{\text{Diffusion Likelihood, Unknown}})$$



Reverse KL
Mode Seeking ✓

$$\nabla_{\phi} J(\phi) \approx \frac{M}{N} \sum_{\mathbf{s}_i \sim \mu} \mathbb{E}_{p(\boldsymbol{\epsilon})} \left[\nabla_{\phi} \log q^{\phi}(\boldsymbol{h}^{\phi}(\boldsymbol{\epsilon}, \mathbf{s}_i)|\mathbf{s}_i) - \nabla_{\phi} \log \pi(\boldsymbol{h}^{\phi}(\boldsymbol{\epsilon}, \mathbf{s}_i)|\mathbf{s}_i)\right]$$

Reparameterization Trick

$$(\underbrace{\nabla_{\mathbf{a}} \log \pi(\mathbf{a}|\mathbf{s}_i)}_{\text{Diffusion Score, Known!}}) \nabla_{\phi} \boldsymbol{h}^{\phi}(\boldsymbol{\epsilon}, \mathbf{s}_i)$$

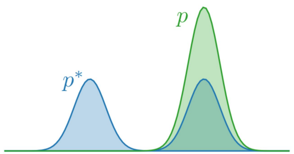
$$q^{\phi}(\mathbf{a}|\mathbf{s}) = \sum_{z} q^{\xi}(z|\mathbf{s}) q^{\nu_z}(\mathbf{a}|\mathbf{s}, z)$$

Similar objectives were also used in recent works[3, 4].
We are the first to distill diffusions into Mixture of Experts (MoEs)

[1] Wang Z, Lu C, Wang Y, Bao F, Li C, Su H, Zhu J. Prolificdreamer: High-fidelity and diverse text-to-3d generation with variational score distillation. NeurIPS 2023.
[2] Chen H, Lu C, Wang Z, Su H, Zhu J. Score regularized policy optimization through diffusion behavior. ICLR 2024.

Policy Distillation via Variational Inference

Variational Inference minimize the expected reverse KL



Reverse KL
Mode Seeking

$$\min_{\phi} J(\phi) = \min_{\phi} \mathbb{E}_{\mu(\mathbf{s})} D_{\mathrm{KL}}(\underbrace{q^{\phi}(\mathbf{a}|\mathbf{s})}_{\text{Distilled Model}} \| \underbrace{\pi(\mathbf{a}|\mathbf{s})}_{\text{Diffusion Likelihood, Unknown}})$$

$$\nabla_{\phi} J(\phi) \approx \frac{M}{N} \sum_{\mathbf{s}_i \sim \mu} \mathbb{E}_{p(\boldsymbol{\epsilon})} \left[\nabla_{\phi} \log q^{\phi}(\boldsymbol{h}^{\phi}(\boldsymbol{\epsilon}, \mathbf{s}_i)|\mathbf{s}_i) - \nabla_{\phi} \log \pi(\boldsymbol{h}^{\phi}(\boldsymbol{\epsilon}, \mathbf{s}_i)|\mathbf{s}_i) \right]$$

Reparameterization Trick

$$\underbrace{(\nabla_{\mathbf{a}} \log \pi(\mathbf{a}|\mathbf{s}_i))}_{\text{Diffusion Score, Known!}} \nabla_{\phi} \boldsymbol{h}^{\phi}(\boldsymbol{\epsilon}, \mathbf{s}_i)$$

$$q^{\phi}(\mathbf{a}|\mathbf{s}) = \sum_{z} \underbrace{q^{\xi}(z|\mathbf{s})}_{\text{Gating (Categorical)}} q^{\nu_z}(\mathbf{a}|\mathbf{s}, z)$$

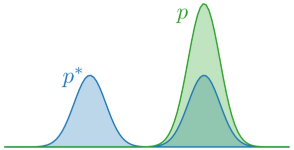
Similar objectives were also used in recent works[3, 4].
We are the first to distill diffusions into Mixture of Experts (MoEs)

[1] Wang Z, Lu C, Wang Y, Bao F, Li C, Su H, Zhu J. Prolificdreamer: High-fidelity and diverse text-to-3d generation with variational score distillation. NeurIPS 2023.
[2] Chen H, Lu C, Wang Z, Su H, Zhu J. Score regularized policy optimization through diffusion behavior. ICLR 2024.

Policy Distillation via Variational Inference

Variational Inference minimize the expected reverse KL

$$\min_{\phi} J(\phi) = \min_{\phi} \mathbb{E}_{\mu(\mathbf{s})} D_{\mathrm{KL}}(\underbrace{q^{\phi}(\mathbf{a}|\mathbf{s})}_{\text{Distilled Model}} \| \underbrace{\pi(\mathbf{a}|\mathbf{s})}_{\text{Diffusion Likelihood, Unknown}})$$



Reverse KL
Mode Seeking

$$\nabla_{\phi} J(\phi) \approx \frac{M}{N} \sum_{\mathbf{s}_i \sim \mu} \mathbb{E}_{p(\boldsymbol{\epsilon})} \left[\nabla_{\phi} \log q^{\phi}(\boldsymbol{h}^{\phi}(\boldsymbol{\epsilon}, \mathbf{s}_i)|\mathbf{s}_i) - \nabla_{\phi} \log \pi(\boldsymbol{h}^{\phi}(\boldsymbol{\epsilon}, \mathbf{s}_i)|\mathbf{s}_i) \right]$$

Reparameterization Trick

$$\underbrace{(\nabla_{\mathbf{a}} \log \pi(\mathbf{a}|\mathbf{s}_i))}_{\text{Diffusion Score, Known!}} \nabla_{\phi} \boldsymbol{h}^{\phi}(\boldsymbol{\epsilon}, \mathbf{s}_i)$$

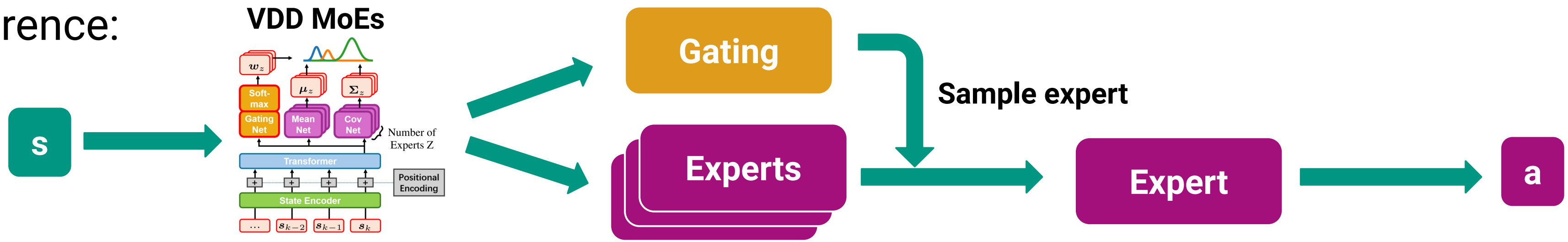
$$q^{\phi}(\mathbf{a}|\mathbf{s}) = \sum_{z} \underbrace{q^{\xi}(z|\mathbf{s})}_{\text{Gating (Categorical)}} \underbrace{q^{\nu_z}(\mathbf{a}|\mathbf{s}, z)}_{\text{Expert (Gaussian)}}$$

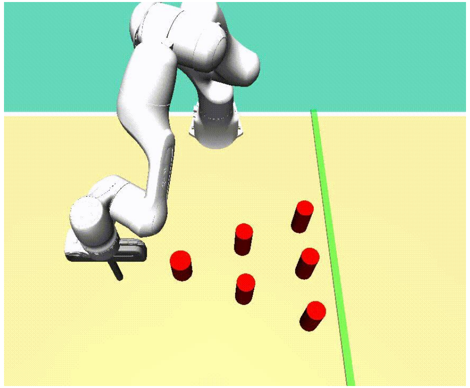
Similar objectives were also used in recent works[3, 4].
We are the first to distill diffusions into Mixture of Experts (MoEs)

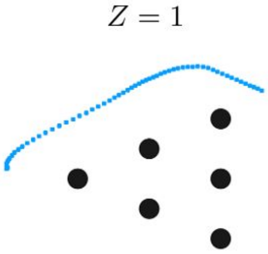
[1] Wang Z, Lu C, Wang Y, Bao F, Li C, Su H, Zhu J. Prolificdreamer: High-fidelity and diverse text-to-3d generation with variational score distillation. NeurIPS 2023.
[2] Chen H, Lu C, Wang Z, Su H, Zhu J. Score regularized policy optimization through diffusion behavior. ICLR 2024.

Variational Diffusion Distillation (VDD)

Inference:

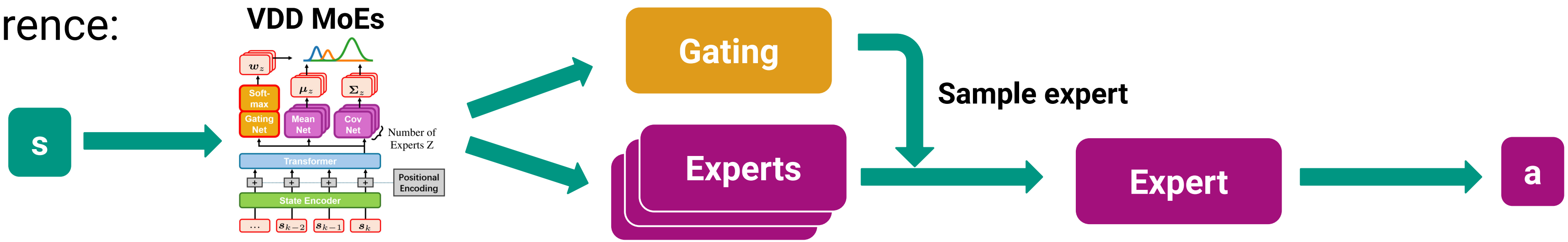


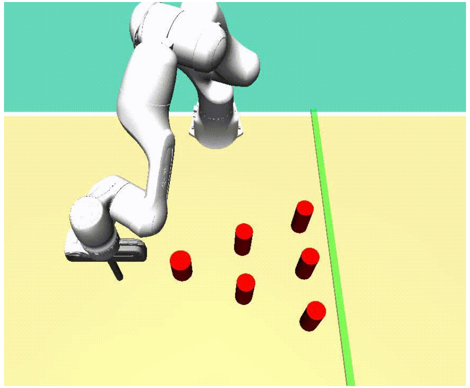


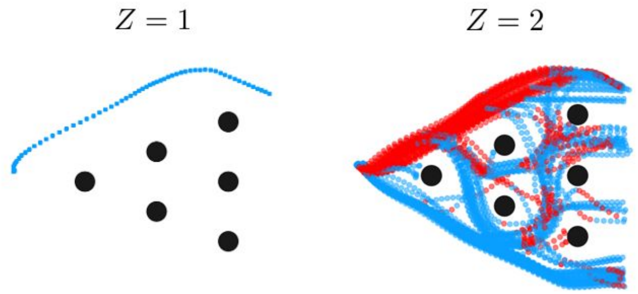


Variational Diffusion Distillation (VDD)

Inference:







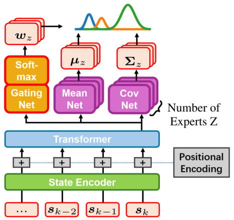
Variational Diffusion Distillation (VDD)

NEURAL INFORMATION PROCESSING SYSTEMS

KIT Karlsruhe Institute of Technology

Inference:

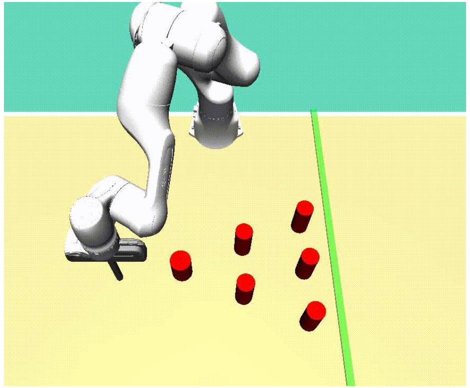
VDD MoEs

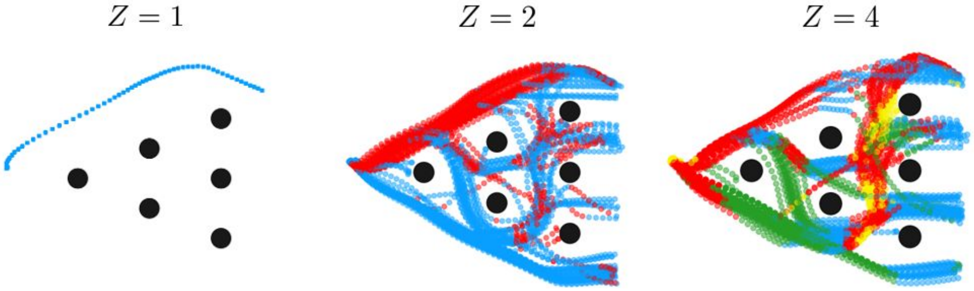


s → Gating / Experts

Sample expert

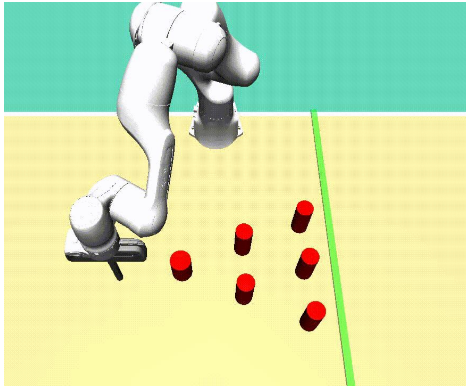
→ Expert → a

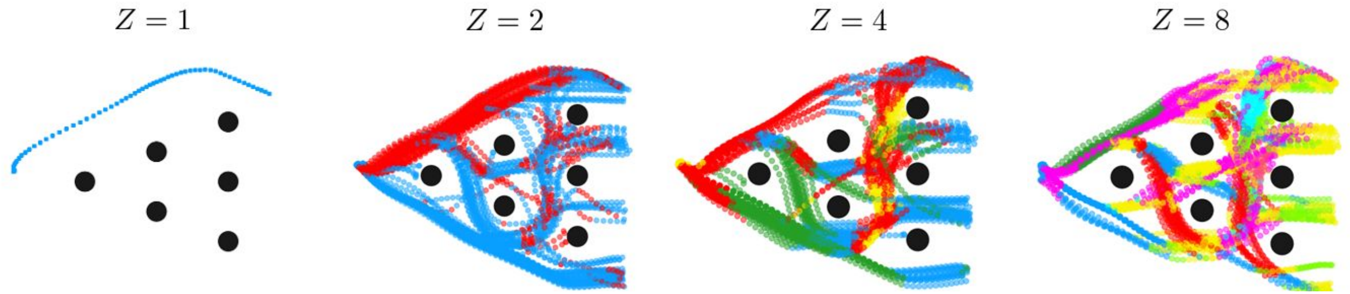




Variational Diffusion Distillation (VDD)

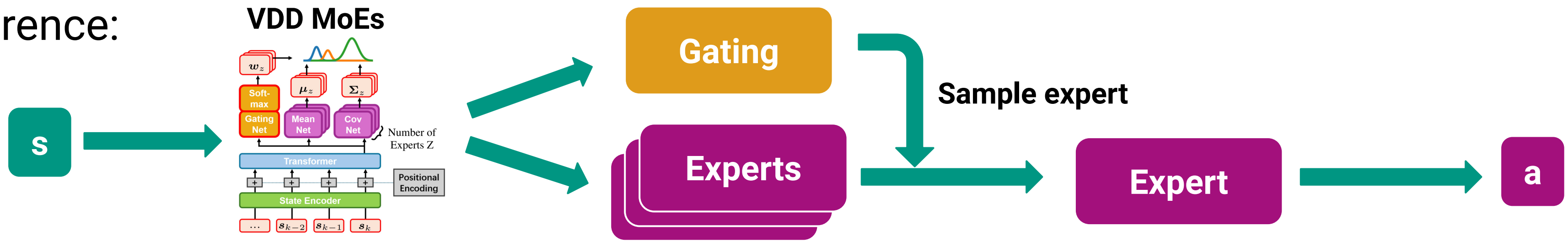
Inference:

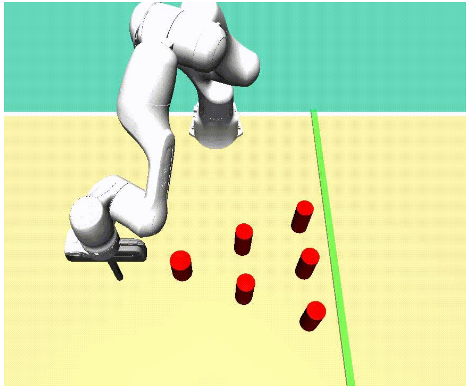


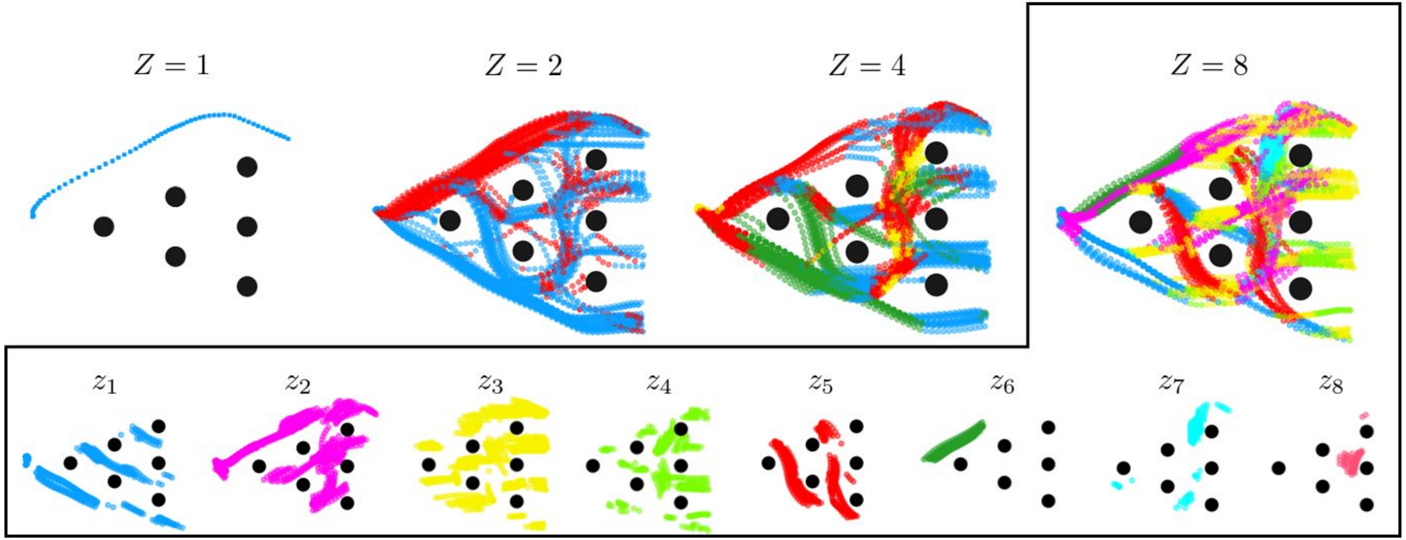


Variational Diffusion Distillation (VDD)

Inference:







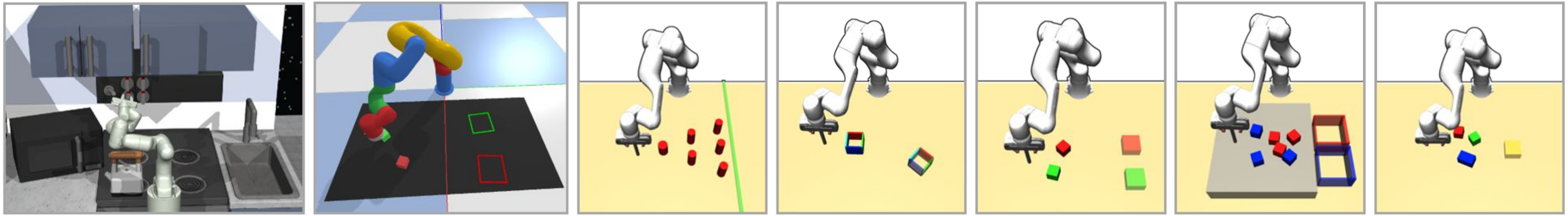
Experiment Results

Better than learning MoEs from scratch [3][4]

Environments	EM-GPT	IMC-GPT	VDD-VP	VDD-VE	EM-GPT	IMC-GPT	VDD-VP	VDD-VE
Avoiding	0.65 ± 0.18	0.75 ± 0.08	0.92 ± 0.02	$\mathbf{0.95 \pm 0.01}$	0.17 ± 0.13	$\mathbf{0.82 \pm 0.05}$	0.37 ± 0.01	0.73 ± 0.09
Aligning	0.78 ± 0.04	0.83 ± 0.02	0.70 ± 0.07	$\mathbf{0.86 \pm 0.04}$	0.38 ± 0.11	0.27 ± 0.09	0.25 ± 0.09	$\mathbf{0.40 \pm 0.04}$
Pushing	0.16 ± 0.07	0.76 ± 0.04	0.61 ± 0.04	$\mathbf{0.85 \pm 0.02}$	0.14 ± 0.10	0.31 ± 0.03	0.66 ± 0.05	$\mathbf{0.69 \pm 0.08}$
Stacking-1	0.58 ± 0.06	0.54 ± 0.05	0.81 ± 0.08	$\mathbf{0.83 \pm 0.09}$	$\mathbf{0.43 \pm 0.08}$	0.37 ± 0.04	0.19 ± 0.05	0.16 ± 0.03
Stacking-2	0.34 ± 0.07	0.29 ± 0.07	$\mathbf{0.60 \pm 0.07}$	0.57 ± 0.06	$\mathbf{0.27 \pm 0.05}$	0.17 ± 0.07	0.07 ± 0.04	0.13 ± 0.06
Sorting (image)	0.69 ± 0.02	0.74 ± 0.04	$\mathbf{0.80 \pm 0.04}$	0.76 ± 0.03	0.13 ± 0.03	0.10 ± 0.03	0.12 ± 0.03	$\mathbf{0.22 \pm 0.03}$
Stacking (image)	0.04 ± 0.03	0.39 ± 0.10	$\mathbf{0.78 \pm 0.02}$	0.60 ± 0.04	0.00 ± 0.00	0.05 ± 0.04	0.08 ± 0.02	$\mathbf{0.11 \pm 0.03}$
Relay Kitchen	3.62 ± 0.10	3.67 ± 0.05	3.24 ± 0.12	$\mathbf{3.85 \pm 0.10}$	-	-	-	-
Block Push	0.88 ± 0.04	0.89 ± 0.04	$\mathbf{0.93 \pm 0.03}$	0.91 ± 0.03	-	-	-	-

Success Rates

Entropy



Experiment Results

Compare with Consistency Distillation [5] and Consistency Trajectory Models [6, 7]

	VP (DDPM)	VE (BESO)	VP-1	VE-1	CD-VE	CTM-VE	VDD-VP(ours)	VDD-VE(ours)
Kitchen	3.35	4.06	0.22	4.02	3.87 ± 0.05	$\mathbf{3.89 \pm 0.11}$	3.24 ± 0.12	3.85 ± 0.10
Block Push	0.96	0.96	0.09	0.94	0.89 ± 0.05	0.89 ± 0.04	0.93 ± 0.03	$\mathbf{0.91 \pm 0.03}$
Avoiding	0.94	0.96	0.09	0.84	0.82 ± 0.05	0.93 ± 0.02	0.92 ± 0.02	$\mathbf{0.95 \pm 0.01}$
Aligning	0.85	0.85	0.00	0.93	$\mathbf{0.94 \pm 0.08}$	0.81 ± 0.11	0.70 ± 0.07	0.86 ± 0.04
Pushing	0.74	0.78	0.00	0.70	0.66 ± 0.05	0.80 ± 0.07	0.61 ± 0.04	$\mathbf{0.85 \pm 0.02}$
Stacking-1	0.89	0.91	0.00	0.75	0.69 ± 0.06	0.54 ± 0.17	0.81 ± 0.08	$\mathbf{0.85 \pm 0.02}$
Stacking-2	0.68	0.70	0.00	0.53	0.46 ± 0.11	0.30 ± 0.09	$\mathbf{0.60 \pm 0.07}$	0.57 ± 0.06
Sorting (Image)	0.69	0.70	0.20	0.68	0.71 ± 0.07	0.70 ± 0.07	$\mathbf{0.80 \pm 0.04}$	0.76 ± 0.04
Stacking (Image)	0.58	0.66	0.00	0.58	0.63 ± 0.01	0.59 ± 0.10	$\mathbf{0.78 \pm 0.02}$	0.60 ± 0.04

(a) Task Success Rate (or Environment Return for Kitchen)

	VP (DDPM)	VE (BESO)	VP-1	VE-1	CD-VE	CTM-VE	VDD-VP(ours)	VDD-VE(ours)
Avoiding	0.89	0.87	0.25	0.76	0.72 ± 0.02	$\mathbf{0.79 \pm 0.04}$	0.37 ± 0.01	0.72 ± 0.12
Aligning	0.62	0.67	0.00	0.34	0.32 ± 0.14	0.31 ± 0.28	0.25 ± 0.09	$\mathbf{0.40 \pm 0.04}$
Pushing	0.74	0.76	0.00	0.50	0.53 ± 0.07	0.54 ± 0.08	0.66 ± 0.05	$\mathbf{0.69 \pm 0.08}$
Stacking-1	0.24	0.30	0.00	0.26	$\mathbf{0.19 \pm 0.12}$	0.18 ± 0.08	$\mathbf{0.19 \pm 0.05}$	0.16 ± 0.03
Stacking-2	0.12	0.13	0.00	0.07	0.03 ± 0.05	0.09 ± 0.06	0.07 ± 0.04	$\mathbf{0.13 \pm 0.06}$
Sorting (Image)	0.16	0.19	0.09	0.14	0.14 ± 0.06	0.08 ± 0.05	0.12 ± 0.03	$\mathbf{0.22 \pm 0.03}$
Stacking (Image)	0.31	0.15	0.00	0.10	0.06 ± 0.01	0.04 ± 0.04	0.05 ± 0.02	$\mathbf{0.11 \pm 0.03}$

(b) Task Entropy

References

[1] Wang Z, Lu C, Wang Y, Bao F, Li C, Su H, Zhu J. Prolificdreamer: High-fidelity and diverse text-to-3d generation with variational score distillation. NeurIPS 2023.
[2] Chen H, Lu C, Wang Z, Su H, Zhu J. Score regularized policy optimization through diffusion behavior. ICLR 2024.
[3]] Todd K Moon. The expectation-maximization algorithm. IEEE Signal processing magazine, 13 (6):47–60, 1996.
[4] Blessing D, Celik O, Jia X, Reuss M, Li MX, Lioutikov R, Neumann G. Information maximizing curriculum: A curriculum-based approach for training mixtures of experts. NeurIPS 2023.
[5]Song Y, Dhariwal P, Chen M, Sutskever I. Consistency models. ICML 2023.
[6]Kim D, Lai CH, Liao WH, Murata N, Takida Y, Uesaka T, He Y, Mitsufuji Y, Ermon S. Consistency trajectory models: Learning probability flow ode trajectory of diffusion. ICLR 2024.
[7]Prasad A, Lin K, Wu J, Zhou L, Bohg J. Consistency policy: Accelerated visuomotor policies via consistency distillation. RSS 2024.