Exact Guidance for Discrete Flow Matching

Liyan Xie

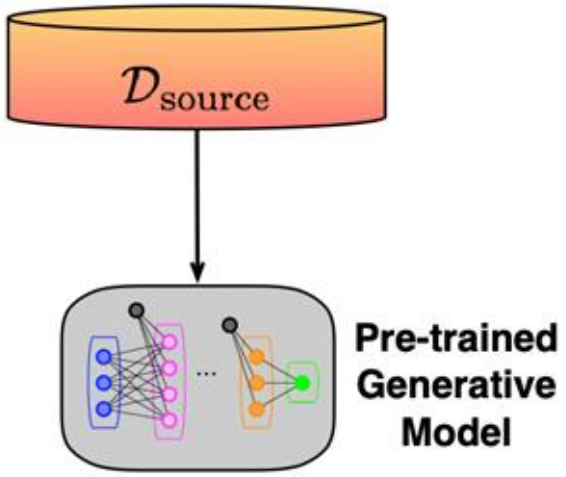
Department of Industrial and Systems Engineering

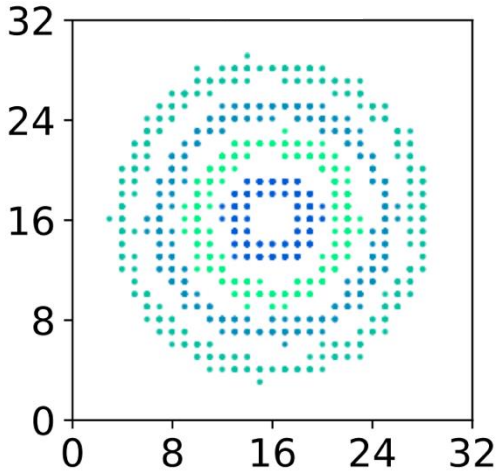
University of Minneasota

Joint work with Zhengyan Wan[1], Yidong Ouyang[2], Fang Fang[1], Hongyuan Zha[3], and Guang Cheng[2]

[1]East China Normal University; [2]UCLA; [3]CUHK-Shenzhen

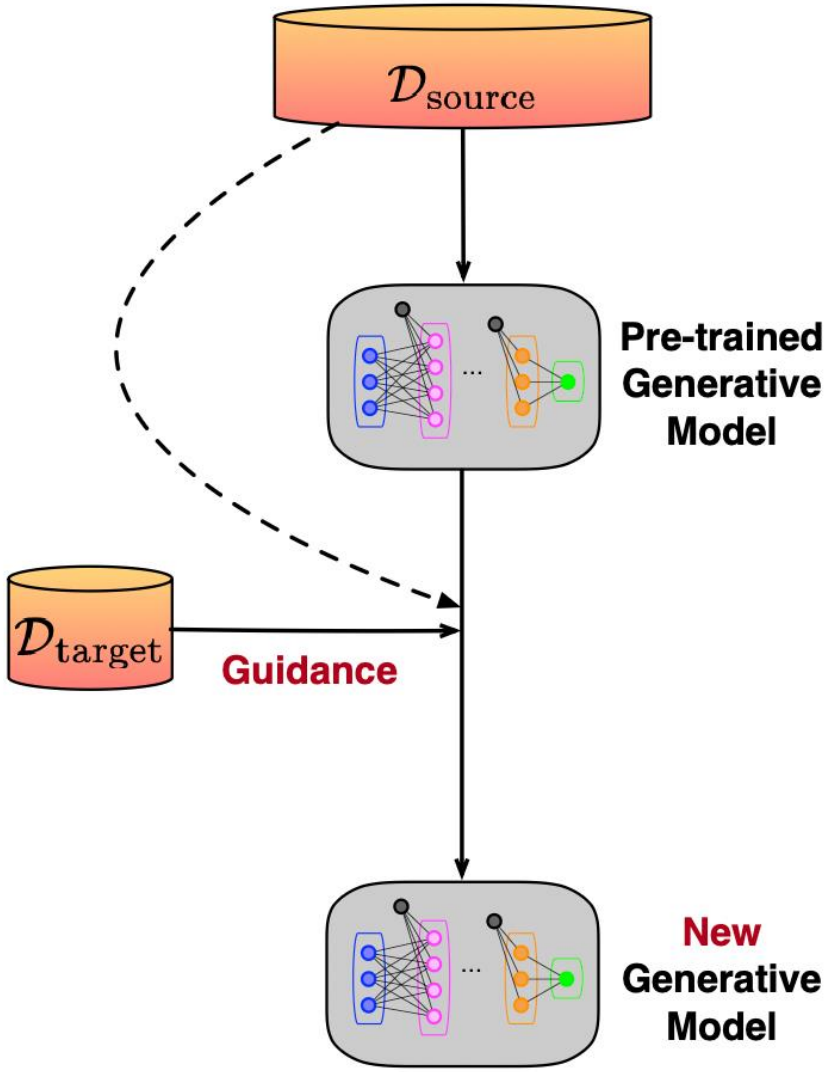
Guidance for Generative Models

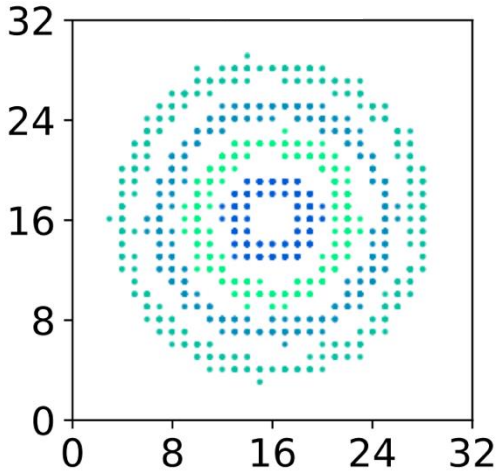




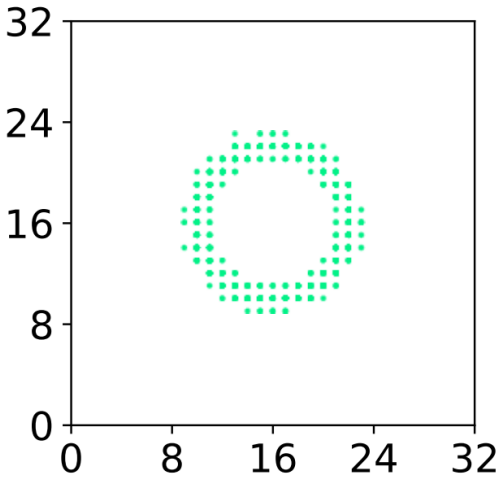
Source Distribution

Guidance for Generative Models



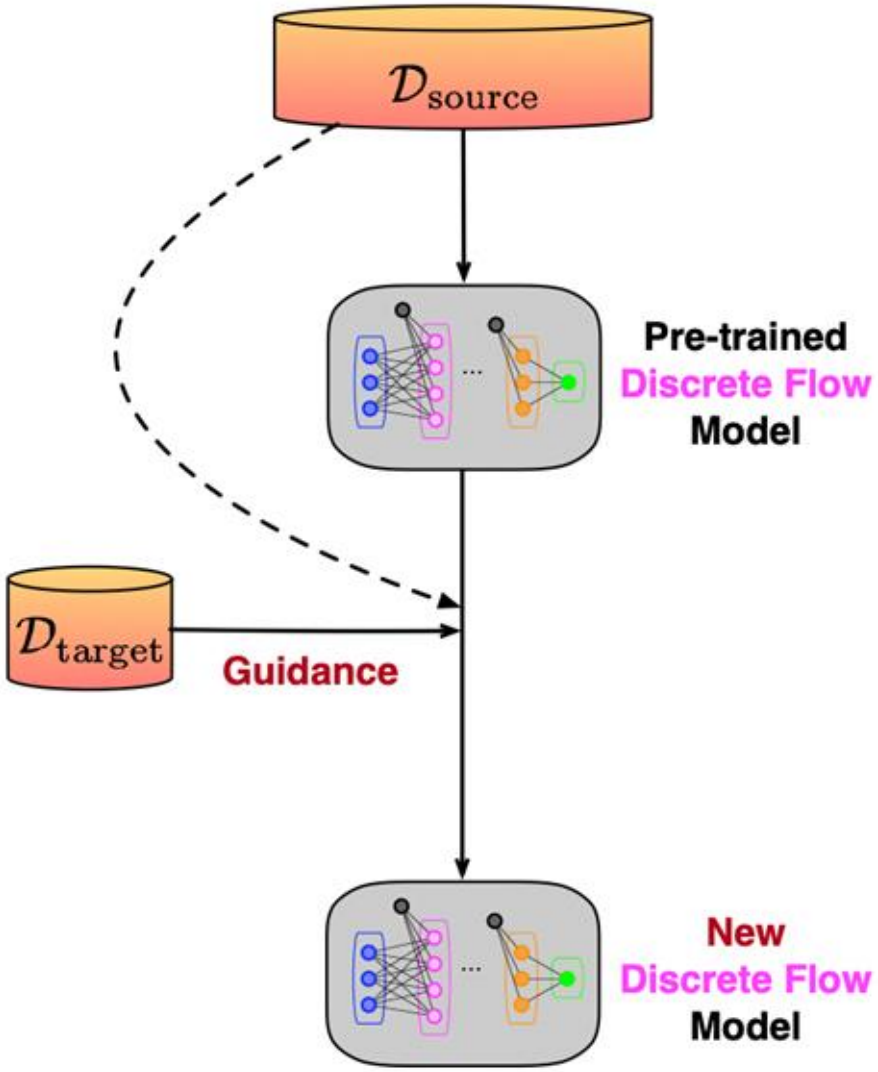


Source Distribution



Target Distribution

This work



Application Example #1: Text-to-Image Generation

A beautiful modern wooden house, close to the lake, in the mountains at sunrise, anime style

FUDOKI

Ours

Application Example #2: Multimodal Understanding

Question: Is the phone number in the picture "0131 555 6363"? Please answer yes or no.

	FUDOKI	Ours
Answer:	No. ❌	Yes. ✅

Wan, Z. et al. "Discrete guidance matching: Exact guidance for discrete flow matching." *ICLR 2026*.

Outline

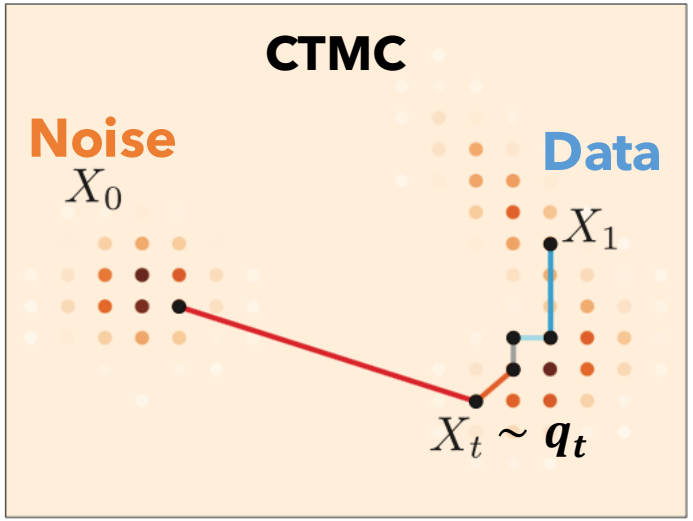
- A brief introduction of discrete flow models
- Proposed algorithm: density-ratio based exact guidance
- Numerical experiments

Discrete Flow Models

- Continuous Time Markov Processes

$$q_{t+h|t}(x \mid z) = \delta_z(x) + u_t^q(x, z)h + o(h)$$

Transition rate
(to be learned)



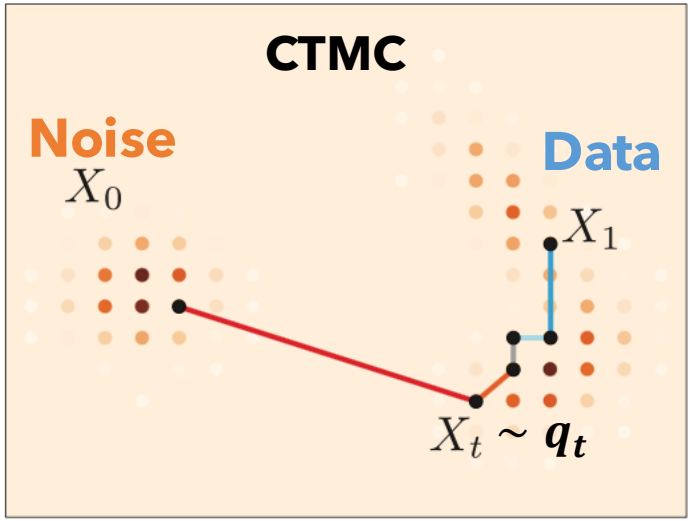
Campbell, A. et al. "Generative flows on discrete state-spaces: Enabling multimodal flows with applications to protein co-design. *ICML 2024*.

Discrete Flow Models

- Continuous Time Markov Processes

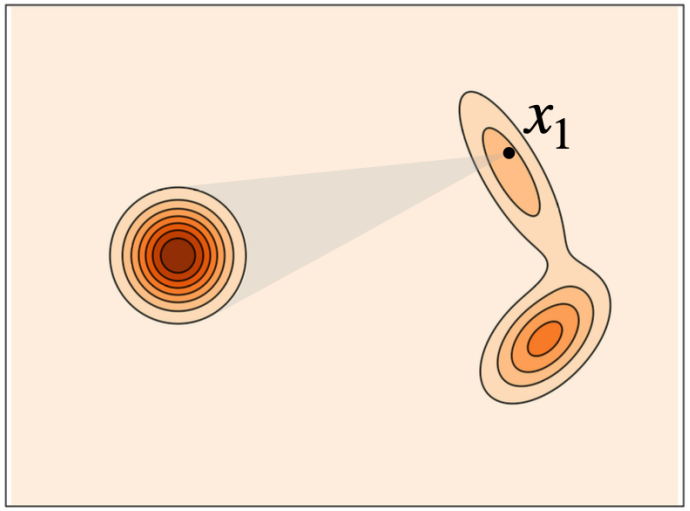
$$q_{t+h|t}(x \mid z) = \delta_z(x) + u_t^q(x, z)h + o(h)$$

Transition rate
(to be learned)



- Building from **conditional** generators

pre-specified $u_t^q(z, x|\mathbf{x}_1)$ ⟶ $q_{t|1}(\cdot|\mathbf{x}_1)$ pre-specified



-- Credit: online

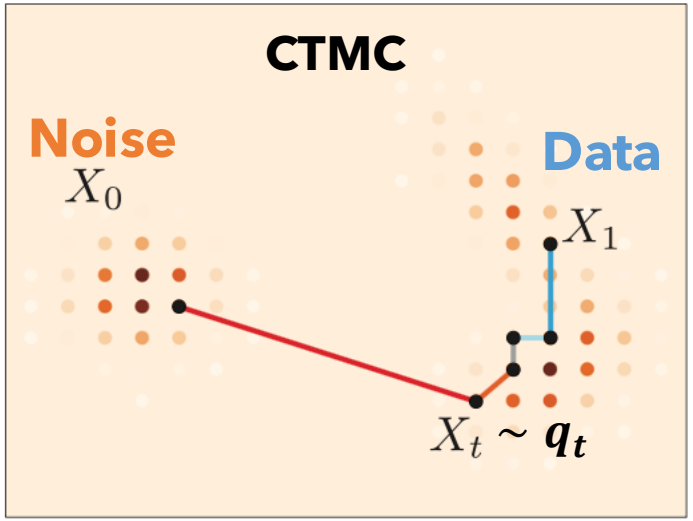
Campbell, A. et al. "Generative flows on discrete state-spaces: Enabling multimodal flows with applications to protein co-design. *ICML 2024*.

Discrete Flow Models

- Continuous Time Markov Processes

$$q_{t+h|t}(x \mid z) = \delta_z(x) + u_t^q(x, z)h + o(h)$$

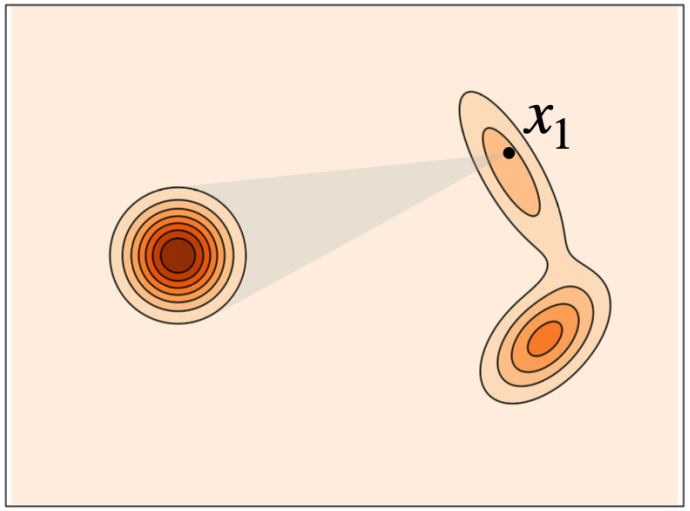
Transition rate
(to be learned)



- Building from **conditional** generators

pre-specified $u_t^q(z, x|\mathbf{x}_1)$ ⟶ $q_{t|1}(\cdot|\mathbf{x}_1)$ pre-specified

$$u_t^q(z, x) = \mathbb{E}_{q_{1|t}(\mathbf{x}_1|x)}[u_t^q(z, x|\mathbf{x}_1)] \longrightarrow q_t(\cdot)$$



-- Credit: online

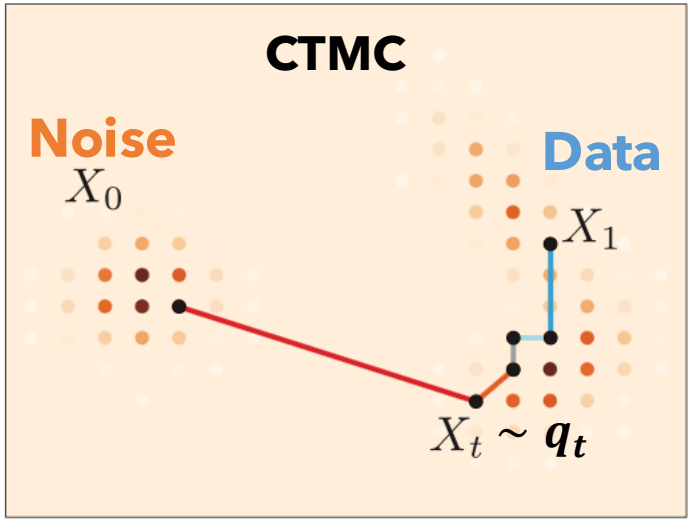
Campbell, A. et al. "Generative flows on discrete state-spaces: Enabling multimodal flows with applications to protein co-design. *ICML 2024*.

Discrete Flow Models

- Continuous Time Markov Processes

$$q_{t+h|t}(x \mid z) = \delta_z(x) + u_t^q(x, z)h + o(h)$$

Transition rate
(to be learned)

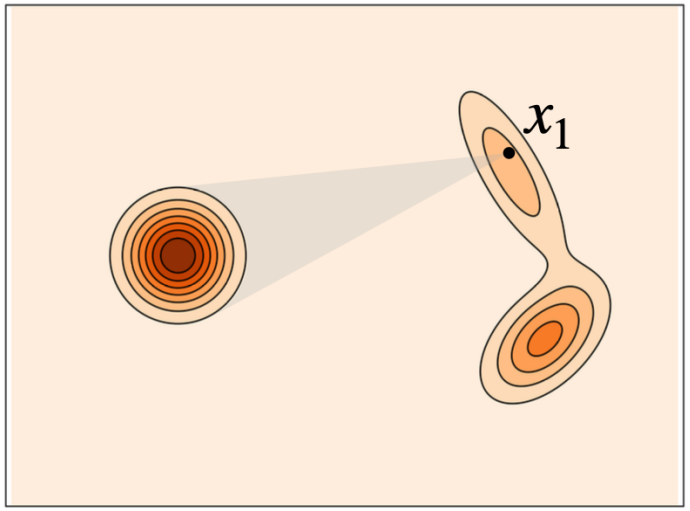


- Building from **conditional** generators

pre-specified $u_t^q(z, x|\mathbf{x}_1)$ ⟶ $q_{t|1}(\cdot|\mathbf{x}_1)$ pre-specified

$$u_t^q(z, x) = \mathbb{E}_{q_{1|t}(\mathbf{x}_1|x)}[u_t^q(z, x|\mathbf{x}_1)] \longrightarrow q_t(\cdot)$$

$$\approx \mathbb{E}_{q_{1|t}^{\theta}(\mathbf{x}_1|x)}[u_t^q(z, x|\mathbf{x}_1)]$$ **learned denoising model**



-- Credit: online

Campbell, A. et al. "Generative flows on discrete state-spaces: Enabling multimodal flows with applications to protein co-design. *ICML 2024*.

Main insights for guidance

Suppose that the conditional probability path of the **source** distribution $p_{t|1}$ and the **target** distribution $q_{t|1}$ are the same for any t. Then we have

$$q_{1|t}(z|x) = \frac{r(z)}{\mathbb{E}_{\mathbf{x}_1 \sim p_{1|t}(\mathbf{x}_1|x)}[r(\mathbf{x}_1)]} p_{1|t}(z|x)$$

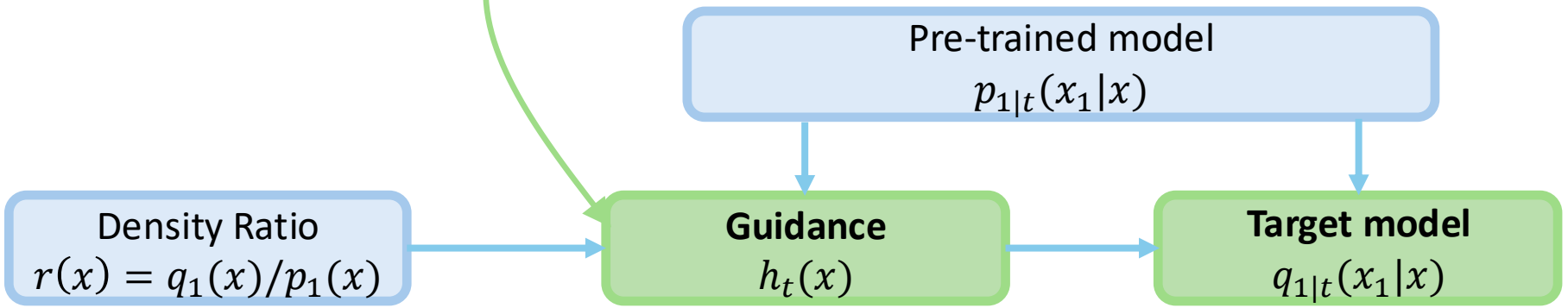
where $r(\cdot) = q_1(\cdot)/p_1(\cdot)$ is the density ratio.

Main insights for guidance

Suppose that the conditional probability path of the **source** distribution $p_{t|1}$ and the **target** distribution $q_{t|1}$ are the same for any t. Then we have

$$q_{1|t}(z|x) = \frac{r(z)}{\mathbb{E}_{\mathbf{x}_1 \sim p_{1|t}(\mathbf{x}_1|x)}[r(\mathbf{x}_1)]} p_{1|t}(z|x)$$

where $r(\cdot) = q_1(\cdot)/p_1(\cdot)$ is the density ratio.

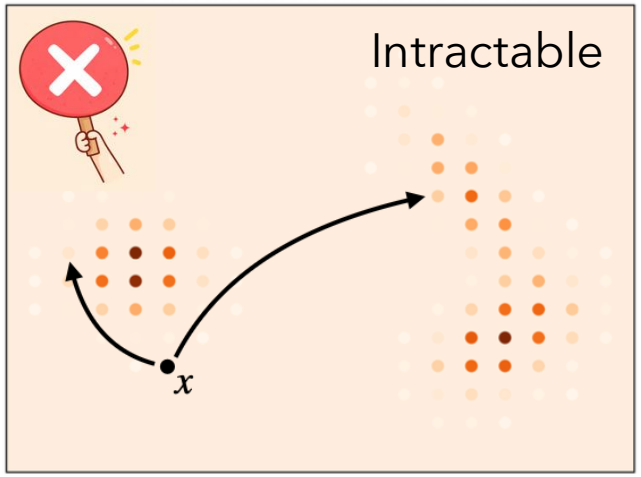


(Approximated by a neural network)

In Practice: Factorized Velocities

- Generate 1000 tokens: $|\mathcal{S}| \approx 50000; \mathcal{D} \approx 1000$

$$u_t(\cdot, x) \in \mathbb{R}^{|\mathcal{S}|^{\mathcal{D}}}$$

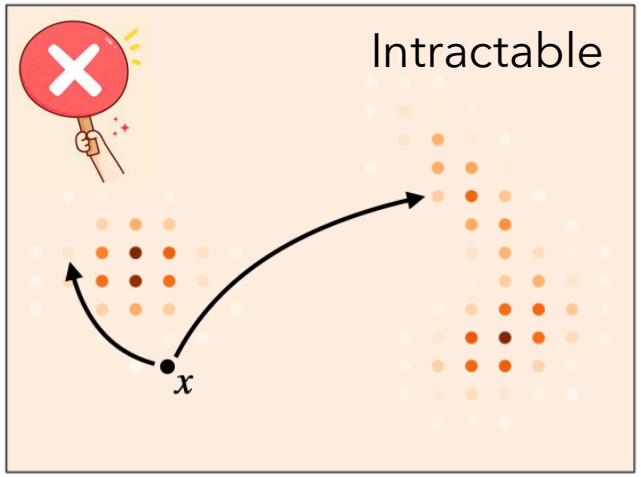


Campbell, A. et al. "Generative flows on discrete state-spaces: Enabling multimodal flows with applications to protein co-design. *ICML 2024*.

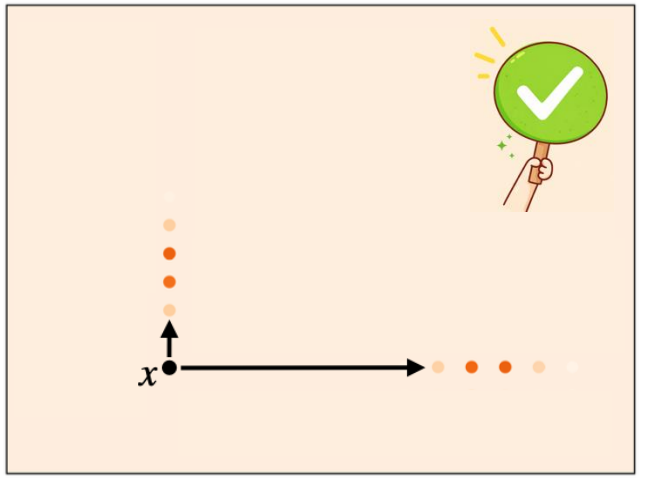
In Practice: Factorized Velocities

- Generate 1000 tokens: $|\mathcal{S}| \approx 50000; \mathcal{D} \approx 1000$

$u_t(\cdot, x) \in \mathbb{R}^{|\mathcal{S}|^{\mathcal{D}}}$



$u_t(\cdot, x) \in \mathbb{R}^{|\mathcal{S}| \cdot \mathcal{D}}$

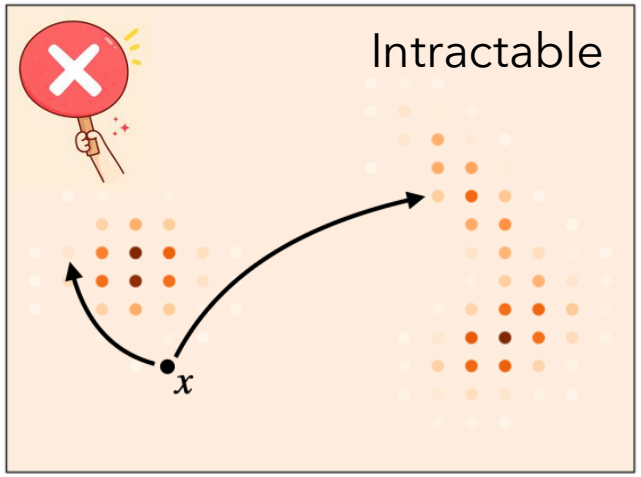


Campbell, A. et al. "Generative flows on discrete state-spaces: Enabling multimodal flows with applications to protein co-design. *ICML 2024*.

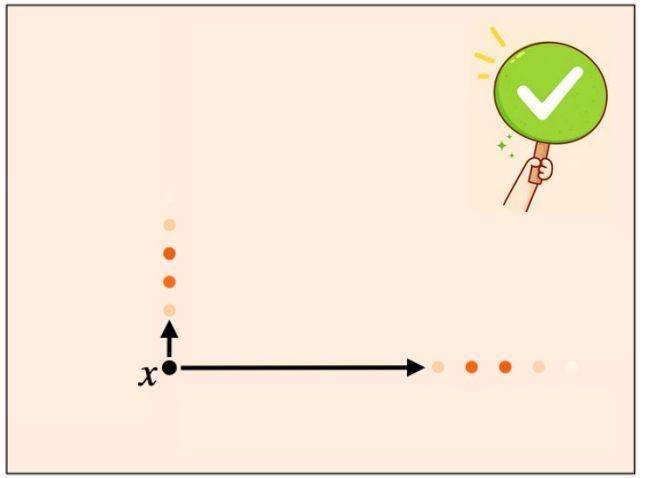
In Practice: Factorized Velocities

- Generate 1000 tokens: $|\mathcal{S}| \approx 50000; \mathcal{D} \approx 1000$

$u_t(\cdot, x) \in \mathbb{R}^{|\mathcal{S}|^{\mathcal{D}}}$



$u_t(\cdot, x) \in \mathbb{R}^{|\mathcal{S}| \cdot \mathcal{D}}$



$$q_{t|1}(x|\mathbf{x}_1) = \prod_{d=1}^{\mathcal{D}} q_{t|1}^d(x^d|x_1^d), \text{ and } u_t^q(z, x|\mathbf{x}_1) = \sum_{d=1}^{\mathcal{D}} \delta_{x^{\backslash d}}(z^{\backslash d}) u_t^{q,d}(z^d, x^d|x_1^d)$$

***d*-th coordinate**

Campbell, A. et al. "Generative flows on discrete state-spaces: Enabling multimodal flows with applications to protein co-design. *ICML 2024*.

Main results

$$q_{1|t}(z|x) = \frac{r(z)}{\mathbb{E}_{\mathbf{x}_1 \sim p_{1|t}(\mathbf{x}_1|x)}[r(\mathbf{x}_1)]} p_{1|t}(z|x)$$

Informal Theorem

Suppose that the conditional probability path of the **source** distribution $p_{t|1}$ and the **target** distribution $q_{t|1}$ are the same for any t. Then we have

$$q_{1|t}(z^d|x) = \frac{\mathbb{E}_{\mathbf{x}_1^{\setminus d} \sim p(\mathbf{x}_1^{\setminus d}|\mathbf{x}_1^d = z^d, \mathbf{x}_t = x)}[r(\mathbf{x}_1)]}{\mathbb{E}_{\mathbf{x}_1 \sim p_{1|t}(\mathbf{x}_1|x)}[r(\mathbf{x}_1)]} p_{1|t}(z^d|x).$$

Main results

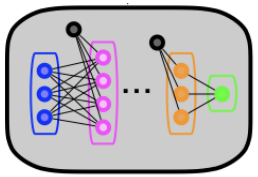
$$q_{1|t}(z|x) = \frac{r(z)}{\mathbb{E}_{\mathbf{x}_1 \sim p_{1|t}(\mathbf{x}_1|x)}[r(\mathbf{x}_1)]} p_{1|t}(z|x)$$

Informal Theorem

Suppose that the conditional probability path of the **source** distribution $p_{t|1}$ and the **target** distribution $q_{t|1}$ are the same for any t. Then we have

$$q_{1|t}(z^d|x) = \frac{\mathbb{E}_{\mathbf{x}_1^{\setminus d} \sim p(\mathbf{x}_1^{\setminus d}|\mathbf{x}_1^d = z^d, \mathbf{x}_t = x)}[r(\mathbf{x}_1)]}{\mathbb{E}_{\mathbf{x}_1 \sim p_{1|t}(\mathbf{x}_1|x)}[r(\mathbf{x}_1)]} p_{1|t}(z^d|x).$$

- Train a **guidance network** to approximate the guidance term



$$h_t^{d,\theta}(x_1^d, x_t) \approx \mathbb{E}_{x_1^{\setminus d} \sim p(x_1^{\setminus d}|x_1^d = z^d, x_t = x)}[r(x_1)]$$

Main results

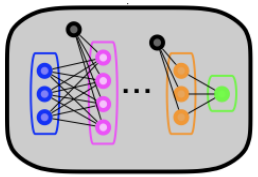
$$q_{1|t}(z|x) = \frac{r(z)}{\mathbb{E}_{\mathbf{x}_1 \sim p_{1|t}(\mathbf{x}_1|x)}[r(\mathbf{x}_1)]} p_{1|t}(z|x)$$

Informal Theorem

Suppose that the conditional probability path of the **source** distribution $p_{t|1}$ and the **target** distribution $q_{t|1}$ are the same for any t. Then we have

$$q_{1|t}(z^d|x) = \frac{\mathbb{E}_{\mathbf{x}_1^{\setminus d} \sim p(\mathbf{x}_1^{\setminus d}|\mathbf{x}_1^d = z^d, \mathbf{x}_t = x)}[r(\mathbf{x}_1)]}{\mathbb{E}_{\mathbf{x}_1 \sim p_{1|t}(\mathbf{x}_1|x)}[r(\mathbf{x}_1)]} p_{1|t}(z^d|x).$$

- Train a **guidance network** to approximate the guidance term



$$h_t^{d,\theta}(x_1^d, x_t) \approx \mathbb{E}_{x_1^{\setminus d} \sim p(x_1^{\setminus d}|x_1^d = z^d, x_t = x)}[r(x_1)]$$

$$\mathbf{h}_t^\theta(\mathbf{x}_1, \mathbf{x}_t) = (h_t^{1,\theta}(\mathbf{x}_1^1, \mathbf{x}_t), \ldots, h_t^{\mathcal{D},\theta}(\mathbf{x}_1^{\mathcal{D}}, \mathbf{x}_t))^\top$$

$$\mathbb{E}_{t \sim \mathcal{U}([0,1]), \mathbf{x}_1 \sim p_1(\mathbf{x}_1), \mathbf{x}_t \sim p_{t|1}(\mathbf{x}_t|\mathbf{x}_1)} \left[D_F\left(\mathbf{r}(\mathbf{x}_1) \middle\| \mathbf{h}_t^\theta(\mathbf{x}_1, \mathbf{x}_t)\right) \right]$$

Bregman Divergence

Sampling Algorithm

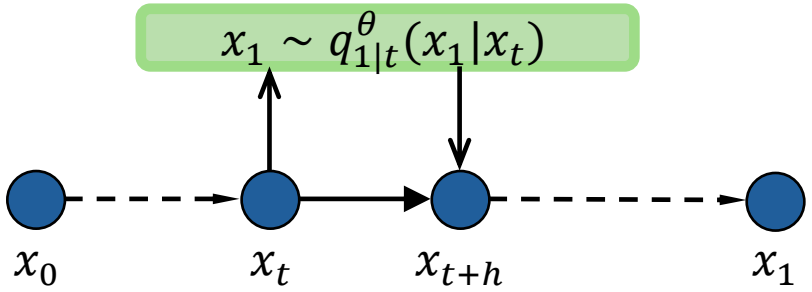
Require: pretrained posterior $p_{1|t}$, conditional transition rate $u_t^{p,d}(z, x|x_1)$, initial value x_0, posterior-based guidance H_t^θ, step size h

$t \leftarrow 0$
$\mathbf{x}_t \leftarrow x_0$
while $t + h < 1$ **do**
  for $d = 1, \dots, \mathcal{D}$ **do** ▷ in parallel
    Calculate the guided posterior $q_{1|t}^{d,\theta}(x_1^d|\mathbf{x}_t) \propto H_t^\theta(\mathbf{x}_t)_{d,x_1^d} p_{1|t}^d(x_1^d|\mathbf{x}_t)$
    Sample $\mathbf{x}_1^d \sim q_{1|t}^{d,\theta}(\cdot \mid \mathbf{x}_t)$
    $\lambda^d \leftarrow \sum_{s \neq \mathbf{x}_t^d} u_t^{p,d}(s, \mathbf{x}_t^d|\mathbf{x}_1^d)$
    Sample $Z_{\text{jump}}^d \sim U[0, 1]$
    if $Z_{\text{jump}}^d \leq 1 - e^{-h\lambda^d}$ **then**
      Sample $\mathbf{x}_t^d \sim \frac{u_t^{p,d}(\cdot, \mathbf{x}_t^d|\mathbf{x}_1^d)}{\lambda^d} \left(1 - \delta_{\mathbf{x}_t^d}(\cdot)\right)$
    end if
  end for
  $t \leftarrow t + h$
end while
$t \leftarrow t - h$
for $d = 1, \dots, \mathcal{D}$ **do** ▷ in parallel
  Calculate the guided posterior $q_{1|t}^{d,\theta}(x_1^d|\mathbf{x}_t) \propto H_t^\theta(\mathbf{x}_t)_{d,x_1^d} p_{1|t}^d(x_1^d|\mathbf{x}_t)$
  Sample $\mathbf{x}_1^d \sim q_{1|t}^{d,\theta}(\cdot \mid \mathbf{x}_t)$
end for
return $\mathbf{x}_1$

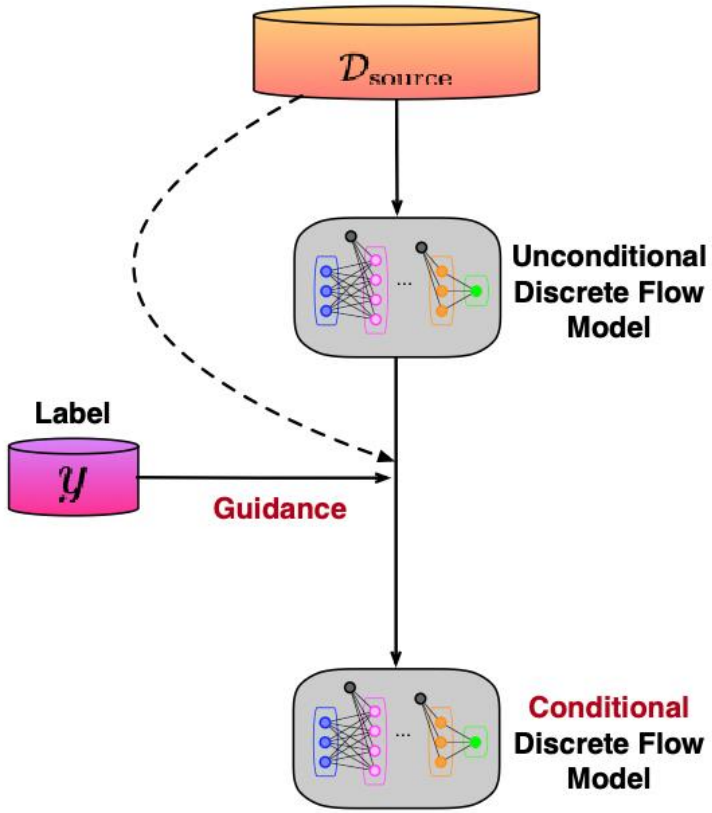
Sampling Algorithm

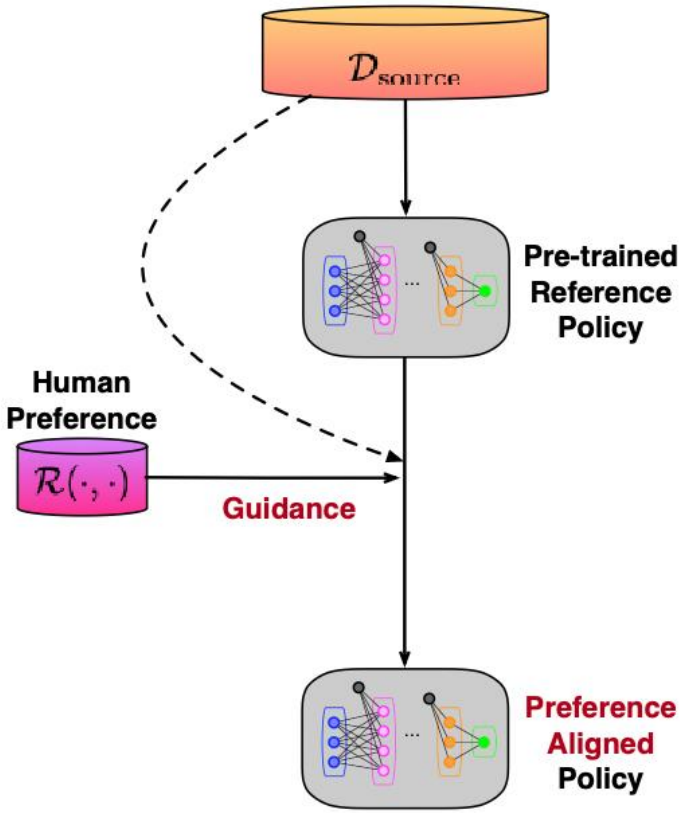
Require: pretrained posterior $p_{1|t}$, conditional transition rate $u_t^{p,d}(z, x|x_1)$, initial value x_0, posterior-based guidance H_t^θ, step size h

$t \leftarrow 0$
$\mathbf{x}_t \leftarrow x_0$
while $t + h < 1$ **do**
  for $d = 1, \dots, \mathcal{D}$ **do** ▷ in parallel
    Calculate the guided posterior $q_{1|t}^{d,\theta}(x_1^d|\mathbf{x}_t) \propto H_t^\theta(\mathbf{x}_t)_{d,x_1^d} p_{1|t}^d(x_1^d|\mathbf{x}_t)$
    Sample $\mathbf{x}_1^d \sim q_{1|t}^{d,\theta}(\cdot \mid \mathbf{x}_t)$
    $\lambda^d \leftarrow \sum_{s \neq \mathbf{x}_t^d} u_t^{p,d}(s, \mathbf{x}_t^d|\mathbf{x}_1^d)$
    Sample $Z_{\text{jump}}^d \sim U[0, 1]$
    if $Z_{\text{jump}}^d \leq 1 - e^{-h\lambda^d}$ **then**
      Sample $\mathbf{x}_t^d \sim \frac{u_t^{p,d}(\cdot, \mathbf{x}_t^d|\mathbf{x}_1^d)}{\lambda^d} \left(1 - \delta_{\mathbf{x}_t^d}(\cdot)\right)$
    end if
  end for
  $t \leftarrow t + h$
end while
$t \leftarrow t - h$
for $d = 1, \dots, \mathcal{D}$ **do** ▷ in parallel
  Calculate the guided posterior $q_{1|t}^{d,\theta}(x_1^d|\mathbf{x}_t) \propto H_t^\theta(\mathbf{x}_t)_{d,x_1^d} p_{1|t}^d(x_1^d|\mathbf{x}_t)$
  Sample $\mathbf{x}_1^d \sim q_{1|t}^{d,\theta}(\cdot \mid \mathbf{x}_t)$
end for
return $\mathbf{x}_1$



Special Cases



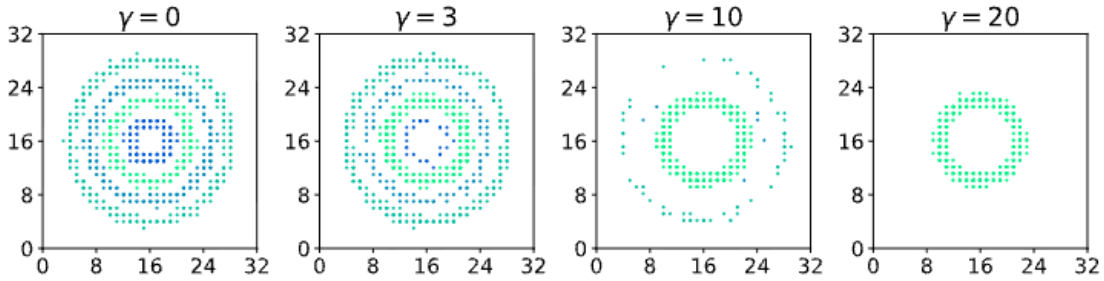


$$q_{1|t}(z^d|x) = p_{1|t}(z^d|x,y) = \frac{p(y|\mathbf{x}_1^d = z^d, \mathbf{x}_t = x)}{p(y|\mathbf{x}_t = x)} p_{1|t}(z^d|x).$$

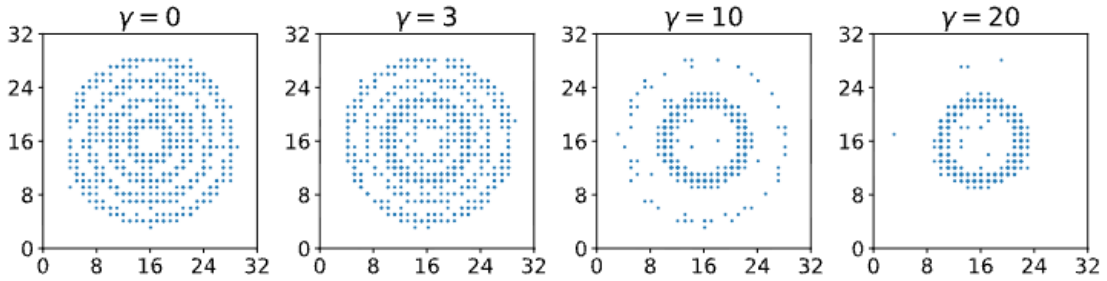
$$\pi^*(\mathbf{o}_1|\mathbf{c}) = \frac{1}{\mathcal{Z}(\mathbf{c})} \pi_{ref}(\mathbf{o}_1|\mathbf{c}) \exp\left(\frac{\mathcal{R}(\mathbf{c},\mathbf{o}_1)}{\tau}\right)$$

Simulation

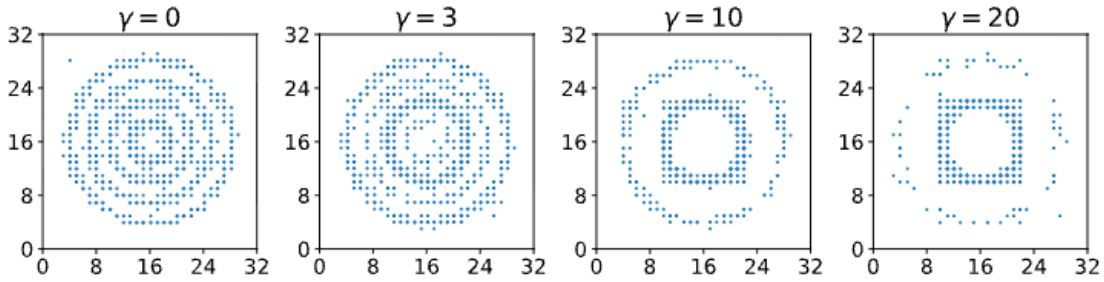
- Sample space $\mathcal{S}^{\mathcal{D}} = \{0, 1, \ldots, 32\}^2$
- Source distribution p_1; Target distribution $p_1^{(\gamma)}(x) \propto p_1(x)e^{-\gamma\mathcal{E}(x)}$



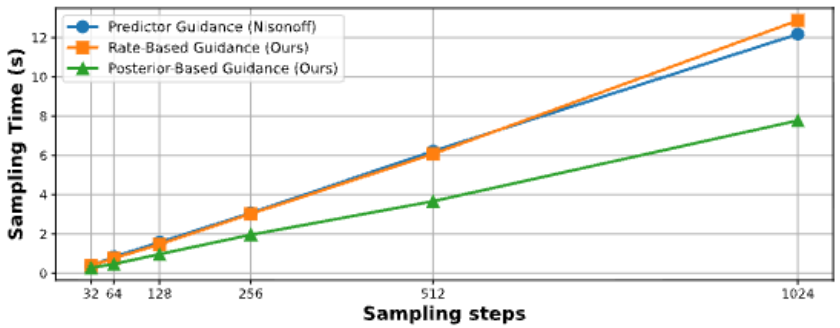
(a) **Ground Truth (rings)**



(c) Posterior-Based (Uniform, Ours)



(b) Rate-Based (Masked, Nisonoff et al. (2025))



(d) Sampling Time Comparison

Real data

- Visual Generation Performance on the GenEval Benchmark.

A beautiful modern wooden house …

Type	Method	Single Obj.	Two Obj.	Counting	Colors	Position	Color Attri.	Overall ↑
Gen. Only	LlamaGen [72]	0.71	0.34	0.21	0.58	0.07	0.04	0.32
	Emu3-Gen [79]	0.98	0.71	0.34	0.81	0.17	0.21	0.54
	LDM [62]	0.92	0.29	0.23	0.70	0.02	0.05	0.37
	SDv1.5 [62]	0.97	0.38	0.35	0.76	0.04	0.06	0.43
	PixArt-alpha [11]	0.98	0.50	0.44	0.80	0.08	0.07	0.48
	SDv2.1 [62]	0.98	0.51	0.44	0.85	0.07	0.17	0.50
	DALL-E 2 [61]	0.94	0.66	0.49	0.77	0.10	0.19	0.52
	SDXL [57]	0.98	0.74	0.39	0.85	0.15	0.23	0.55
	DALL-E 3 [7]	0.96	0.87	0.47	0.83	0.43	0.45	0.67
	SD3-Medium [18]	0.99	0.94	0.72	0.89	0.33	0.60	0.74
Und. + Gen.	SEED-† [23]	0.97	0.58	0.26	0.80	0.19	0.14	0.49
	LWM [41]	0.93	0.41	0.46	0.79	0.09	0.15	0.47
	ILLUME [77]	0.99	0.86	0.45	0.71	0.39	0.28	0.61
	TokenFlow-XL [59]	0.95	0.60	0.41	0.81	0.16	0.24	0.55
	Chameleon [74]	–	–	–	–	–	–	0.39
	Janus [80]	0.97	0.68	0.30	0.84	0.46	0.42	0.61
	Janus-Pro-1B [12]	0.98	0.82	0.51	0.89	0.65	0.56	0.73
	Show-o [82]	0.95	0.52	0.49	0.82	0.11	0.28	0.53
	Transfusion [89]	–	–	–	–	–	–	0.63
	UniDisc [73]	0.92	0.47	0.15	0.67	0.13	0.19	0.42
	D-DiT [39]	0.97	0.80	0.54	0.76	0.32	0.50	0.65
	FUDOKI [78]	0.96	0.85	0.56	0.88	0.68	0.67	0.77
	Ours	0.94	0.86	0.53	0.89	0.70	0.77	0.78

Note: "Und." = Understanding, "Gen." = Generation. † = models integrating an external pretrained diffusion model. Values exceeding the baseline FUDOKI are highlighted in gray.

Real data

Q: Is the phone number in…?

- Multimodal Understanding Performance on Various Benchmarks.

Type	Model	# Params	POPE ↑	MME-P ↑	MMB ↑	GQA ↑	MMMU ↑	MM-Vet ↑
Und. Only	LLaVA-Phi-1.5 [43]	1.3B	84.1	1128.0	-	56.5	30.7	-
	MobileVLM [13]	1.4B	84.5	1196.2	53.2	56.1	-	-
	MobileVLM-V2 [14]	1.4B	84.3	1302.8	57.7	59.3	-	-
	MobileVLM [13]	2.7B	84.9	1288.9	59.6	59.0	-	-
	MobileVLM-V2 [14]	2.7B	84.7	1440.5	63.2	61.1	-	-
	LLaVA-Phi [91]	2.7B	85.0	1335.1	59.8	-	-	28.9
	LLaVA [43]	7B	76.3	809.6	38.7	-	-	25.5
	LLaVA-v1.5 [42]	7B	85.9	1510.7	64.3	62.0	35.4	31.1
	InstructBLIP [15]	7B	-	-	36.0	49.2	-	26.2
	Qwen-VL-Chat [4]	7B	-	1487.5	60.6	57.5	-	-
	IDEFICS [35]	8B	-	-	48.2	38.4	-	-
	Emu3-Chat [79]	8B	85.2	1244.0	58.5	60.3	31.6	37.2
	InstructBLIP [15]	13B	78.9	1212.8	-	49.5	-	25.6
Und. & Gen.	LaVIT† [32]	7B	-	-	-	46.8	-	-
	MetaMorph† [75]	8B	-	-	75.2	-	-	-
	Gemini-Nano-1 [19]	1.8B	-	-	-	-	26.3	-
	ILLUME [77]	7B	88.5	1445.3	65.1	-	38.2	37.0
	TokenFlow-XL [59]	13B	86.8	1545.9	68.9	62.7	38.7	40.7
	LWM [41]	7B	75.2	-	-	44.8	-	9.6
	VILA-U [81]	7B	85.8	1401.8	-	60.8	-	33.5
	Chameleon [74]	7B	-	-	-	-	22.4	8.3
	Janus [80]	1.5B	87.0	1338.0	69.4	59.1	30.5	34.3
	Janus-Pro-1B [12]	1.5B	86.2	1444.0	75.5	59.3	36.3	39.8
	Show-o-256 [82]	1.3B	73.8	948.4	-	48.7	25.1	-
	Show-o-512 [82]	1.3B	80.0	1097.2	-	58.0	26.7	-
	D-Dit [39]	2.0B	84.0	1124.7	-	59.2	-	-
	FUDOKI [78]	1.5B	86.1	1485.4	73.9	57.6	34.3	38.0
	Ours	1.5B	86.8	1492.7	74.2	58.2	35.4	38.6

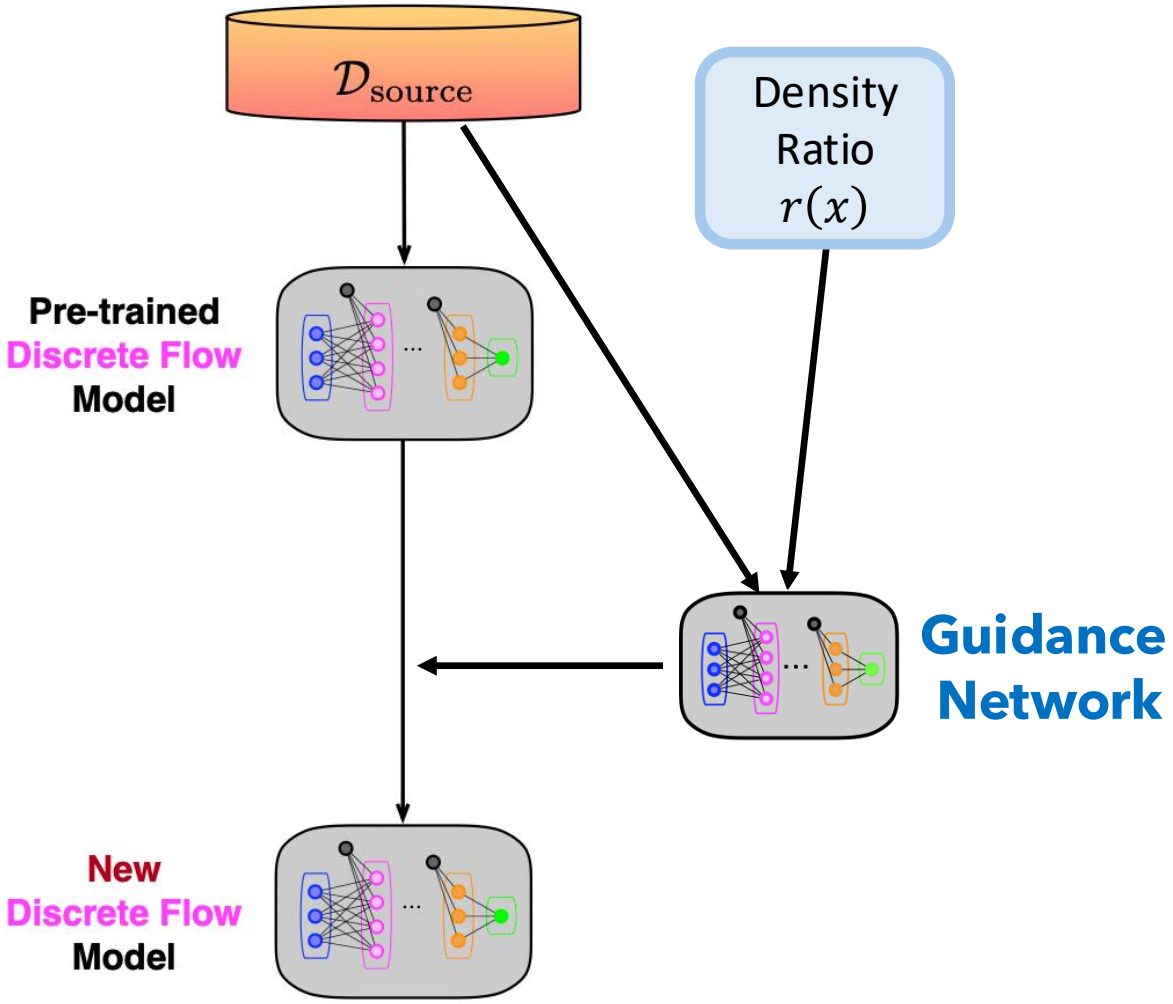
Note: "Und." = Understanding, "Gen." = Generation. † = models integrating an external pretrained diffusion model. Values exceeding the baseline FUDOKI are highlighted in gray.

Comparison to literature

- Discrete Diffusion / Flow Models
  - Discrete diffusion models [Austin et al, NeurIPS 2021] [Campbell et al, NeurIPS 2022] [Sun et al, ICLR 2023] [Lou et al, ICML 2024] etc.
  - Discrete flow matching / CTMC models [Gat et al, NeurIPS 2024] [Campbell et al, ICML 2024] [Shaul et al, ICLR 2025] etc.

Guidance Method	**Exact Guidance**	# of Function Calls
Posterior-Based (Ours)	✓	1
Rate-Based [Nisonoff et al, ICLR 2025]	✓	Absorbing: $\mathcal{D}+1$ Uniform: $\mathcal{D} \times (\lvert\mathcal{S}\rvert - 1) + 1$
First-Order Approximated [Schiff et al, ICLR 2025] [Vignac et al, ICLR 2023] etc.	✗	2

Summary



- Derive the exact guidance for discrete flow-based models, steering a pre-trained generation process toward the target distribution
- Efficient to compute: a single forward pass in each sampling step
- Unified framework: with many existing guidance methods as special cases.

Thanks!

liyanxie@umn.edu

Zhengyan Wang, Yidong Ouyang, X., Fang Fang, Hongyuan Zha, and Guang Cheng. "Discrete guidance matching: Exact guidance for discrete flow matching." *ICLR 2026, also on arXiv:2509.21912.*