NEURAL INFORMATION PROCESSING SYSTEMS

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

NeurIPS 2025 | Spotlight

Towards a Golden Classifier-Free Guidance Path via Foresight Fixed Point Iterations

Kaibo Wang[1], Jianda Mao[1], Tong Wu[1], Yang Xiang[1,2]

1. Department of Mathematics, The Hong Kong University of Science and Technology
2. HKUST Shenzhen-Hong Kong Collaborative Innovation Research Institute

Better text-to-image guidance

Text prompt
e.g. A fierce dragon → Diffusion model + Guidance →




Insight 1: Existing guidance is one-step fixed-point iteration

Better text-to-image guidance

Text prompt
e.g. A fierce dragon → Diffusion model + Guidance →

Insight 2: We upgrade it to foresight fixed-point iterations with more steps.

Background: conditional generation

香港科技大學 THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

- Unconditional diffusion:

$$x_T \sim \mathcal{N}(0, I) \rightarrow \epsilon_\theta(x_T, t = T) \rightarrow x_{T-1} \rightarrow \cdots \rightarrow \epsilon_\theta(x_1, t = 1) \rightarrow x_0$$

- Conditional diffusion (adds condition c as an input):

$$x_T \sim \mathcal{N}(0, I) \rightarrow \epsilon_\theta(x_T, c, t = T) \rightarrow x_{T-1} \rightarrow \cdots \rightarrow \epsilon_\theta(x_1, c, t = 1) \rightarrow x_0$$

- Training: $(x, c) \sim D, \min_\theta \left\| \epsilon_\theta\left(\boldsymbol{x_t} \coloneqq \sqrt{\boldsymbol{\bar{\alpha}_t}} \boldsymbol{x} + \sqrt{\boldsymbol{1 - \bar{\alpha}_t}} \boldsymbol{\epsilon},\ t,\ \boldsymbol{c} \textbf{ with } \boldsymbol{prob}\ \boldsymbol{p} \textbf{ else } \emptyset \right) - \epsilon \right\|_2^2$

Due to the sparsity of the conditional data, the model tends to **underfit**.

Background: classifier-free guidance

- Classifier-free guidance (CFG):

$$\epsilon_{CFG}^{\gamma} = \epsilon(x_t, t, \emptyset) + \gamma[\epsilon(x_t, t, c) - \epsilon(x_t, t, \emptyset)] \qquad \gamma > 1$$

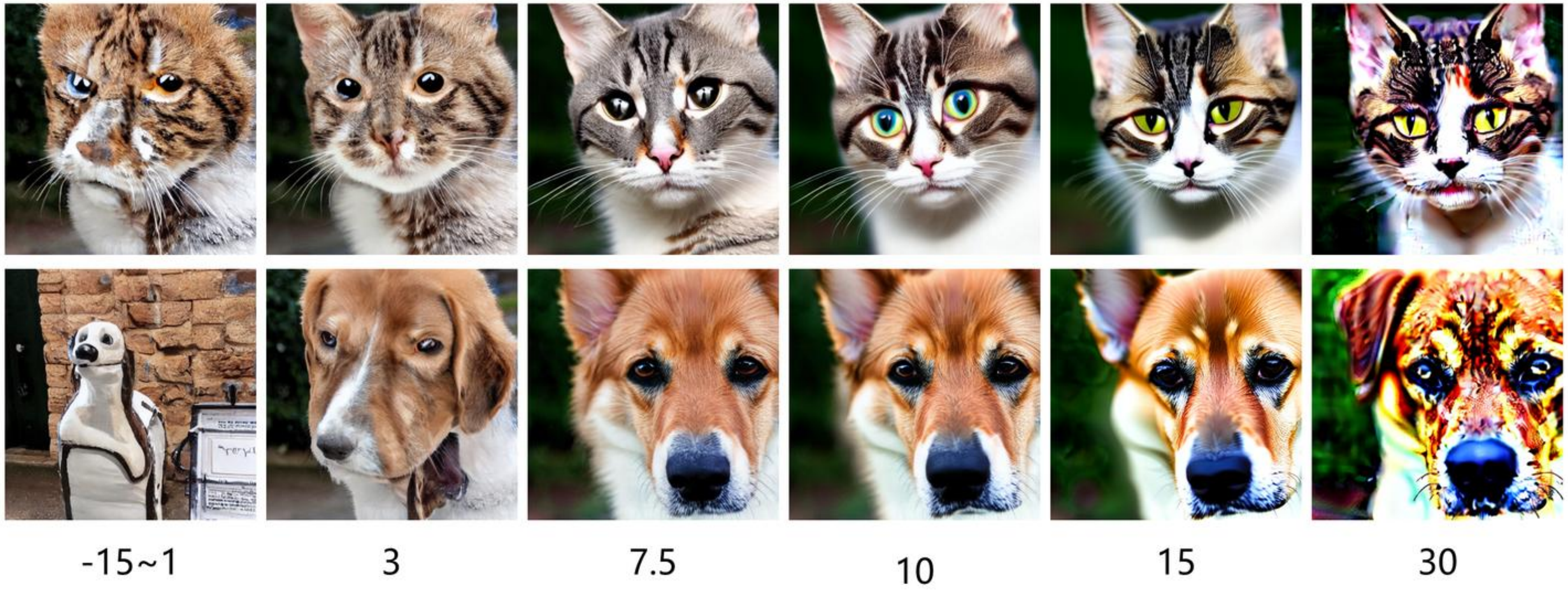
Class: mushroom **cheeseburger**

Naïve conditional samples

CFG Samples

CFG amplifies the difference between conditional and unconditional outputs

Background: Cons of CFG



CFG improves quality, but reduces diversity and oversaturate with increasing γ

Better CFG: given prediction $\epsilon^{u/c}(x_t)$, what's better update $x_{t-1} = h\big(x_t, \epsilon^u(x_t), \epsilon^c(x_t)\big)$?

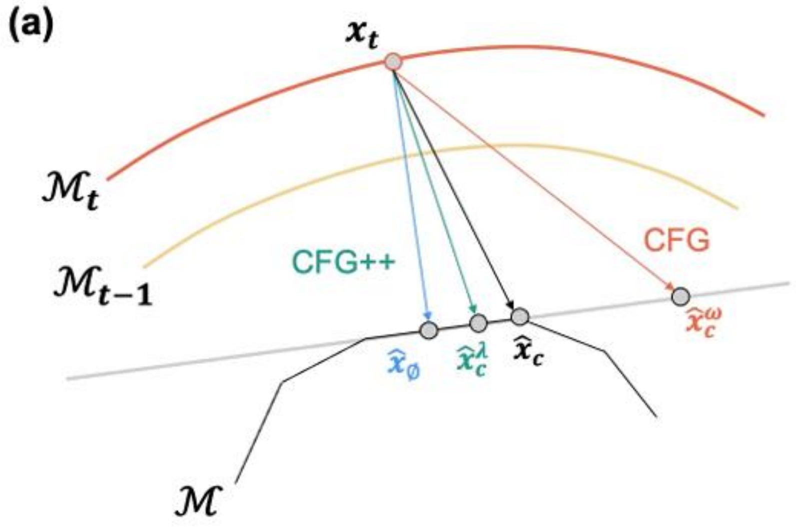
Related works: improving CFG

香港科技大學 THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

- CFG++ 2406.08070

Target: Off-manifold phenomenon of CFG

Route: inverse problem sampling



Algorithm 1 Reverse Diffusion with CFG

Require: $\boldsymbol{x}_T \sim \mathcal{N}(0, \mathbf{I}_d), 0 \leq \omega \in \mathbb{R}$

1: **for** $i = T$ **to** 1 **do**

2: $\quad \hat{\boldsymbol{\epsilon}}_{\boldsymbol{c}}^{\omega}(\boldsymbol{x}_t) = \hat{\boldsymbol{\epsilon}}_{\varnothing}(\boldsymbol{x}_t) + \omega[\hat{\boldsymbol{\epsilon}}_{\boldsymbol{c}}(\boldsymbol{x}_t) - \hat{\boldsymbol{\epsilon}}_{\varnothing}(\boldsymbol{x}_t)]$

3: $\quad \hat{\boldsymbol{x}}_{\boldsymbol{c}}^{\omega}(\boldsymbol{x}_t) \leftarrow (\boldsymbol{x}_t - \sqrt{1-\bar{\alpha}_t}\hat{\boldsymbol{\epsilon}}_{\boldsymbol{c}}^{\omega}(\boldsymbol{x}_t))/\sqrt{\bar{\alpha}_t}$

4: $\quad \boldsymbol{x}_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\hat{\boldsymbol{x}}_{\boldsymbol{c}}^{\omega}(\boldsymbol{x}_t) + \sqrt{1-\bar{\alpha}_{t-1}}\hat{\boldsymbol{\epsilon}}_{\boldsymbol{c}}^{\omega}(\boldsymbol{x}_t)$

5: **end for**

6: **return** $\boldsymbol{x}_0$

Algorithm 2 Reverse Diffusion with CFG++

Require: $\boldsymbol{x}_T \sim \mathcal{N}(0, \mathbf{I}_d), \lambda \in [0, 1]$

1: **for** $i = T$ **to** 1 **do**

2: $\quad \hat{\boldsymbol{\epsilon}}_{\boldsymbol{c}}^{\lambda}(\boldsymbol{x}_t) = \hat{\boldsymbol{\epsilon}}_{\varnothing}(\boldsymbol{x}_t) + \lambda[\hat{\boldsymbol{\epsilon}}_{\boldsymbol{c}}(\boldsymbol{x}_t) - \hat{\boldsymbol{\epsilon}}_{\varnothing}(\boldsymbol{x}_t)]$

3: $\quad \hat{\boldsymbol{x}}_{\boldsymbol{c}}^{\lambda}(\boldsymbol{x}_t) \leftarrow (\boldsymbol{x}_t - \sqrt{1-\bar{\alpha}_t}\hat{\boldsymbol{\epsilon}}_{\boldsymbol{c}}^{\lambda}(\boldsymbol{x}_t))/\sqrt{\bar{\alpha}_t}$

4: $\quad \boldsymbol{x}_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\hat{\boldsymbol{x}}_{\boldsymbol{c}}^{\lambda}(\boldsymbol{x}_t) + \sqrt{1-\bar{\alpha}_{t-1}}\hat{\boldsymbol{\epsilon}}_{\varnothing}(\boldsymbol{x}_t)$

5: **end for**

6: **return** $\boldsymbol{x}_0$

Related works: improving CFG

➢ Z-Sampling [2412.10891]

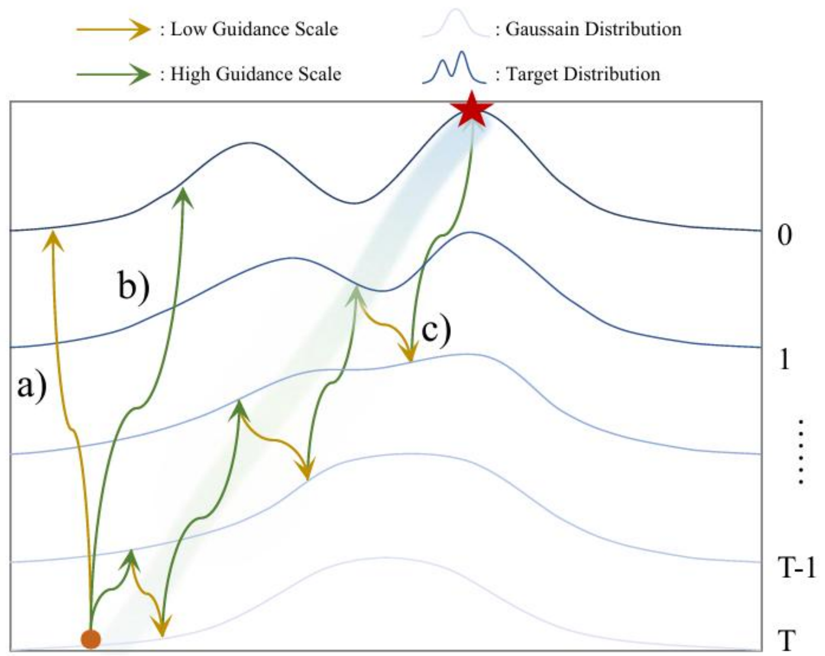
Target: Semantic injection in CFG is insufficient.

Route: Backward-forward in each step x_t

1. ODE backward to x_{t+1} with $\epsilon^{\emptyset}$
2. ODE forward to $\hat{x}_t$ with ϵ^{γ} (CFG with γ)

➢ Resample [2201.09865]

Diff: Backward with noising function



Guidance design

Different methods become a **tightly coupled package**

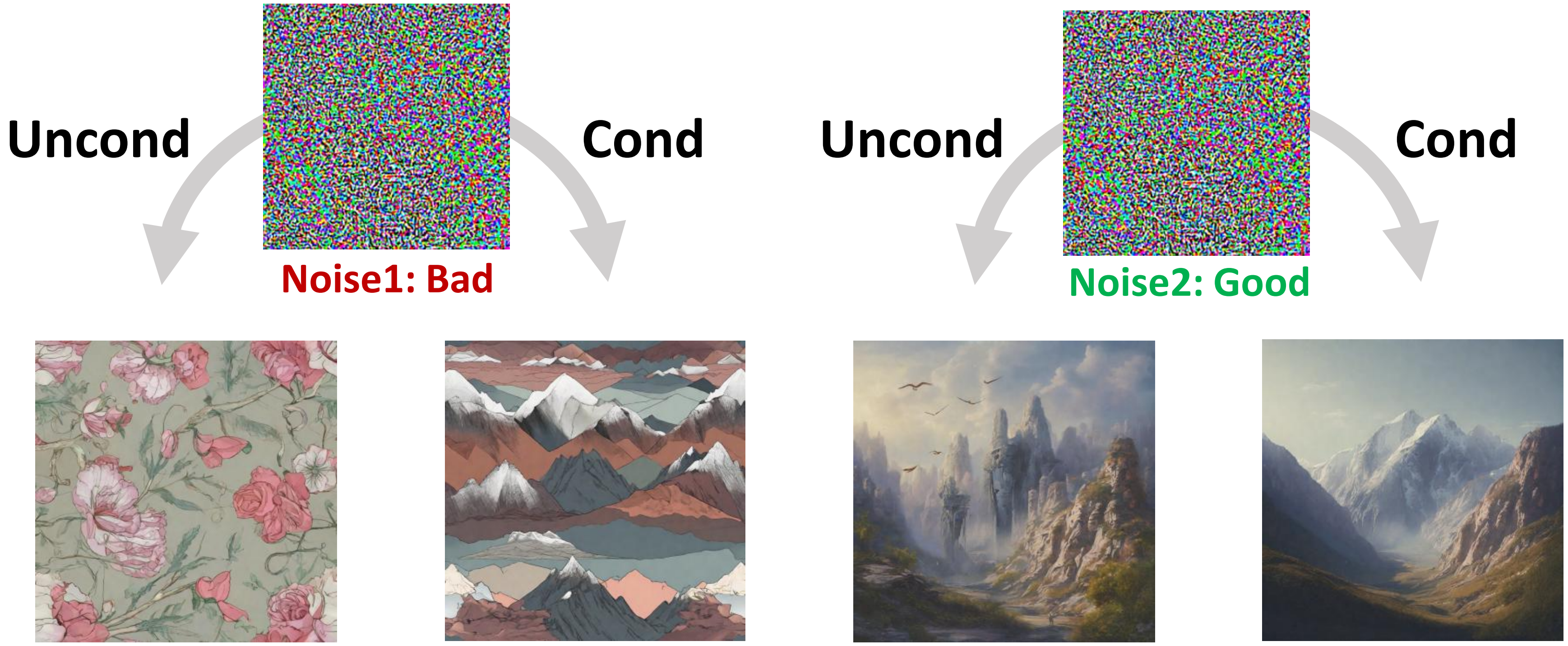
Z-sampling for better alignment

CFG++ avoids oversaturate

Resample for stochasticity

How can we systematically explore the design space for guidance and develop more effective algorithms from a unified framework?

Golden path hypothesis



Key Observation: If $\hat{x}_t$ unconditionally generates image of mountains $f^u_{t\to 0}(\hat{x}_t)$, it performs better on mountain-related prompts than other x_t.

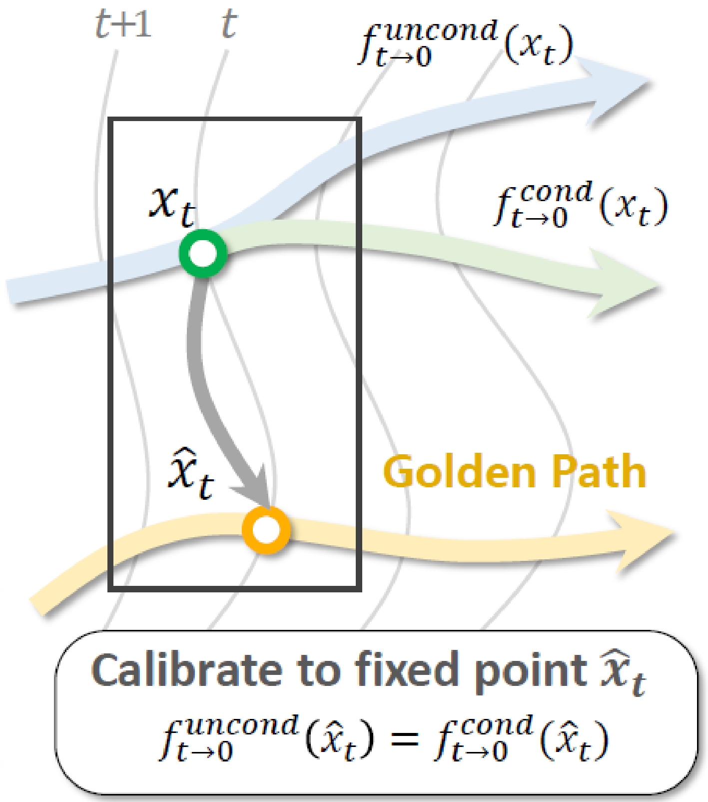
Golden path via fixed point iterations

- **Problem:** Given t and initial value x_t, find $\hat{x}_t$ such that $f^u_{t\to 0}(\hat{x}_t) = f^c_{t\to 0}(\hat{x}_t)$

  $f^{u/c}_{t\to 0}(\hat{x}_t)$ denotes denotes (un)conditional denoising function from x_t to x_0 using DDIM (ODE)
- **Intuition:** When cond. and uncond. produce different outputs, the model requires a sharp turn.
- **Solver:** Fixed point iteration framework

$$\hat{x}_t = F^{(K)}(x_t) = F \circ \cdots \circ F(x_t) \qquad \triangleright \text{ Calibrate}$$
$$x_{t-1} = \text{Sampler}\left(\hat{x}_t, \epsilon^u(\hat{x}_t)\right) \qquad \triangleright \text{ Denoise}$$

Existing guidance can be unified into the framework



Design space of fixed point iterations

香港科技大學 THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

1. Consistency interval $f^u_{a\to b}(\hat{x}_t) = f^c_{a\to b}(\hat{x}_t)$ (Pursuing $t \to 0$ is too difficult)
2. Fixed point operator F, $F(\hat{x}_t) = \hat{x}_t \Rightarrow f^u_{a\to b}(\hat{x}_t) = f^c_{a\to b}(\hat{x}_t)$

   e.g., linear: $x_t = x_t + w\left(f^u_{a\to b}(x_t) - f^c_{a\to b}(x_t)\right)$, backward-forward: $x_t = f^{u-1}_{a\to b} f^c_{a\to b}(x_t)$
3. Guidance strength/scheduler e.g. $w\xi_t$
4. Number of iterations K

Methods	Fixed point operator $F(x_t)$	Interval	Iters.
CFG	$x_t - w\xi_t[\epsilon^c(x_t) - \epsilon^u(x_t)]$	$t \to t-1$	1
CFG++	$x_t - \lambda\tilde{\xi}_t[\epsilon^c(x_t) - \epsilon^u(x_t)]$	$t \to t-1$	1
Z-sampling	$(\mathrm{id} - w\xi_t\Delta\epsilon) \circ f^\gamma_{t+1\to t} \circ f^u_{t\to t+1}(x_t)$	$t+1 \to t-1$	1
Resampling	$(\mathrm{id} - w\xi_t\Delta\epsilon) \circ f^\gamma_{t+1\to t} \circ n_{t\to t+1}(x_t)$	$t+1 \to t-1$	N
FSG (ours)	$(\mathrm{id} - \lambda\tilde{\xi}_t\Delta\epsilon) \circ f^u_{t-\Delta t\to t} \circ f^\gamma_{t\to t-\Delta t}(x_t)$	$t \to t-\Delta t$	N

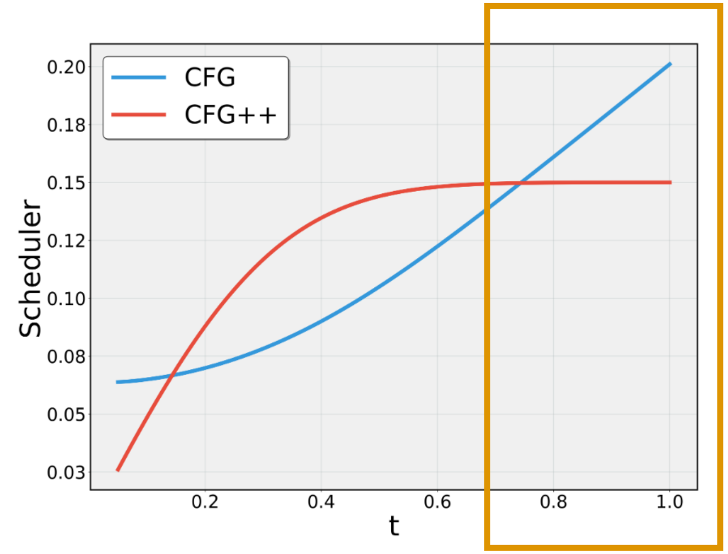
Comparison of design choices

- **Scheduler** (difference of CFG & CFG++):

During the critical early stages of generation, the scheduler of CFG++ provides more stable guidance

- **Fixed point operator:**

Forward-backward operators with longer intervals exhibit lower empirical contraction rates (ℓ_2 norm)



Operators	$t = 0.2$	0.4	0.6	0.8
id - $w_t\Delta\epsilon$ (CFG, CFG++)	1.00	1.00	1.00	1.00
$f^{\gamma}_{t+dt\to t} \circ f^{u}_{t\to t+dt}$ (Z-sampling)	1.04	0.99	0.97	0.99
$f^{\gamma}_{t+dt\to t} \circ n^{u}_{t\to t+dt}$ (Resample)	0.89	0.97	1.03	1.07
$f^{u}_{t/2\to t} \circ f^{\gamma}_{t\to t/2}$	1.03	0.95	0.96	0.98
$f^{u}_{t/4\to t} \circ f^{\gamma}_{t\to t/4}$	0.61	0.91	0.88	0.91
$f^{u}_{0\to t} \circ f^{\gamma}_{t\to 0}$	0.62	0.70	0.75	0.79

Foresight guidance

- We adopt scheduler of CFG++ and Forward-backward operator

Three modifications:

1. **More iterations**
2. Use **longer consistency interval**
   $f^{u}_{t\to t-\Delta t}(\hat{x}_t) = f^{c}_{t\to t-\Delta t}(\hat{x}_t)$
3. Prioritize fixed point iterations and longer intervals **in the early stages**

FSG will not add computational overhead.

Algorithm 1: Foresight Guidance (FSG) Sampling

Input : Initial noise x_T, Condition c, Timesteps T, Iteration set $\mathcal{S} = \{(t_i, K_i, \Delta t_i)\}_{i=1}^{M}$, Strengths γ, λ.
Output : Generated image x_0

for $t \leftarrow T$ **to** 1 **do**
  $\hat{x}_t = x_t$;
  if $(t, K, \Delta t) \in \mathcal{S}$ **then**
    Foresight Fixed Point Calibration;
    for $k \leftarrow 1$ **to** K **do**
      $x'_{t-\Delta t} = f^{\gamma}_{t\to t-\Delta t}(\hat{x}_t; c)$;
      $x'_t = f^{u}_{t-\Delta t\to t}(x'_{t-\Delta t})$;
    end
  end
  CFG++ Calibration;
  $\hat{x}_t = x_t - \lambda\tilde{\xi}_t(\epsilon^c(x_t) - \epsilon^u(x_t))$;
  Denoising Step;
  $x_{t-1} = \text{Sampler}(\hat{x}_t, \epsilon^u(\hat{x}_t))$;
end
return x_0

Subproblem decomposition

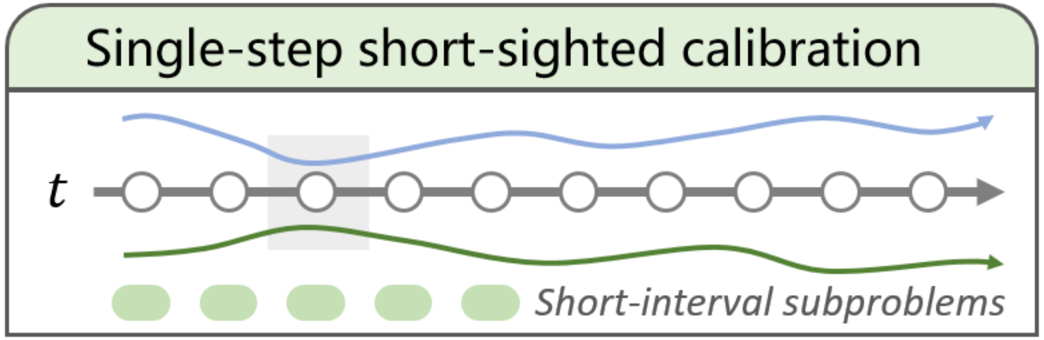
香港科技大學 THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

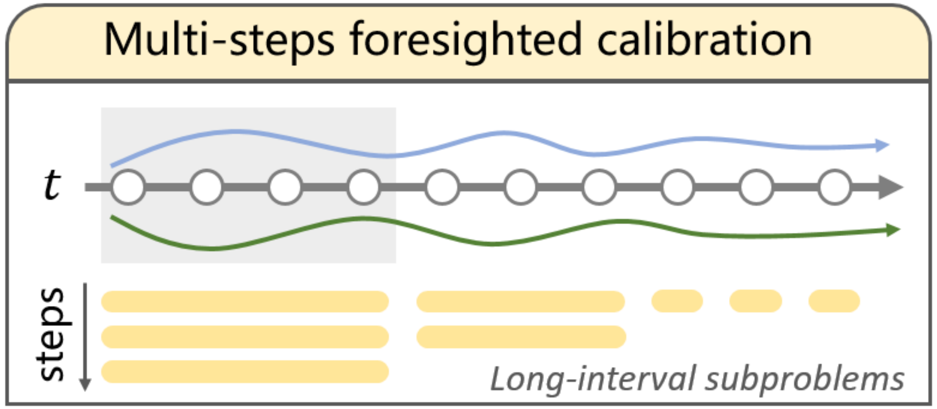
For the consistency of the entire trajectory, we can

1. Short interval subproblem: Solving the $f^c_{t \to t-1}(x_t) = f^u_{t \to t-1}(x_t)$ in a single iteration

2. Long interval subproblem: Solving the $f^c_{t \to t-\Delta t}(x_t) = f^u_{t \to t-\Delta t}(x_t)$ through multi-step iterations

Intuitively, multi-step iteration can impose stronger consistency at both ends of the entire trajectory.





Theoretical analysis

Theorem 1. *Given the total iteration budget N and timesteps T, uniformly divide $f^u_{T\to 0} = f^c_{T\to 0}$ into M subproblems ($f^u_{iT/M\to(i-1)T/M} = f^c_{iT/M\to(i-1)T/M}$, $i \in [M]$), each solved with N/M fixed point iterations. Let $\mathcal{L} = \frac{1}{T}\sum_{t=1}^{T}\|\epsilon^c(\hat{x}_t) - \epsilon^u(\hat{x}_t)\|_2^2$ denote the average gap over calibrated trajectories $\hat{x}_t \in \mathbb{R}^d$, with B as the Euclidean norm bound for $\hat{x}_t$ and $\epsilon^{c/u}(\hat{x}_t)$, L as the smoothness constant of $\epsilon^{c/u}(\cdot)$, and $r \in (0,1)$ as the upper bound of the contraction rate of $_i$, $i \in [M]$. Under mild assumptions , there exists a constant $C > 0$ such that*

$$\mathcal{L} = \frac{1}{T}\sum_{t=1}^{T}\|\epsilon^c(\hat{x}_t) - \epsilon^u(\hat{x}_t)\|_2^2 \leq B^2\left(Cr^{\frac{2N}{M}} + \frac{2L^2}{M^2}\right).$$

When the model is smoother, the strategy with **longer intervals and more iterations** proves superior

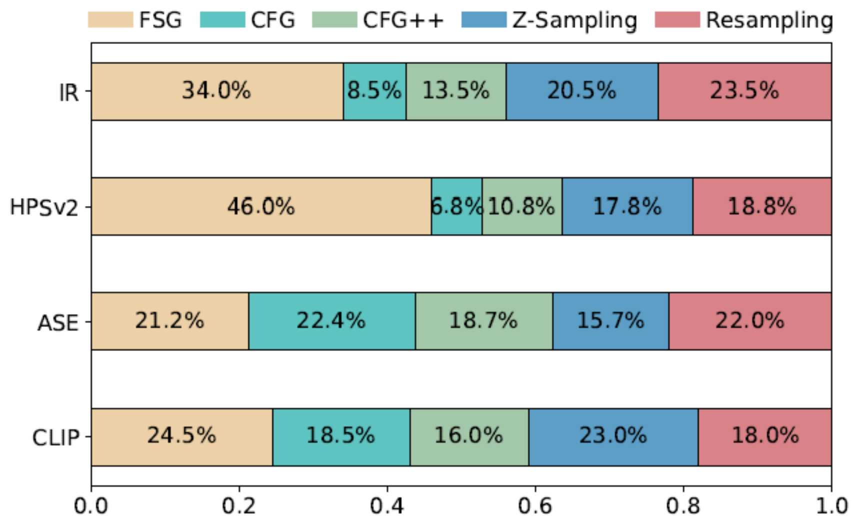
When resources are sufficient, tend towards strategy with **shorter intervals and single iteration**

Quantitative Analysis

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

FSG enhance aesthetics and alignment

CFG/CFG++ benefit from increased iterations (× 2 / 3)



(a) Top-1 Rate.

Datasets		DrawBench				Pick-a-Pic			
Method	NFE	IR↑	HPSv2↑	AES↑	CLIP↑	IR↑	HPSv2↑	AES↑	CLIP↑
CFG	50	59.02	28.73	6.07	32.29	82.14	28.46	**6.73**	33.53
CFG++	50	65.21	28.98	**6.08**	32.60	89.75	28.72	6.67	33.86
Z-Sampling	50	72.75	29.08	6.00	32.59	96.77	28.68	6.59	33.97
Resampling	50	59.99	28.80	5.99	32.21	82.65	28.46	6.61	33.46
FSG (ours)	50	**82.81**	**29.42**	6.01	**32.65**	**98.59**	**28.89**	6.60	**34.32**
CFG×2	100	77.71	29.36	**6.06**	32.44	96.06	28.84	6.64	34.13
CFG++×2	100	79.42	29.42	6.01	32.61	99.90	29.00	6.61	34.18
Z-Sampling	100	77.46	29.26	6.03	32.39	94.98	28.79	6.61	34.01
Resampling	100	77.26	29.12	6.00	32.46	79.36	28.61	6.02	33.61
FSG (ours)	100	**84.12**	**29.54**	6.02	**32.76**	**102.82**	**29.05**	**6.66**	**34.30**
CFG×3	150	83.56	**29.51**	5.95	32.66	102.13	29.04	6.61	**34.28**
CFG++×3	150	82.58	29.45	5.93	32.66	103.32	**29.05**	6.57	34.20
Z-Sampling	150	78.35	29.40	**6.06**	32.43	97.25	28.90	**6.67**	34.20
Resampling	150	79.98	29.23	6.05	32.32	87.48	28.70	6.59	33.49
FSG (ours)	150	**88.18**	29.44	5.96	**32.70**	**104.86**	29.04	6.65	**34.28**

Quantitative Analysis

FSG and CFG(++) × 3 improve fine-grained instruction compliance

Method	Overall↑	Single Object↑	Two Object↑	Counting↑	Colors↑	Position↑	Color Attribution↑
CFG	48.39 %	97.50 %	61.62 %	22.50 %	78.72 %	**14.00 %**	16.00 %
CFG×3	55.94 %	98.75 %	75.76 %	40.00 %	85.11 %	8.00 %	**28.00 %**
CFG++×3	56.03 %	97.50 %	78.79 %	45.00 %	81.91 %	10.00 %	23.00 %
Z-sampling	56.70 %	**100.00 %**	75.76 %	**46.25 %**	**86.17 %**	12.00 %	20.00 %
Resampling	56.65 %	**100.00 %**	**84.85 %**	40.00 %	84.04 %	7.00 %	24.00 %
FSG (ours)	**57.95 %**	**100.00 %**	79.80 %	43.75 %	**86.17 %**	12.00 %	**28.00 %**

Our framework easily extend to other models and sampler

Models	SD-2.1, DDIM				Hunyuan-DiT, DDIM				SDXL, DDPM			
Method	IR↑	HPSv2↑	AES↑	CLIP↑	IR↑	HPSv2↑	AES↑	CLIP↑	IR↑	HPSv2↑	AES↑	CLIP↑
CFG	-62.83	25.51	5.88	30.38	115.63	29.00	6.82	33.09	73.57	28.42	6.73	33.62
CFG ×3	1.08	27.25	5.92	32.50	115.32	28.98	6.50	32.88	91.37	**28.69**	6.62	33.73
CFG++×3	3.98	27.24	5.96	32.51	115.63	29.03	6.63	32.97	89.17	28.79	6.60	33.60
Z-sampling	3.65	27.24	6.07	32.60	128.72	29.23	6.72	33.42	90.35	28.63	**6.65**	33.58
Resampling	8.07	27.03	5.85	32.31	117.65	29.28	**6.73**	33.18	90.92	28.64	**6.65**	33.73
FSG (ours)	**16.26**	**27.60**	**6.10**	**32.80**	**132.88**	**29.37**	6.68	**33.48**	**91.53**	28.56	**6.65**	**33.79**

Qualitative analysis

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Prompts:

A bowl of soup that looks like a monster with tofu says deep learning
↑ Prompt details and text

Five frosted glass bottles
↑ Counting

A cat in a space suit walking on the moon
↑ Image details

A storefront with 'Google Research Pizza Cafe' written on it.
↑ Long text

One computer technical sketch white background
↑ Prompt alignment

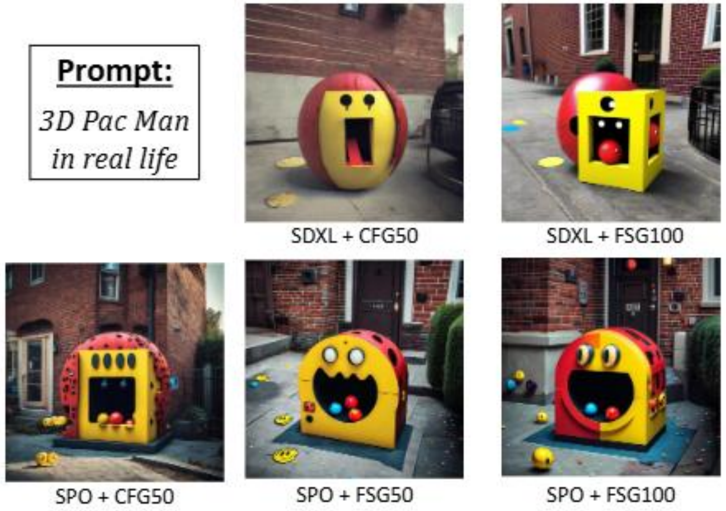
Combination with orthogonal methods

FSG can achieve “1+1>2” gains when combined with
noise optimization / preference-aligned models

Better initial values for fixed-point iteration

Methods	IR ↑	HPSv2 ↑	AES ↑	CLIP ↑
Standard (CFG)	82.14	28.46	6.73	33.53
+NPNet	91.66	28.60	6.70	33.57
FSG50	98.59	28.89	6.60	34.32
FSG100	102.82	29.05	6.66	34.30
FSG50+NPNet	112.64	29.04	6.54	34.09
FSG100+NPNet	111.83	29.15	6.57	34.13

Method	IR↑	HPSv2↑	AES↑	CLIP↑
CFG50	82.14	28.46	6.73	33.53
FSG100	102.82	29.05	6.66	34.30
SPO	111.86	29.08	6.91	33.22
+FSG50	115.86	29.16	6.91	33.12
+FSG100	**117.93**	**29.20**	**6.93**	33.24
+FSG150	116.49	28.74	6.85	**33.30**



Conclusion

香港科技大學 THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

- We propose fixed point iteration as unified guidance framework with broader design applicability.
- We unified existing CFG and its variants as short-sighted single-step approaches.
- We present Foresight Guidance (FSG), a multi-step iteration paradigm with longer interval, achieving enhanced alignment-quality balance.

Code: https://github.com/Ka1b0/Foresight-Guidance